

**Towards numerical modelling of natural
subduction systems with an application to
Eastern Caribbean subduction**

Menno R. T. Fraters

UTRECHT STUDIES IN EARTH SCIENCES

No. 185

Members of the dissertation committee:

Prof. Dr. Magali Billen
Dept of Earth & Planetary Sciences
UC Davis, USA

Prof. Dr. Nicolas Coltice
Département de Géosciences
Ecole Normale Supérieure de Paris, France

Prof. Dr. Clint Conrad
Centre for Earth Evolution and Dynamics (CEED)
Department of Geosciences
University of Oslo, Norway

Prof. Dr. Boris Kaus
Institute of Geosciences
Johannes Gutenberg University Mainz, Germany

Prof. Dr. Wouter Schellart
Faculty of Science, Geology and Geochemistry
Vrije Universiteit Amsterdam, Netherlands

Copyright © 2019 Menno Fraters, Utrecht University

All rights reserved. No part of this publication may be reproduced in any form, by print or photographic print, microfilm or any other means, without written permission by the author.

Printed in the Netherlands by Ipskamp Printing, Enschede

ISBN: 978-90-6266-540-2

Towards numerical modelling of natural subduction systems with an application to Eastern Caribbean subduction

Een opstap naar numerieke modellering van natuurlijke subductie
systemen met een toepassing op het oostlijke Caribisch gebied
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H. R. B. M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
woensdag 15 mei 2019 des ochtends te 10.30 uur

door

Menno Roeland Theodorus Fraters
geboren op 3 oktober 1989 te Wageningen

Promotoren:

Prof. dr. W. Spakman

Prof. dr. D.J.J van Hinsbergen

Prof. dr. W. Bangerth

Copromotor:

Dr. C.A.P. Thieulot

This work is funded by the Netherlands Organization for Scientific Research (NWO), as part of the Caribbean Research program, grant 858.14.070. This work was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 223272., the Netherlands Research Centre for Integrated Solid Earth Science (ISES), the Computational Infrastructure in Geodynamics initiative (CIG), through the National Science Foundation under Award No. EAR-0949446 and The University of California - Davis, and by the National Science Foundation under awards OCI-1148116 and OAC-1835673 as part of the Software Infrastructure for Sustained Innovation (SI2) program (now the Cyberinfrastructure for Sustained Scientific Innovation, CSSI).

**Il est encore plus facile de juger de l'esprit d'un homme
par ses questions que par ses réponses.**

*It is easier to judge the mind of a man by his questions
than his answers.*

Pierre-Marc-Gaston de Lévis

Maximes, préceptes et réflexions sur différens sujets
de morale et de politique. Paris (1808), Maxim xviii

Contents

Contents	vii
1 Introduction	1
2 Efficient and Practical Newton Solvers for Nonlinear Stokes Systems in Geodynamic Problems	7
2.1 Summary	8
2.2 Introduction	8
2.3 Problem statement and numerical methods	11
2.3.1 The model	11
2.3.2 Discretization	13
2.3.3 Newton linearization	14
2.3.4 Restoring symmetry of J^{uu}	16
2.3.5 Restoring well-posedness of the Newton step	18
2.3.6 Algorithms for the solution of the nonlinear problem	22
2.4 Numerical experiments using common benchmarks	26
2.4.1 Nonlinear channel flow	26
2.4.2 Spiegelman et al. benchmark	28
2.4.3 Tosi et al. benchmark	33
2.4.4 A 3d subduction test case	34
2.5 Conclusions	40
2.6 Acknowledgments	41
2.7 Appendix A: The connection between elliptic operators, well-posedness of the Newton update equation, and eigenvalues of coefficients	41
2.8 Appendix B: A look at some common rheologies	42
2.8.1 Power law rheology	43
2.8.2 Drucker-Prager rheology	44
2.8.3 The rheology of the Spiegelman et al. benchmark	45

2.8.4	The rheology of the Tosi et al. benchmark	46
2.9	Appendix C: Parameters for the 3d subduction test case	46
3	The Geodynamic World Builder: a solution for complex initial conditions in numerical modelling	49
3.1	Abstract	50
3.2	Introduction	50
3.3	Geodynamic World Builder Philosophy	51
3.3.1	User Philosophy	51
3.3.2	Code philosophy	53
3.4	Using the World Builder	54
3.4.1	Standalone examples	55
3.4.2	Using the GWB with SEPRAN	58
3.4.3	Using the GWB with ELEFANT	59
3.4.4	Using the GWB with ASPECT	59
3.4.5	Performance	64
3.5	Discussion	64
3.6	Code availability	64
3.7	Acknowledgements	65
3.8	2D subduction examples	65
3.8.1	Cartesian input file	65
3.8.2	Spherical input file	66
3.9	3D ocean spreading example input file	68
3.10	3D subduction example input file	68
3.11	SEPRAN 2D subduction	70
3.12	ELEFANT 3D Double subduction setup	72
3.13	ASPECT 3d curved subduction	73
4	Assessing the geodynamics of strongly arcuate subduction zones: the eastern Caribbean subduction setting.	75
4.1	Summary	76
4.2	Introduction	76
4.3	Three-dimensional initial model of arcuate subduction based on the tectonic setting of the Lesser Antilles slab	79
4.4	Model setup	84
4.4.1	Numerical model setup	84
4.4.2	Boundary conditions	86
4.4.3	Rheological model	86
4.5	Experiments	90

4.5.1	The reference model	90
4.5.2	The influence of the crustal rheology of the subducting plate	96
4.5.3	The influence of mantle viscosity and the temperature of the slab	97
4.5.4	The influence of vertical weak zones in the slab	102
4.5.5	The influence of the subducting plate velocity	104
4.6	Discussion and conclusions	107
4.7	Appendix A: GWB input	110
4.8	Appendix B: ASPECT input	113
5	Conclusions and outlook	119
	Bibliography	123
	Summary	135
	Samenvatting	137
	Acknowledgements	139
	Funding	143
	Curriculum Vitae	145

1

Introduction

1

It has long been understood that there is a strong coupling between processes of Earth's mantle and the tectonic evolution but the nature of the dynamic processes involved and their interaction continues to be a main topic in geodynamic research. Sub-disciplines of Solid Earth Science, like geology or seismology, provide various valuable observations of the internal workings of our planet but usually provide only indirect information on acting processes. Direct access to the operation of deep processes is provided by forward numerical modelling, solving the pertinent conservation laws of continuum mechanics, which ultimately may provide useful predictions of geological, seismological, or geodetic observations such that these are better understood and that hypotheses of acting mantle processes can be tested against observations.

My goal in this thesis is to make new steps toward assessing how mantle processes operate and affect tectonic evolution, particularly, by creating steps toward numerically simulating natural plate boundary regions involving lithosphere subduction. By far most of current 3D subduction modelling is of a strongly generic nature aimed at understanding the process and its interaction with the mantle but not so much on simulating natural subduction in past or present. Subduction plate boundaries are simplified in initial models to straight subduction trenches and pure trench-perpendicular convergence instead of implementing the natural geometry of subduction zones that comprises highly variable trench geometry and the generally oblique convergence of the subducting plate towards the trench. Also, often simplified linear material-rheology is used. Of course, there are several reasons why this characterizes the current stage of 3D subduction modelling. Some of these are: subduction in 3D space is a complex process that requires step-by-step investigation; obtaining a feasible computation time; exploiting the possibility to use simple free-slip and no-slip boundaries along the side-walls of the model domain; or the practical difficulty of constructing 3-D initial models that mimic the geometrical complexity of natural subduction in an embryonic, or advanced stage of evolution.

In this thesis, I aim to address the research problems of 1) decreasing the computation time, while improving the accuracy, of subduction modelling involving realistic non-linear material models, here visco-plasticity, and more advanced boundary conditions, of 2) establishing a versatile method for building complex 3D geodynamic models of subduction for use as initial condition of subduction modelling, and 3) the research problem of the geodynamics of strongly arcuate subduction. Addressing the latter research problem builds on providing solutions to the first problems and is intended as a showcase of the

practical feasibility of coming much closer to simulating natural evolution in all its geometrical complexity.

Chapter 2 is devoted to providing a solution to the first research problem. Depending on ambient temperature and pressure, mantle materials exhibit often a nonlinear response of strain-rate to deforming stress. Using nonlinear rheology causes the Stokes equation, describing the force balance, to become nonlinear which consequently leads in Finite Element as well as Finite Difference applications to a system of nonlinear equations to be resolved. The standard and stable way to tackle the nonlinearity is by setting up, for each time step, an iteration scheme by the Picard method which is robust and numerically the easiest way to solve the nonlinear system of equations. However, complex geodynamic problems, e.g., simulating natural mantle processes, are characterized by strong nonlinear viscosity and strain-rate variation of several orders of magnitudes occurring over short distances, often causing the Picard iteration to slowly converge in each time step, i.e. considerably increasing the overall computation time.

A more advanced method is the Newton method that uses information about the derivatives of the nonlinear equations. Under several conditions, this method can converge much faster to the correct solution of the nonlinear system of equations. However, the Newton method for the Stokes equations is not always stable and may lead to systems of equations for which the linear solver can not converge, because the involved matrix is singular. To develop a stable and (almost always) converging Newton method that also leads to a considerable decrease in computation time, is the main goal of Chapter 2.

Chapter 3 tackles the second research problem of finding an effective and versatile method for constructing 3D geodynamic initial models that allow to approximate the geometrical complexity of natural plate boundaries and 3D subduction. Setting up initial conditions for geodynamic models, especially in 3D, is so far done by personally developed methods and code, which can be a very time-consuming process and so far has not led to a more general, easy to use, open-source solution accessible to the modelling community. Chapter 3 aims to offer exactly that versatile solution, called the Geodynamic World Builder, such that it is not only relatively easy to use but also provides a simple interface for use with the various community, or personal, numerical modelling codes.

Chapter 4 concerns the problem of the geodynamics involved in arcuate subduction along strongly curved trenches. Trench curvature is still not well understood but is a common characteristic of natural subduction zones across

the globe. The negative buoyancy of the subducting slab exerts a vertical force on the lithosphere entering the trench, slab pull, that basically acts trench-perpendicular but still has an arc-forming potential by its interaction with lateral lithosphere heterogeneity and with slab-induced mantle flow. Slab pull, however, cannot easily explain the generally observed trench-oblique convergence of tectonic plates along strongly arcuate trenches. Particularly, arcuate subduction systems, such as the Marianas-Izu-Bonin, or Aleutians-Alaska systems, are characterized by trench-perpendicular convergence at one trench segment (Marianas, Alaska) while a strong trench-parallel component of subducting plate motion occurs elsewhere along the trench. This suggests that the slab subducting at the latter trench segment may be involved in lateral transport through the mantle. In Chapter 4, I investigate this possible geodynamic aspect of arcuate subduction for the Lesser Antilles system of the eastern Caribbean where near trench-parallel westward subducting plate motion occurs along the northern trench below which a south-dipping slab is observed by seismic tomography. This application to a geometrically complex natural subduction system showcases the benefits of the Newton method developed in Chapter 2, i.e. stable and fast convergence, and the benefits of the Geodynamic World Builder for setting up an application to natural subduction. It also intends to demonstrate that more detailed and elaborate applications to natural subduction systems are now within reach.

A common thread across all chapters is the use and development of the ASPECT code, which I would like to introduce here. ASPECT (short for Advanced Solver for Problems in Earth's ConvecTion) is a modern C++ code designed to solve the equations underpinning the geodynamic processes of the crust-mantle system (Kronbichler et al., 2012; Heister et al., 2017). Although originally designed to model convection related problems, its modular design has allowed a growing core of developers and users that have quickly expanded ASPECT's range of applications to thermo-mechanical modelling of the crust-mantle system using visco-plastic rheology (Glerum et al., 2018), free surface (Rose et al., 2017), melt transport (Dannberg and Heister, 2016; Dannberg and Gasmöller, 2018), or even grain size evolution (Dannberg et al., 2017). ASPECT is hosted by the Computational Infrastructure for Geodynamics¹ (CIG). The code is based on the deal.II Finite Element library² (Bangerth et al., 2007;

¹www.geodynamics.org

²<https://www.dealii.org/>

Alzetta et al., 2018) and employs numerical methods which are at the forefront of research, such as adaptive mesh refinement (AMR) based on the p4est library³ (Burstedde et al., 2011), linear and nonlinear solvers from Trilinos⁴ (Heroux and Willenbring, 2012), and stabilization of transport-dominated processes (Guermont et al., 2010, 2011). It also allows for various domain geometries, such as 2D and 3D Cartesian, 2D Cylindrical, and 3D Spherical domains encompassing the entire crust-mantle system or parts of it, (hollow) spheres and even (parts of) ellipsoidal meshes. From the start, the code has been designed to support high levels of parallelism and has been run and shown to scale on up to tens of thousands of cores. The current version is hosted on the github platform⁵ while stable releases are also available on the CIG website⁶. The code comes with an embedded and automatically generated manual which is extensive and up-to-date. Users are encouraged to get involved and to submit either code or so-called cookbooks (geodynamically relevant example cases) which are thoroughly reviewed by core developers before being merged into the main version. Finally, every change to the code is tested before becoming part of the main code by more than 600 integration tests (and growing) and more than 100 unit tests assertions (and growing) on a dedicated server to ensure that none of the existing features of the code is ever broken, thereby ensuring backwards compatibility.

My own investments in the development of ASPECT have provided me with the advanced research modelling tool that constitutes the solid base of the research describe in my thesis.

³<http://www.p4est.org/>

⁴<http://trilinos.org/>

⁵<https://github.com/geodynamics/aspect>

⁶<https://aspect.geodynamics.org/>

2

Efficient and Practical Newton Solvers for Nonlinear Stokes Systems in Geodynamic Problems

2.1 Summary

Many problems in geodynamic modeling result in a nonlinear Stokes problem in which the viscosity depends on the strain rate and pressure (in addition to other variables). After discretization, the resulting nonlinear system is most commonly solved using a Picard fixed-point iteration. However, it is well understood that Newton's method – when augmented by globalization strategies to ensure convergence even from points far from the solution – can be substantially more efficient and accurate than a Picard solver.

In this contribution, we evaluate how a straight-forward Newton method must be modified to allow for the kinds of rheologies common in geodynamics. Specifically, we show that the Newton step is not actually well-posed for strain rate-weakening models without modifications to the Newton matrix. We derive modifications that guarantee well-posedness and that also allow for efficient solution strategies by ensuring that the top-left block of the Newton matrix is symmetric and positive definite. We demonstrate the applicability and relevance of these modifications with a sequence of benchmarks and a test case of realistic complexity.

2.2 Introduction

Geodynamics aims to understand the dynamics of processes in and on the Earth on a wide range of spatial and temporal scales, typically by connecting physical processes to geological observations through either analogue or numerical modeling. The physical basis of most numerical modeling codes in the geodynamics community are continuum mechanics conservation laws for momentum, mass, and energy. A fundamental assumption that underlies most models is that we can average over the small scales at which natural materials exhibit heterogeneity, and that we can approximate the *macroscopic* material properties to obtain equations that are well understood.

When considering long enough time scales, the dynamics of the mantle – and, to some degree, the crust – can then be described as a slow-moving fluid that is governed by the Stokes equations, together with advection-diffusion equations for the temperature, chemical compositions, and possibly other

This chapter is currently under review at Geophysical Journal International as *Efficient and Practical Newton Solvers for Nonlinear Stokes Systems in Geodynamic Problems*, M. Fraters, W. Bangerth, C. Thieulot and W. Spakman.

quantities. In the case of the Stokes equations, the fluid's effective viscosity will then depend on the material's temperature, pressure, composition, and possibly other factors such as mechanical stress. Rheology – the science of determining how a material flows – is therefore of key importance to this approach. Unfortunately, the rheology of Earth materials over geological timescales is also one of the least constrained ingredients in modeling the physical processes of the solid Earth. For both philosophical and computational reasons, many studies use linear rheologies (e.g. Baumann et al., 2014; Fritzell et al., 2016; Pusok and Kaus, 2015), i.e., a viscosity that may depend on the external temperature and chemical composition, but not on the fluid variables velocity (or its derivatives, e.g., the strain rate) and pressure. However, experiments have shown that the rheology of Earth materials can behave in a very nonlinear way (Karato and Wu, 1993). Specifically, in deformation regimes, the mechanical stress leads to material weakening with increasing strain-rate and consequently an effective viscosity that is a decreasing function of the strain rate. Furthermore, many widely used rheological models – in particular if they try to incorporate plastic effects – include a pressure dependence of the viscosity. This nonlinearity of the rheology results in models best described by a nonlinear variation of the Stokes equations. An additional source of nonlinearities arises from the fact that Earth materials are compressible, i.e., that their density depends on the pressure. Because there is no convenient way of solving this kind of nonlinear partial differential equation exactly, it is important to develop numerical methods that can discretize and iteratively resolve the nonlinearity in the equations.

A simple and frequently used way to solve such nonlinear problems is to use Picard iterations, a particular form of fixed point iterations (Kelly, 1995). In it, one computes the viscosity and density as a function of the previous iteration's strain rate and pressure, solves for a new velocity and pressure field, and then repeats the process. The Picard iteration owes its popularity to the fact that it is relatively easy to implement in codes that only support linear rheologies because it only requires the repeated solution of linear problems. It is also often globally convergent, i.e., with sufficiently many iterations it is possible to approximate the solution of the nonlinear problem regardless of the choice of initial guess. Consequently, it is the method that is likely used in the majority of mantle convection papers that actually iteratively resolve the nonlinearity in each time step; most papers do not explicitly state so, but van Keken et al. (2008); Tosi et al. (2015); Buitert et al. (2016); Glerum et al. (2018) are some examples.

On the other hand, Picard iterations are often slow to converge, requiring

dozens or hundreds of iterations for strongly nonlinear problems – something we also observe in our results in Section 2.4. This slow convergence may make the solution of nonlinear problems to high accuracy prohibitively expensive. Consequently, commonly used approaches to cope with the high computational cost are, for example, limiting the allowed number of Picard iterations per timestep (e.g. Lemiale et al., 2008), combining Picard iterations with small timesteps to ensure good starting guesses (e.g. Ruh et al., 2013), or other mostly *ad hoc* approaches. In practice, however, many studies do not adequately document the exact algorithm used and how this affects the accuracy of the solutions of the equations considered.

Here, we address the slow convergence of nonlinear solvers by replacing the Picard solver by a Newton solver (Kelly, 1995). Previous applications of Newton’s method to geodynamics problems can be found in May et al. (2015); Rudi et al. (2015); Kaus et al. (2016); Spiegelman et al. (2016). Newton’s method promises quadratic convergence towards the solution, compared to the linear convergence of the Picard iteration, when the initial guess is close enough to the solution of the nonlinear problem and therefore offers the prospect of vastly faster solution procedures. On the other hand, the implementation of Newton’s method is substantially more involved than a Picard iteration. Furthermore, requiring an initial guess that is close enough to the exact solution is often impractical, and may require running a number of initial Picard iterations before starting the Newton iteration.

In this paper we present the details of an improvement on the Newton method for the nonlinear Stokes problem, and discuss an implementation of this improved Newton solver along with recommendations on how to use it. Specifically, and going beyond what is available in the literature, we will show that a naive application of Newton’s method may break both the symmetry and the positive definiteness of the elliptic part of the (linearized) Jacobian of the Stokes operator. While the lack of symmetry is annoying from a practical perspective because it makes the solution of the linear system associated with each Newton step more complicated, a lack of positive definiteness implies that the Newton step is ill-posed and may not have a solution. We will analyze both of these issues in detail and propose modifications to the Newton equations that retain the symmetry and restore the positive definiteness. We will also consider whether there are special classes of material models where these modifications are not necessary. Unfortunately, as we will show, many rheologies that have been used extensively in the literature do not fall into these classes; our methods are therefore strict improvements over the current state-of-the-art and will

allow solving problems that were not previously solvable with an unmodified Newton method.

While there are previous reports on using a Newton method for Stokes problems in geodynamics applications (see, for example, May et al. (2015); Rudi et al. (2015); Kaus et al. (2016); Spiegelman et al. (2016)), we will provide a more in-depth discussion of the mathematical properties of the operators and linear systems associated with each Newton step. We will underpin our claims with numerical experiments and demonstrate that the approach advocated herein is, indeed, more efficient and robust than previous approaches. In particular, we will show that our implementation of the Newton solver significantly decreases computational time for realistic problems, with greatly improved accuracy. Our implementation is available as open source as part of the ASPECT code (Kronbichler et al., 2012; Heister et al., 2017), an open source geodynamics community code.

The layout of the remainder of this paper is as follows: We will first describe the mathematical formulation of the nonlinear Stokes problem we consider here, its discretization, and linearization in Section 2.3. This section also contains our main results on how the Newton method has to be modified (“stabilized”) in order to make it well-posed, as well as a discussion of practical aspects of how this method can be embedded in efficient nonlinear and linear solvers. We then show how the above works in practice in Section 2.4, first using three artificial test cases and then using a realistic application of modeling subduction. We conclude in Section 2.5.

2.3 Problem statement and numerical methods

2.3.1 The model

Let us begin by concisely stating the equations we want to solve herein. We are concerned with modeling convection in the Earth mantle, a process that is typically described by a coupled system of differential equations. Under commonly used assumptions – see for example Schubert et al. (2001) – typical models include a Stokes-like, compressible fluid flow system for the velocity \mathbf{u} and pressure p defined in the volume $\Omega \subset \mathbb{R}^d$ (where the space dimension $d = 2$ or 3) under consideration,

$$-\nabla \cdot \left[2\eta \left(\epsilon(\mathbf{u}) - \frac{1}{3}(\nabla \cdot \mathbf{u})\mathbf{I} \right) \right] + \nabla p = \rho \mathbf{g} \quad \text{in } \Omega, \quad (2.1)$$

$$-\nabla \cdot (\rho \mathbf{u}) = 0 \quad \text{in } \Omega, \quad (2.2)$$

where η is the viscosity, ρ the density, \mathbf{g} the gravity vector, $\epsilon(\cdot)$ denotes the symmetric gradient operator defined by $\epsilon(\mathbf{v}) = \frac{1}{2}(\nabla\mathbf{v} + \nabla\mathbf{v}^T)$, and I is the $d \times d$ identity matrix. (The sign in (2.2) is chosen in this way because $-\nabla \cdot$ is the adjoint operator to the gradient in the first equation, leading to a symmetric system if the density is constant, as shown below.)

While these equations describe a compressible model, we will assume for the purposes of this paper that the fluid is in fact incompressible, i.e., that $\nabla \cdot (\rho\mathbf{u}) = \rho\nabla \cdot \mathbf{u} = 0$. We do so because we can illustrate all difficulties associated with the Newton method using this simplification already, and because many of the approximations used in geodynamics (for example, the Boussinesq approximation) also assume incompressibility. In addition to this simplification, we have to scale the equations to ensure that we can numerically compare the residuals of the two equations and consequently have a basis for numerically stable algorithms. Consequently, we multiply the second equation by a constant $s_p = \frac{\eta_0}{L}$ where η_0 is a “reference viscosity” and L a length scale of the domain we are solving the equations in. (See Kronbichler et al. (2012) for a more detailed discussion.) In order to retain the symmetry between the divergence in the second equation and the gradient in the first, we also replace the pressure by a scaled version, $\bar{p} = \frac{1}{s_p}p$. The properly scaled, incompressible equations then read as follows:

$$-\nabla \cdot [2\eta\epsilon(\mathbf{u})] + s_p\nabla\bar{p} = \rho\mathbf{g}, \quad (2.3)$$

$$-s_p\nabla \cdot \mathbf{u} = 0. \quad (2.4)$$

It is this form of the equations we will attempt to solve, using \mathbf{u}, \bar{p} as the primary variables. Of course, the physical pressure can be recovered as $p = s_p\bar{p}$ after the system has been solved.

In geodynamic models, the fluid flow model is coupled to an equation for the temperature T ,

$$\begin{aligned} \rho C_p \left(\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T \right) - \nabla \cdot k \nabla T = \rho H \\ + 2\eta \left(\epsilon(\mathbf{u}) - \frac{1}{3}(\nabla \cdot \mathbf{u})\mathbf{I} \right) : \left(\epsilon(\mathbf{u}) - \frac{1}{3}(\nabla \cdot \mathbf{u})\mathbf{I} \right) \\ + \alpha_T T (\mathbf{u} \cdot \nabla p) \\ + \rho T \Delta S \left(\frac{\partial X}{\partial t} + \mathbf{u} \cdot \nabla X \right) \end{aligned} \quad \text{in } \Omega, \quad (2.5)$$

and possibly other equations that describe the transport of chemical compositions. Here, C_p is the specific heat, α_T is the thermal expansion coefficient, k the thermal conductivity, H is the internal heat production, and ΔS and X are related to the entropic effects of phase changes. All coefficients that appear in these equations typically depend on the pressure, temperature, chemical composition, and – in the case of the viscosity – the strain rate $\varepsilon(\mathbf{u})$.

Even though the *entire* system is coupled in nonlinear ways, in this paper, we will only concern ourselves with the first set of these equations, (2.3)–(2.4), and how they can efficiently be solved through a Newton scheme. In principle, one may want to solve the entire system with a Newton scheme, given that the velocity appears in (2.5), the temperature in (2.3)–(2.4) via the temperature dependence of the viscosity and density, and more generally all coefficients may depend on pressure and temperature. While this is beyond the scope of the current paper, being able to apply a Newton method to the Stokes sub-system is clearly a necessary ingredient to the larger goal. Consequently, the efficient solution of nonlinear Stokes problems is of interest in itself. As we will show, this alone is not trivial, and will therefore serve as a worthwhile target for the investigations in this paper. In fact, the incompressible formulation already poses all of the mathematical difficulties we will encounter in deriving well-posed Newton schemes. In other words, it serves as a good model problem to illustrate and understand both difficulties and solutions related to the linearization. The incorporation of compressible terms (i.e., solving (2.1)–(2.2)) would then only complicate the exposition of our methods. At the same time, we point out that our methods immediately carry over to compressible models – albeit with significantly more cumbersome formulas; we will investigate this generalization in future work.

2.3.2 Discretization

We convert equation (2.3)–(2.4) above into a finite-dimensional system by utilizing the finite element method for discretization. To this end, we seek approximations

$$\mathbf{u}_h(\mathbf{x}) = \sum_j U_j \varphi_j^u(\mathbf{x}) \tag{2.6}$$

$$\bar{p}_h(\mathbf{x}) = \sum_j \bar{P}_j \varphi_j^p(\mathbf{x}) \tag{2.7}$$

where φ_j^u and φ_j^p are the finite element basis functions for the velocity and pressure, respectively.

The expansion coefficients U_j, \bar{P}_j are found by solving the discrete weak form of the equations. Discretization of the incompressible system then leads to a nonlinear system in $\mathbf{X} = (\mathbf{U}, \bar{\mathbf{P}})$,

$$\mathbf{Q}(\mathbf{X}) \mathbf{X} = \mathbf{b}(\mathbf{X}), \quad (2.8)$$

where the matrix \mathbf{Q} and right hand side \mathbf{b} have an internal sub-structure. For our incompressible formulation, this substructure has the form

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{U} \\ \bar{\mathbf{P}} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{h} \end{pmatrix}. \quad (2.9)$$

Here, the matrix and right hand side blocks are defined as

$$A_{ij} = (\varepsilon(\varphi_i^u), 2\eta\varepsilon(\varphi_j^u)), \quad B_{ij} = -s_p(\varphi_i^q, \nabla \cdot \varphi_j^u), \quad (2.10)$$

$$f_i = (\varphi_i^u, \rho g), \quad h_i = 0, \quad (2.11)$$

where as usual we denote $(\alpha, \beta) = \int_{\Omega} \alpha(\mathbf{x}) \beta(\mathbf{x}) \, d\mathbf{x}$. Because the viscosity η may depend on the pressure and strain rate, and the density ρ on the pressure, the system is in general nonlinear in the coefficients U_j, \bar{P}_j as both $\mathbf{A} = \mathbf{A}(\mathbf{X})$ and $\mathbf{f} = \mathbf{f}(\mathbf{X})$. (The coefficients η, ρ may of course also depend on the temperature or other factors, but we consider these fixed for the purposes of the current paper.)

Much of the content of this paper is concerned with the question of how to solve the nonlinear system (2.8) *in practice*, i.e., how a naive application of the standard Newton iteration solver needs to be adapted to make it practical and efficient.

2.3.3 Newton linearization

In order to resolve the nonlinearity in equation (2.8), let us introduce the residual $\mathbf{r}(\mathbf{X}) = \mathbf{Q}(\mathbf{X})\mathbf{X} - \mathbf{b}(\mathbf{X})$. In Newton iteration $k + 1$, starting with the previous guess \mathbf{X}_k , we then need to solve

$$\mathbf{J}_k \delta \mathbf{X}_k = -\mathbf{r}_k \quad (2.12)$$

where $\mathbf{r}_k = \mathbf{r}(\mathbf{X}_k)$ and $\mathbf{J}_k = \nabla_{\mathbf{X}} \mathbf{r}(\mathbf{X}_k)$. This system has the internal sub-structure

$$\begin{pmatrix} \mathbf{J}_k^{uu} & \mathbf{J}_k^{up} \\ \mathbf{J}_k^{pu} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \delta \mathbf{U}_k \\ \delta \bar{\mathbf{P}}_k \end{pmatrix} = - \begin{pmatrix} \mathbf{r}_k^u \\ \mathbf{r}_k^p \end{pmatrix}. \quad (2.13)$$

After solving for $\delta \mathbf{X}_k$, we can compute $\mathbf{X}_{k+1} = \mathbf{X}_k + \alpha_k \delta \mathbf{X}_k$ where α_k is a step length parameter that can be determined, for example, using a line search (Kelly, 1995; Nocedal and Wright, 1999).

There are a number of approaches to determining the entries of the matrix \mathbf{J}_k and to solving the resulting linear system. For example, in the geodynamics community alone, May et al. (2015) and Kaus et al. (2016) make use of a Jacobian-free Newton-Krylov (JFNK) framework (see Knoll and Keyes (2004)), which essentially computes a finite difference approximation of \mathbf{J} by evaluating \mathbf{r} at different values of its argument, and integrates this directly into the solver so that the full Jacobian matrix is never built. On the other hand, Rudi et al. (2015) and Spiegelman et al. (2016) use the same approach as we will take here and compute derivatives analytically or semi-analytically, except that Rudi et al. (2015) implemented this in a Jacobian-free manner.

Regardless of how exactly these derivatives are computed, the blocks of the linear system for the Newton updates will have to have the following form (again omitting dependencies on quantities we consider frozen, such as the temperature):

$$\begin{aligned} (\mathbf{J}_k^{uu})_{ij} &= \frac{\partial}{\partial U_j} \left(\mathbf{A}_k \mathbf{U}_k + \mathbf{B}^\top \bar{\mathbf{P}}_k - \mathbf{f}_k \right)_i \\ &= (\mathbf{A}_k)_{ij} + \left(\varepsilon(\varphi_i^u), 2 \left(\frac{\partial \eta(\varepsilon(\mathbf{u}_k), p_k)}{\partial \varepsilon} : \varepsilon(\varphi_j^u) \right) \varepsilon(\mathbf{u}_k) \right), \end{aligned} \quad (2.14)$$

$$\begin{aligned} (\mathbf{J}_k^{up})_{ij} &= \frac{\partial}{\partial \bar{p}_j} \left(\mathbf{A}_k \mathbf{U}_k + \mathbf{B}^\top \bar{\mathbf{P}}_k - \mathbf{f}_k \right)_i \\ &= \mathbf{B}_{ij}^\top + \left(\varepsilon(\varphi_i^u), 2 \left(\frac{\partial \eta(\varepsilon(\mathbf{u}_k), p_k)}{\partial \bar{p}} \varphi_j^p \right) \varepsilon(\mathbf{u}_k) \right), \\ &= \mathbf{B}_{ij}^\top + \left(\varepsilon(\varphi_i^u), 2 \left(\frac{\partial \eta(\varepsilon(\mathbf{u}_k), p_k)}{\partial p} \frac{\partial p}{\partial \bar{p}} \varphi_j^p \right) \varepsilon(\mathbf{u}_k) \right), \\ &= \mathbf{B}_{ij}^\top + s_p \left(\varepsilon(\varphi_i^u), 2 \left(\frac{\partial \eta(\varepsilon(\mathbf{u}_k), p_k)}{\partial p} \varphi_j^p \right) \varepsilon(\mathbf{u}_k) \right), \end{aligned} \quad (2.15)$$

$$\begin{aligned} (\mathbf{J}_k^{pu})_{ij} &= \frac{\partial}{\partial U_j} (\mathbf{B} \mathbf{U}_k - \mathbf{h}_k)_i \\ &= \mathbf{B}_{ij}. \end{aligned} \quad (2.16)$$

It is easy to see that – as expected – the Newton system (2.13) reverts to the simple Stokes problem if the viscosity does not depend on strain rate or pressure, i.e., if the system is linear.

As we will show below, while equation (2.12) (and its block structure (2.13)) is the correct linearization of the (discretized) original, nonlinear system (2.3)

and (2.4), it turns out that this does not necessarily lead to a well-posed problem. This is not uncommon in optimization problems where a function $f(x)$ may have a well-defined minimizer, but the Hessian matrix $\mathbf{H}_k = \nabla^2 f(x_k)$ at early iterates may be singular or have negative eigenvalues; consequently the solution of the linear system $\mathbf{H}_k \delta x_k = -\nabla f(x_k)$ may not have a solution δx_k or the solution may not be a direction of descent. There are standard techniques described in the optimization literature for these cases (see, for example, the section on “Hessian modification” methods in Nocedal and Wright (1999)) that we will adapt in the following sections, though we will work at the level of the partial differential equations that give rise to the Newton matrix, rather than at the algebraic level of the matrix we wish to modify. Furthermore, the linear system we obtain in each Newton step may be difficult to solve for practical reasons if it is not symmetric.

We will therefore discuss the practical implications of Newton linearization in Sections 2.3.4 and 2.3.5 below, along with remedies to the problems we identify. It is important to stress that the modifications we propose only change the matrix \mathbf{J}_k in (2.12) but not the right hand side. As a consequence, we can hope that the iterates \mathbf{X}_k still converge to the correct solution \mathbf{X} of (2.8), and this is indeed the case in our numerical experiments as we observe that $\|r_k\| \rightarrow 0$ as the iterations proceed. In other words, we replace an exact (though potentially ill-posed) Newton iteration by an approximate (and well-posed) Newton iteration, but we continue to solve the original physical problem.

2.3.4 Restoring symmetry of \mathbf{J}^{uu}

Even for incompressible models, given the form of the individual blocks in (2.14)–(2.16), the Newton system (2.13) is in general not symmetric. This is despite the fact that the matrix \mathbf{Q} in the nonlinear model (2.8) and in particular \mathbf{A} in (2.9) are of course symmetric, as shown in (2.10).

On the other hand, symmetry of matrices is an important property from a practical perspective because it allows for the construction of efficient solvers and preconditioners. As a consequence, we advocate replacing (2.12) by an approximation. This of course yields a different Newton update δx_k and may destroy the quadratic convergence order of the Newton method. On the other hand, we retain our ability to construct efficient solvers and preconditioners; in practice, one does not often run a large number of Newton iterations in each time step, and consequently a reduction from quadratic to possibly only super-linear convergence order may be acceptable. As pointed out above, we do not modify the right hand side of the Newton update equation and consequently converge to the solution of the original nonlinear problem.

Specifically, then, we advocate for the following approximation of (2.14):

$$\begin{aligned}
 (\mathbf{J}_k^{uu})_{ij} \approx & (\mathbf{A}_k)_{ij} + \left(\varepsilon(\varphi_i^u), \left(\frac{\partial \eta(\varepsilon(\mathbf{u}_k), p_k)}{\partial \varepsilon} : \varepsilon(\varphi_j^u) \right) \varepsilon(\mathbf{u}_k) \right) \\
 & + \left(\varepsilon(\varphi_j^u), \left(\frac{\partial \eta(\varepsilon(\mathbf{u}_k), p_k)}{\partial \varepsilon} : \varepsilon(\varphi_i^u) \right) \varepsilon(\mathbf{u}_k) \right).
 \end{aligned}$$

This approximation ensures that the top left block in (2.13) is indeed symmetric, and as we will see below, this and the modification discussed in the next section will then allow for the construction of efficient, multigrid-based preconditioners and the use of the Conjugate Gradient method. Indeed, the modification simply symmetrizes the second term in (2.14). In order to analyze the effect of the underlying approximation, it is useful to rewrite the original term in (2.14) in sum notation:

$$\begin{aligned}
 & \left(\varepsilon(\varphi_i^u), 2 \left(\frac{\partial \eta(\varepsilon(\mathbf{u}_k), p_k)}{\partial \varepsilon} : \varepsilon(\varphi_j^u) \right) \varepsilon(\mathbf{u}_k) \right) \\
 &= \int_{\Omega} \sum_{mn} \varepsilon(\varphi_i^u)_{mn} \left[\sum_{pq} 2 \frac{\partial \eta(\varepsilon(\mathbf{u}_k), p_k)}{\partial \varepsilon_{pq}} \varepsilon(\varphi_j^u)_{pq} \right] \varepsilon(\mathbf{u}_k)_{mn} \\
 &= \int_{\Omega} \sum_{mn,pq} \varepsilon(\varphi_i^u)_{mn} E(\varepsilon(\mathbf{u}_k))_{mnpq} \varepsilon(\varphi_j^u)_{pq}
 \end{aligned}$$

where the rank-4 tensor E is defined as $E(\varepsilon(\mathbf{u}))_{mnpq} = \left[2 \varepsilon(\mathbf{u})_{mn} \frac{\partial \eta(\varepsilon(\mathbf{u}), p)}{\partial \varepsilon_{pq}} \right]$. Clearly, the matrix \mathbf{J}_k^{uu} is symmetric if the tensor E is symmetric, i.e., $E_{mnpq} = E_{pqmn}$, but this is not always the case. (By its definition, we already have $E_{mnpq} = E_{nmnp} = E_{mnpq}$.) The modification we propose is equivalent to explicitly symmetrizing this tensor, i.e., replacing E_{mnpq} by $E_{mnpq}^{\text{sym}} = \frac{1}{2} (E_{mnpq} + E_{pqmn})$ and replacing the matrix in (2.14) by

$$(\mathbf{J}_k^{uu})_{ij} = (\mathbf{A}_k)_{ij} + \left(\varepsilon(\varphi_i^u), E^{\text{sym}}(\varepsilon(\mathbf{u}_k)) \varepsilon(\varphi_j^u) \right). \quad (2.17)$$

It is instructive to consider whether there are cases in which the tensor E is *already* symmetric, and replacing it by its symmetrized version consequently does not change anything. Specifically, this is the case if the viscosity $\eta(\varepsilon(\mathbf{u}))$ can be written as a scalar function of the square of the strain rate, i.e., $\eta(\varepsilon(\mathbf{u})) = f(\|\varepsilon(\mathbf{u})\|^2)$ where $\|\varepsilon\|^2 = \sum_{ij} \varepsilon_{ij}^2$. In this case, the chain rule implies that

$$\frac{\partial \eta(\varepsilon(\mathbf{u}))}{\partial \varepsilon_{pq}} = f'(\|\varepsilon(\mathbf{u})\|^2) \frac{\partial \|\varepsilon(\mathbf{u})\|^2}{\partial \varepsilon_{pq}} = 2 f'(\|\varepsilon(\mathbf{u})\|^2) \varepsilon(\mathbf{u})_{pq}.$$

We then have that $E(\boldsymbol{\varepsilon}(\mathbf{u}))_{mnpq} = 4f'(\|\boldsymbol{\varepsilon}(\mathbf{u})\|^2)\boldsymbol{\varepsilon}(\mathbf{u}_k)_{mn}\boldsymbol{\varepsilon}(\mathbf{u})_{pq}$, which satisfies the desired symmetry condition.

Furthermore, for incompressible materials, we have that $\text{trace } \boldsymbol{\varepsilon}(\mathbf{u}) = \text{div } \mathbf{u} = 0$, and in that case, the second invariant of the strain rate can be simplified to $\mathbb{I}_2(\boldsymbol{\varepsilon}(\mathbf{u})) = \frac{1}{2} [(\text{trace } \boldsymbol{\varepsilon}(\mathbf{u}))^2 - \text{trace } (\boldsymbol{\varepsilon}(\mathbf{u})^2)] = -\frac{1}{2} \text{trace } (\boldsymbol{\varepsilon}(\mathbf{u})^2) = -\frac{1}{2} \|\boldsymbol{\varepsilon}(\mathbf{u})\|^2$. In other words, for incompressible materials, the second invariant is a function of the square of the norm of the strain rate, and consequently any material model that only depends on the second invariant then also satisfies the criteria for cases where the explicit symmetrization does not actually change anything. Indeed, many incompressible material models define the viscosity only in terms of the second invariant of the strain rate, see for example Schellart and Moresi (2013). (We note that the geodynamics literature uses varying definitions for the second invariant. In contrast to the one used above, some papers use the definition $\mathbb{I}_2(\boldsymbol{\varepsilon}(\mathbf{u})) = (\frac{1}{2}\boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{u}))^{1/2} = (\frac{1}{2}\|\boldsymbol{\varepsilon}(\mathbf{u})\|^2)^{1/2}$ – see, for example, Gerya (2010, p. 56) or May et al. (2015). However, even with this convention the second invariant is a function of the square of the norm of the strain rate, and the conclusion above about material models that are functions of only the second invariant of the strain rate remains valid.)

We end by pointing out that the entire Jacobian remains non-symmetric since, in general, $\mathbf{J}^{up} \neq (\mathbf{J}^{pu})^\top$ because of the added term due to the derivative of the viscosity with regard to the pressure (see (2.15) and (2.16)). We will come back to this in Section 2.3.6.

2.3.5 Restoring well-posedness of the Newton step

The Stokes-like system (2.13) that arises from Newton linearization can only be well-posed if the top-left block is invertible. However, it turns out that this is not always the case, as we will see shortly. It is important to realize, however, that a lack of well-posedness of the Newton step is not equivalent to a lack of well-posedness of the original, nonlinear problem from which it arises. Indeed, it is easy to conceive of situations where a Newton method applied to finding solutions of one-dimensional equations $f(x) = 0$ fails because one of the intermediate iterates x_k happens to land at a location where $f'(x_k) = 0$ and the next iteration fails because there is no δx_k so that $f'(x_k)\delta x_k = -f(x_k)$. In multiple dimensions, and in particular in the case of the infinite dimensional operator from which the top-left matrix block \mathbf{J}^{uu} is derived, the situation is clearly more complex, but not much more complicated to understand.

To this end, recall that after the symmetrization discussed in the previous section, the matrix \mathbf{J}^{uu} has entries

$$\begin{aligned} (\mathbf{J}^{uu})_{ij} &= \left(\varepsilon(\varphi_i^u), 2\eta(\varepsilon(\mathbf{u}))\varepsilon(\varphi_j^u) \right) + \left(\varepsilon(\varphi_i^u), E^{\text{sym}}(\varepsilon(\mathbf{u}))\varepsilon(\varphi_j^u) \right) \\ &= \left(\varepsilon(\varphi_i^u), \underbrace{[2\eta(\varepsilon(\mathbf{u}))\mathbf{I} \otimes \mathbf{I} + E^{\text{sym}}(\varepsilon(\mathbf{u}))]}_{=:H} \varepsilon(\varphi_j^u) \right), \end{aligned}$$

where the rank-4 tensor $(\mathbf{I} \otimes \mathbf{I})_{ijkl} = \delta_{ik}\delta_{jl}$ maps a symmetric rank-2 tensor onto itself. A sufficient (though not necessary) condition for the matrix \mathbf{J}^{uu} to be invertible (i.e., to have no zero eigenvalues) is if the corresponding differential operator, $-\nabla \cdot [H\varepsilon(\bullet)]$ is elliptic. This is the case if and only if the tensor H (as a map from rank-2 symmetric tensors to rank-2 symmetric tensors) has only positive eigenvalues, i.e., if $\varepsilon : (H\varepsilon) > 0$ for all symmetric, non-zero rank-2 tensors ε . (We provide a bit more mathematical background for this connection between the coefficient H and the ellipticity of the corresponding differential equation in Appendix 2.7.)

Informally, for a strain hardening material model, $\frac{\partial\eta(\varepsilon(\mathbf{u}))}{\partial\varepsilon}$ is positive, and then so is $2\eta(\varepsilon(\mathbf{u}))\mathbf{I} \otimes \mathbf{I} + E^{\text{sym}}(\varepsilon(\mathbf{u}))$ because E^{sym} is a *positive* correction to the already positive definite tensor $2\eta\mathbf{I} \otimes \mathbf{I}$. In other words, H would then be positive as would the differential operator, and \mathbf{J}^{uu} would be an invertible matrix. The same would be true if the material model is strain weakening and if the amount of weakening is “small enough” because then the “small correction” E^{sym} does not offset the positive definiteness of $2\eta\mathbf{I} \otimes \mathbf{I}$. That said, we will need to be more formal with arguments as we are dealing with tensors instead of scalars; the remainder of the section is therefore devoted to formalizing these arguments and providing a solution to the problem.

Specifically, given the definitions above, the tensor H can be written as

$$\begin{aligned} H &= 2\eta(\varepsilon(\mathbf{u}))\mathbf{I} \otimes \mathbf{I} + E^{\text{sym}}(\varepsilon(\mathbf{u})) \\ &= 2\eta(\varepsilon(\mathbf{u}))\mathbf{I} \otimes \mathbf{I} + \varepsilon(\mathbf{u}) \otimes \frac{\partial\eta(\varepsilon(\mathbf{u}), p)}{\partial\varepsilon} + \frac{\partial\eta(\varepsilon(\mathbf{u}), p)}{\partial\varepsilon} \otimes \varepsilon(\mathbf{u}), \end{aligned}$$

i.e., H is a rank-2 update of a multiple of the identity operator. The first of these three terms has all eigenvalues equal to 2η , and the other two terms then lead to a perturbation of two of these eigenvalues corresponding to eigendirections that are spanned by $\varepsilon(\mathbf{u})$ and $\frac{\partial\eta(\varepsilon(\mathbf{u}), p)}{\partial\varepsilon}$. As mentioned above, unless a material’s strain weakening rate is sufficiently small, these perturbations may be strong enough to make one or both of the perturbed eigenvalues negative, and in this case the Newton-step fails to be well-posed.

To avoid this, we introduce a tensor

$$\begin{aligned} H^{\text{spd}} &= 2\eta(\varepsilon(\mathbf{u}))\mathbf{I} \otimes \mathbf{I} + \alpha E^{\text{sym}}(\varepsilon(\mathbf{u})) \\ &= 2\eta(\varepsilon(\mathbf{u}))\mathbf{I} \otimes \mathbf{I} + \alpha \left[\varepsilon(\mathbf{u}) \otimes \frac{\partial \eta(\varepsilon(\mathbf{u}), p)}{\partial \varepsilon} + \frac{\partial \eta(\varepsilon(\mathbf{u}), p)}{\partial \varepsilon} \otimes \varepsilon(\mathbf{u}) \right], \end{aligned}$$

where $0 < \alpha \leq 1$ is chosen in such a way that H^{spd} is positive definite. Using this modified form of H^{spd} at every quadrature point at which we perform the integration of the bilinear form for the Newton matrix, we then build the matrix \mathbf{J}^{uu} used in the iteration. As before, since we do not change the right hand side of the Newton update equation, we converge to the solution of the original nonlinear problem.

Clearly, if $\alpha = 0$, then H^{sym} is the identity operator times 2η and has positive eigenvalues. Because the eigenvalues depend continuously on α , there must be an $\alpha > 0$ so that H^{spd} is indeed positive definite. Ideally, to retain the convergence rate of Newton's method, we would like to choose $\alpha = 1$. We therefore propose the following choice: we want to choose α so that (i) we have $\alpha = 1$ if H is already positive definite, (ii) α is as large as possible so that H^{spd} is positive definite. In practice, however, we will also choose α small enough to avoid the case where one of the eigenvalues of H^{spd} is positive but very small compared to 2η , to avoid the numerical difficulties resulting from trying to solve a linear problem with a poorly conditioned matrix \mathbf{J}^{uu} .

It turns out that we can use the rank-2 update form of H and H^{spd} to explicitly compute the value of α . Let us abbreviate $E^{\text{sym}} = a \otimes b + b \otimes a$ where $a = \varepsilon(\mathbf{u})$ and $b = \frac{\partial \eta(\varepsilon(\mathbf{u}), p)}{\partial \varepsilon}$. Then it is clear that the (non-trivial) eigenvectors of E^{sym} must lie in the plane spanned by a, b , i.e., have the form $v = \cos(\theta) \frac{a}{\|a\|} + \sin(\theta) \frac{b}{\|b\|}$. The two non-trivial eigenvalues of E^{sym} are then the ex-

tremal values of the Rayleigh quotient

$$\begin{aligned}
 R(\alpha) &= v : (E^{\text{sym}} : v) \\
 &= \left[\cos(\theta) \frac{a}{\|a\|} + \sin(\theta) \frac{b}{\|b\|} \right] : \left([a \otimes b + b \otimes a] : \left[\cos(\theta) \frac{a}{\|a\|} + \sin(\theta) \frac{b}{\|b\|} \right] \right) \\
 &= 2 \left[\cos(\theta) \|a\| + \sin(\theta) \frac{b : a}{\|b\|} \right] \left[\cos(\theta) \frac{b : a}{\|a\|} + \sin(\theta) \|b\| \right] \\
 &= 2 \left[(b : a) \cos(\theta)^2 + \left(\frac{(b : a)^2}{\|b\| \|a\|} + \|a\| \|b\| \right) \sin(\theta) \cos(\theta) + (b : a) \sin(\theta)^2 \right] \\
 &= 2 \left[(b : a) + \left(\frac{(b : a)^2}{\|b\| \|a\|} + \|a\| \|b\| \right) \sin(\theta) \cos(\theta) \right] \\
 &= 2 \left[(b : a) + \frac{1}{2} \left(\frac{(b : a)^2}{\|b\| \|a\|} + \|a\| \|b\| \right) \sin(2\theta) \right] \\
 &= 2 \left[\frac{b : a}{\|a\| \|b\|} + \frac{1}{2} \left(\frac{(b : a)^2}{\|b\|^2 \|a\|^2} + 1 \right) \sin(2\theta) \right] \|a\| \|b\|.
 \end{aligned}$$

Thus, the eigenvalues of E^{sym} are given by $\left[2 \frac{b:a}{\|a\|\|b\|} \pm \left(\frac{(b:a)^2}{\|b\|^2\|a\|^2} + 1 \right) \right] \|a\| \|b\|$. In other words, there is one positive eigenvalue $\lambda_{\max}(E^{\text{sym}}) = \left[1 + \frac{b:a}{\|a\|\|b\|} \right]^2 \|a\| \|b\|$ and one negative or zero eigenvalue $\lambda_{\min}(E^{\text{sym}}) = - \left[1 - \frac{b:a}{\|a\|\|b\|} \right]^2 \|a\| \|b\|$.

The only eigenvalue of H^{spd} we have to worry about becoming negative is therefore the one associated with the (possibly) negative eigenvalue of E^{sym} , i.e., $2\eta(\varepsilon(\mathbf{u})) - \alpha \left[1 - \frac{b:a}{\|a\|\|b\|} \right]^2 \|a\| \|b\|$. This implies that we can choose the damping factor α as follows to ensure positive semi-definiteness:

$$\alpha = \begin{cases} 1 & \text{if } \left[1 - \frac{b:a}{\|a\|\|b\|} \right]^2 \|a\| \|b\| < 2\eta(\varepsilon(\mathbf{u})) \\ \frac{2\eta(\varepsilon(\mathbf{u}))}{\left[1 - \frac{b:a}{\|a\|\|b\|} \right]^2 \|a\| \|b\|} & \text{otherwise.} \end{cases}$$

In practice, we would like to stay well away from a zero eigenvalue and instead choose α as follows:

$$\alpha = \begin{cases} 1 & \text{if } \left[1 - \frac{b:a}{\|a\|\|b\|} \right]^2 \|a\| \|b\| < c_{\text{safety}} 2\eta(\varepsilon(\mathbf{u})) \\ c_{\text{safety}} \frac{2\eta(\varepsilon(\mathbf{u}))}{\left[1 - \frac{b:a}{\|a\|\|b\|} \right]^2 \|a\| \|b\|} & \text{otherwise,} \end{cases} \quad (2.18)$$

where $0 \leq c_{\text{safety}} < 1$ is a safety factor that ensures that the smaller eigenvalue of H^{spd} is at least $(1 - c_{\text{safety}})2\eta$ and thus bounded away from zero. This computation is easily performed at every quadrature point during the assembly of

J^{uu} . This procedure then guarantees that the resulting matrix is symmetric and positive definite, implying that the Newton direction is well defined.

It is again instructive to consider whether there are cases where we can always choose $\alpha = 1$, i.e., use the unmodified Newton step (possibly up to the symmetrization discussed in the previous section). The simplest case is if $a : b = \|a\| \|b\|$ because in that case the definition of α in (2.18) always ends up in the first branch, regardless of the size of $\eta(\varepsilon(\mathbf{u}))$. Given the definition of a, b , this is specifically the case if $\frac{\partial \eta(\varepsilon(\mathbf{u}), p)}{\partial \varepsilon}$ is a *positive* multiple of $\varepsilon(\mathbf{u})$. Similarly to the discussion in the previous section, this is the case if $\eta(\varepsilon(\mathbf{u})) = f(\|\varepsilon(\mathbf{u})\|^2)$ and if $f' \geq 0$, i.e., for a strain-hardening material. It is not difficult to show that this extends to the case where the viscosity is given by a non-decreasing function $\eta(\varepsilon(\mathbf{u})) = f(\|P\varepsilon(\mathbf{u})\|^2)$ where P is an orthogonal projection applied to the strain rate; an example is the operator that extracts the deviatoric component of the strain rate.

A more interesting case is where the material exhibits strain weakening. In that case, intuitively the conditions in (2.18) imply that we can only choose $\alpha = 1$ if the material “weakens slowly enough”. Let us, for example, consider the class of materials for which $\eta(\varepsilon(\mathbf{u})) = \eta_0 \|\varepsilon(\mathbf{u})\|^{\frac{1}{n}-1}$. Such laws are typically used to describe either diffusion ($n = 1$) or dislocation creep ($n > 1$), see Karato (2012). Indeed, we show in Appendix 2.8 that in these cases one has to *always* choose $\alpha < 1$ if n exceeds a certain threshold.

2.3.6 Algorithms for the solution of the nonlinear problem

The discussions of the previous sections show that a naive application of Newton’s method may lead to matrices that are neither symmetric nor positive definite. Indeed, in some cases the equations for the Newton update may not be well-posed at all (see, for example, the discussion in Appendix 2.8), even if the original, nonlinear model has all of these properties.

The remedies outlined above restore symmetry, positive definiteness, and well-posedness, and consequently lend themselves for a practical implementation. On the other hand, the resulting equations for the update are different from the ones obtained by linearizing the residual, and consequently we may not be able to expect quadratic convergence of the resulting nonlinear iteration. Indeed, this is what we will observe in the experiments we show in Section 2.4. Regardless, the modifications have to be incorporated into an actual algorithm to solve the nonlinear problem. The algorithm we propose for this – which is also the one implemented in the ASPECT code (Kronbichler et al., 2012; Heister et al., 2017) – is therefore outlined below. As for many other

nonlinear problems, it is not easy to universally achieve convergence, and the resulting algorithm is therefore complicated.

Nonlinear iteration

As with many other nonlinear problems, it is not generally possible to solve the nonlinear Stokes equation we consider here using only a Newton iteration. Rather, we use a strategy where we use the following sequence to solve the nonlinear Stokes problem in each time step:

1. We always use one initial Picard step. That is, we solve the original Stokes equations in which we “freeze” all coefficient using values for the strain rate and pressure extrapolated from previous time steps; this corresponds to solving $Q(\tilde{X})X_1 = b(\tilde{X})$ (in analogy to (2.8)) where \tilde{X} is the extrapolated solution. This allows us, in particular, to enforce the correct boundary conditions on all boundaries where the velocity is prescribed.
2. We then solve $N_{DC} \geq 0$ steps using the Picard method written in Defect Correction (DC) form. This corresponds to equation (2.12) if one were to omit all terms that contain derivatives of η in the definition of the blocks in (2.14)–(2.16). Equivalently, this corresponds to solving an update form of (2.8), namely $Q(X_k)\delta X_k = b(X) - Q(X_k)X_k = -r_k$ followed by computing $X_{k+1} = X_k + \delta X_k$. It is well-known that the Picard iteration is more stable than a pure Newton method and often converges even in cases where Newton’s method does not. It therefore allows us to compute an iterate close enough to the exact solution from which we can then successfully start the Newton iteration. (For this second set of iterations, we use the defect correction form because the updates δX_k then have a zero velocity on all boundaries where the velocity is prescribed.)
3. We continue with full Newton steps, i.e., we attempt to solve the unmodified Newton equations stated in (2.12) with blocks defined as in (2.14)–(2.16). We know that these equations will eventually lead to quadratic convergence, but they may not be symmetric, positive definite, or even solvable. Consequently, the linear solvers we will discuss in the next subsection may fail to converge.
4. If the linear solver failed in one of the previous, unmodified Newton steps, we continue with Newton-like steps that modify the matrix blocks as shown in (2.14)–(2.16) by the methods of Sections 2.3.4 and 2.3.5. By construction, the resulting linear system is then guaranteed to be invertible, and indeed our linear solvers always succeed in our experiments.



These iterations are terminated once the nonlinear residual $\|\mathbf{r}_k\|$ has been reduced by a user-defined factor compared to the starting nonlinear residual at the beginning of each time step. We use a line search (see Kelly (1995)) to determine an acceptable step length for all Newton-type steps to further globalize convergence.

In addition to the outline above, we have tried a method suggested to us by Riad Hassani (personal communication, 2017) in which the switch-over between Picard defect-correction iteration as defined above in (ii) (corresponding to using a Newton matrix in which we have dropped all terms involving derivatives of the coefficients) and Newton iterations (i.e., the same blocks but *including* the derivative terms) is done gradually by scaling the derivatives in overall iteration k by a factor c_k between zero and one. We will in the rest of this paper refer to this as the Residual Scaling Method (RSM). The initial N_{DC} iterations can then be interpreted as using $c_k = 0$, after which we choose

$$c_k = \max\left(0.0, 1 - \frac{\|\mathbf{r}_k\|}{\|\mathbf{r}_{N_{DC}}\|}\right)$$

where \mathbf{r}_k is the current nonlinear residual and $\mathbf{r}_{N_{DC}}$ the residual in the first iteration after switching to the Newton or Newton-like method. This choice guarantees that $c_k \approx 1$ once Newton's method has reduced the residual significantly, i.e., once we are close to the solution.

This variation often allows us to choose N_{DC} smaller, i.e., to try a method with a faster convergence rate earlier in the process. On the other hand, it sometimes requires more Newton-type iterations. Using this variation leads to somewhat mixed improvements over the strategy outlined above, as will be shown in our numerical results below.

Linear solvers

Regardless of whether we solve the Picard or any of the Newton-type problems above, we always end up with having to solve a linear system with the same block structure as (2.13) in each nonlinear step. This problem may or may not be symmetric, and the top left block \mathbf{J}^{uu} may or may not be positive definite. However, regardless of these details, we use variations of the solvers discussed in Kronbichler et al. (2012) and Heister et al. (2017) to solve the linear problem.

More specifically, we use F-GMRES as the outer solver, with the following matrix as a preconditioner:

$$\mathbf{P}^{-1} = \begin{pmatrix} (\widetilde{\mathbf{J}^{uu}})^{-1} & (\widetilde{\mathbf{J}^{uu}})^{-1} \mathbf{J}^{up} \widetilde{\mathbf{S}}^{-1} \\ 0 & -\widetilde{\mathbf{S}}^{-1} \end{pmatrix}, \quad (2.19)$$

where a tilde indicates an approximation of the matrix under the tilde, and $\mathbf{S} = \mathbf{J}^{pu}(\mathbf{J}^{uu})^{-1}\mathbf{J}^{up}$ is the Schur complement of the system. Specifically, motivated by the discussions in Kronbichler et al. (2012) and Heister et al. (2017), we use the following approximations for each of these blocks:

- $(\widetilde{\mathbf{J}^{uu}})^{-1}$: We approximate this matrix using either one multigrid cycle or a full solve with an approximation $\widetilde{\mathbf{J}^{uu}}$ of \mathbf{J}^{uu} that is constructed in a similar way as discussed in Kronbichler et al. (2012). In addition, because both multigrid and the Conjugate Gradient method used here require $\widetilde{\mathbf{J}^{uu}}$ to be symmetric and positive definite, we always apply the modifications of Sections 2.3.4 and 2.3.5, even if they are not applied to \mathbf{J}^{uu} itself.
- $\widetilde{\mathbf{S}}^{-1}$: This block is an approximation to the inverse of the Schur complement $\mathbf{S} = \mathbf{J}^{pu}(\mathbf{J}^{uu})^{-1}\mathbf{J}^{up}$. Like for the original Stokes problem, the appropriate approximation is to use $\widetilde{\mathbf{S}}^{-1} = \mathbf{M}_p^{-1}$ where $(\mathbf{M}_p)_{ij} = (\varphi_i^p, \frac{1}{\eta(\epsilon(u))}\varphi_j^p)$ is the mass matrix on the pressure space scaled by the inverse of the viscosity; the inversion of \mathbf{M}_p is facilitated by a Conjugate Gradient solve.

The approximation $\widetilde{\mathbf{S}}^{-1} = \mathbf{M}_p^{-1}$ is known to be good if $\mathbf{J}^{pu} = (\mathbf{J}^{up})^\top$, see Silvester and Wathen (1994). On the other hand, this is not the case if the viscosity depends on the pressure, given the additional term in (2.15). However, the difference between the two matrices is small if the viscosity does not strongly depend on the pressure. This is, in fact, a commonly made assumption, at least for deep Earth mantle models, though it may not be valid for crustal models that employ pressure-dependent plasticity models.

It is conceivable that one can construct a better approximation $\widetilde{\mathbf{S}}^{-1}$ for \mathbf{S}^{-1} – leading to fewer outer F-GMRES iterations – by also incorporating the viscosity derivative terms somehow, but we did not pursue this direction as it is tangential to the purpose of this paper.

It is, in general, not necessary to solve the linear systems in the first few non-linear iterations with high accuracy. Rather, without significant loss of nonlinear solver performance, one can solve with a loose tolerance and terminate F-GMRES substantially earlier. Consequently, we have implemented both choice one and two of Eisenstat and Walker (1996) for stopping criteria for the linear solver, where for choice 2 we followed Kelly (1995) in using $\gamma = 0.9$ and $\alpha = 2$. For the definition of these symbols see the original paper. We noticed for some

of the problems that the difference between these two approaches where significant, where the first choice allowed for a much looser tolerance. Eisenstat and Walker (1996) stated that choice one represents a direct relation between the Newton right hand side F and its local linear model at the previous nonlinear iteration, while choice two is only an approximation of this. Therefore we have chosen the first of these approaches for this paper.

Computation of derivatives

Implementations of Newton solvers require concrete implementations of the formulas for the derivatives $\frac{\partial \eta(\boldsymbol{\varepsilon}(\mathbf{u}))}{\partial \boldsymbol{\varepsilon}}$ and $\frac{\partial \eta(\boldsymbol{\varepsilon}(\mathbf{u}))}{\partial p}$. These can be computed either using simple finite differencing approaches or analytically. Fortunately, even for relatively complicated material models, exact formulas for these derivatives can be derived with modest effort. Examples for the material models we consider in our numerical results below are provided in Appendix 2.8.

2.4 Numerical experiments using common benchmarks

In this section, let us illustrate the performance of the methods layed out above, using several benchmarks that vary both in which specific elements of the solver they test as well as in the difficulty they present to solvers. In particular, we will assess whether and how fast different variations of our algorithms converge. This includes ensuring that the nonlinear residual can be reduced to any small value desired. Furthermore, we will investigate optimal values and relative trade-offs for a variety of parameters that affect the nonlinear solver scheme, as discussed in Section 2.3.6.

The benchmarks we describe here have all been used for similar purposes in the literature. Details of all of our experiments are, sometimes in a simplified form, also part of the ASPECT test suite. All codes necessary to run these experiments are available among the benchmarks included in ASPECT releases starting from version 2.1. The ASPECT repository can be found at <https://github.com/geodynamics/aspect>.

2.4.1 Nonlinear channel flow

The simplest nonlinear Stokes flow one can think of is probably a generalization of incompressible Poiseuille flow to include a strain-rate dependent viscosity. In it, one forces a fluid through a pipe or channel where the velocity is zero at the pipe sides and in- and outflow velocities are prescribed in such a way that the result is a flow field parallel to the pipe axis and constant in the along-pipe direction. The across-pipe variation of the velocity field can then

be computed easily once a rheology law is chosen, leading to an analytically known flow field from which the in- and outflow boundary conditions can also be drawn either via prescribed velocities or prescribed tractions.

A visualization of the solution can be found, for example, in Gerya (2010); Turcotte and Schubert (2002).

Setup. We use the two-dimensional benchmark setup of Gerya (2010, Section 16.4). In it, the viscosity is chosen in accordance with a power law approach as

$$\begin{aligned}
 \eta(\boldsymbol{\varepsilon}(\mathbf{u})) &= C^{-\frac{1}{n}} [2 \mathbb{I}_2(\boldsymbol{\varepsilon}(\mathbf{u}))]^{\frac{1}{n}-1} \\
 &= C^{-\frac{1}{n}} \left[2 \sqrt{\frac{1}{2} \boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{u})} \right]^{\frac{1}{n}-1} \\
 &= \left[\sqrt{2} \right]^{\frac{1}{n}-1} C^{-\frac{1}{n}} \|\boldsymbol{\varepsilon}(\mathbf{u})\|^{\frac{1}{n}-1},
 \end{aligned} \tag{2.20}$$

using the definition of the second invariant found in Gerya (2010, p. 59, equation (4.14)). Here, C is a prefactor, and n is a stress exponent that allows for easy tuning of the nonlinearity of the problem. The model geometry we use here is a box of 10,000 m by 8,000 m, subdivided into 16×16 cells; we use quadratic finite elements for the velocity.

Results. Figures 2.1 and 2.2 show results for a number of methods and settings when the in- and outflow boundary conditions are either prescribed through tractions or velocity values. The latter turns out to generally be a more difficult problem to solve, but all methods eventually converge to a residual whose size is related to the tolerance with which we solve the linear systems.

Figures 2.1 shows that for this problem, when boundary values are given as tractions, line search is neither necessary nor useful, and similarly it is not necessary to run many initial Picard iterations to get close enough to the solution for the Newton method to start working. In addition, the Newton matrix modifications of Section 2.3.5 (right two panels of Fig. 2.1) actually destroy the quadratic convergence rate of Newton’s method and result in only linear convergence as speculated at the beginning of Section 2.3.6 – though with a substantially better linear rate than Picard iterations.

On the other hand, Fig. 2.2 shows that for the more complicated problem when the flow is driven by prescribed velocity boundary conditions, either a line search method or sufficiently many initial Picard iterations are necessary to achieve convergence. Alternatively, the matrix modifications also yield a

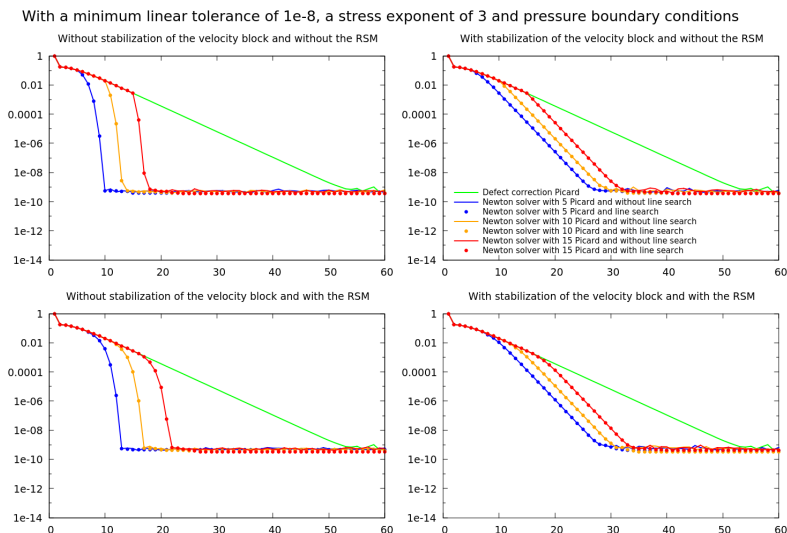


Figure 2.1: Nonlinear channel flow benchmark: Convergence history for several methods for a rheology with $n = 3$ where in- and outflow are described by prescribing the traction. Top row: Computations in which we switch abruptly from Picard iterations to Newton iterations. Bottom row: With the Residual Scaling Method (RSM) in which we switch continuously between the Picard iteration and the Newton method. Left column: Unmodified Newton iterations. Right column: Results where we applied the modifications of Sections 2.3.4 and 2.3.5 to the Newton matrix. Horizontal axes: Number of the nonlinear (outer) iteration. Vertical axes: Nonlinear residual.

convergent scheme. The Residual Scaling Method (RSM) in conjunction with the matrix modifications appears the most robust method, though not always the fastest. Indeed, as explained in Appendix 2.8.1, a stress exponent of $n = 3$ causes the matrix modifications to *always* scale down the derivative terms in the Newton matrix, resulting in a similar effect as the RSM.

In all cases, a pure Picard iteration always converges linearly, though at a rate that is not competitive with well-designed Newton iterations.

2.4.2 Spiegelman et al. benchmark

The Spiegelman et al. benchmark (see Spiegelman et al. (2016)) is an extended form of the brick benchmark of Lemiale et al. (2008) and focuses on solving for the behavior of a material with plastic rheology under compression.

Setup. The benchmark specifies two layers, see Fig. 2.3. The lower layer, which includes a regularized weak seed, has a constant viscosity. The upper

2.4. Numerical experiments using common benchmarks

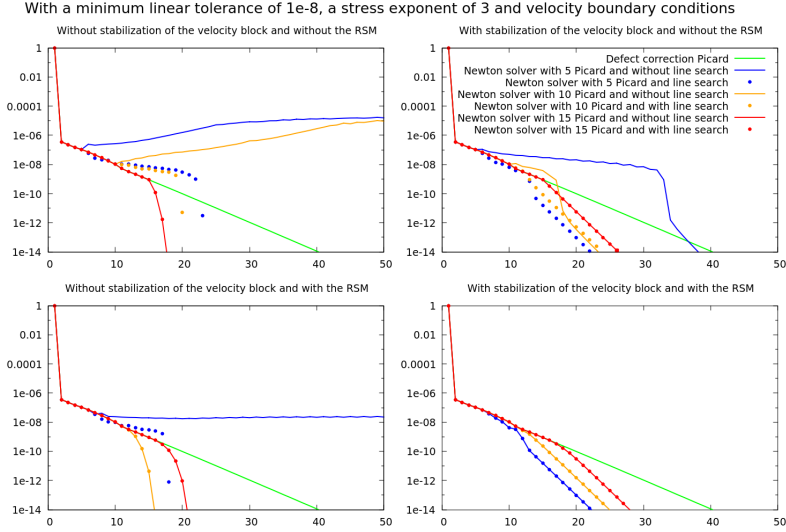


Figure 2.2: Nonlinear channel flow benchmark: Convergence history for several methods for a rheology with $n = 3$ where in- and outflow are described by prescribing the velocity. Panels as in Fig. 2.1.

layer has a viscosity given by the harmonic mean $\eta_{\text{eff}} = \frac{\eta_1 \eta_p}{\eta_1 + \eta_p}$. Here, η_1 is the background viscosity of the upper layer, and

$$\eta_p = \frac{A + B(p_{\text{lith}} + \alpha p')}{2 \|\varepsilon(\mathbf{u})\|},$$

where p_{lith} is the depth dependent lithostatic pressure and $p' = p - p_{\text{lith}}$ is the dynamic component of the total pressure. η_1 can have three different values: $1e23$, $1e24$ and $5e24$ Pa s. For von Mises plasticity, one would choose $A = C, B = 0$ where C is the cohesion of the material; in this case, the viscosity is strain-rate but not pressure dependent. For a depth-dependent von Mises-type model, one would choose $A = C \cos(\phi), B = \sin(\phi), \alpha = 0$ where ϕ is the friction angle; in this case, the viscosity depends on the static, lithospheric pressure but not the dynamic pressure component. Finally, Drucker-Prager plasticity fits this formula with $A = C \cos(\phi), B = \sin(\phi), \alpha = 1$ where the viscosity now depends on both the strain rate and the (total) pressure $p = p_{\text{lith}} + p'$. We will only consider the von Mises and Drucker-Prager cases of the Spiegelman benchmark because these are the most interesting ones.

The benchmark is completed by prescribing an inbound velocity (2.5, 5 and 12.5 mm/yr) on the two sides of the geometry, requiring tangential flow at the

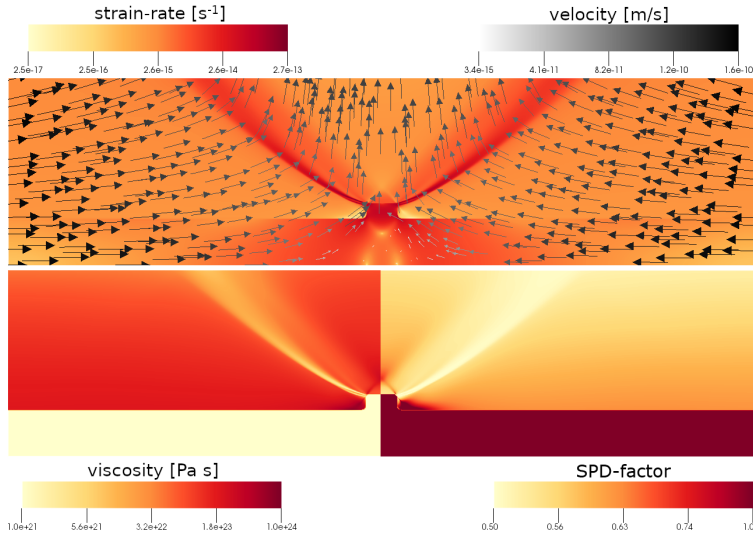


Figure 2.3: Spiegelman et al. benchmark: Depiction of the strain rate and velocity field (top), viscosity (bottom left) and the SPD factor α (bottom right) as a result of the deformation induced by the prescribed horizontal velocity on the sides of the domain. The results shown here are for the case where the velocities are 5 mm/year (the color bar shows it in m/s, $\eta_{\text{ref}} = 10^{24}$ Pa s, the angle of internal friction is 30° and the mesh consists of 1024×256 cells. This is one of the more difficult cases (Drucker-Prager) of the benchmark, and the data shown here are the result of simulations that are only converged to a relative nonlinear residual of about 10^{-6} .

bottom, and no stress at the top, allowing material to leave the domain. The three options for the inbound velocity and the three options for the reference background viscosity together form a set of nine test cases whose difficulty increases with velocity and reference background viscosity. In the original paper, an unstructured grid of stable Taylor-Hood elements was used. Based on their Fig. 1, the results shown there should, based on the length of the edges, correspond to a uniform ASPECT mesh that has been refined globally approximately 7 or 8 times (i.e., 512×128 or 1024×256 cells). This does not, however, account for details of the unstructured mesh used in Spiegelman et al. (2016).

Results. We compare the results of our Newton implementation to the results of Spiegelman et al. (2016) and the pre-existing Picard solver in ASPECT. The von Mises case results are quite similar to the results from Spiegelman et al. (2016); consequently, we will focus on the more difficult Drucker-Prager case. We have exhaustively explored the space of parameters affecting the nonlinear solver (see Section 2.3.6) at different mesh resolutions. In general, we found

2.4. Numerical experiments using common benchmarks

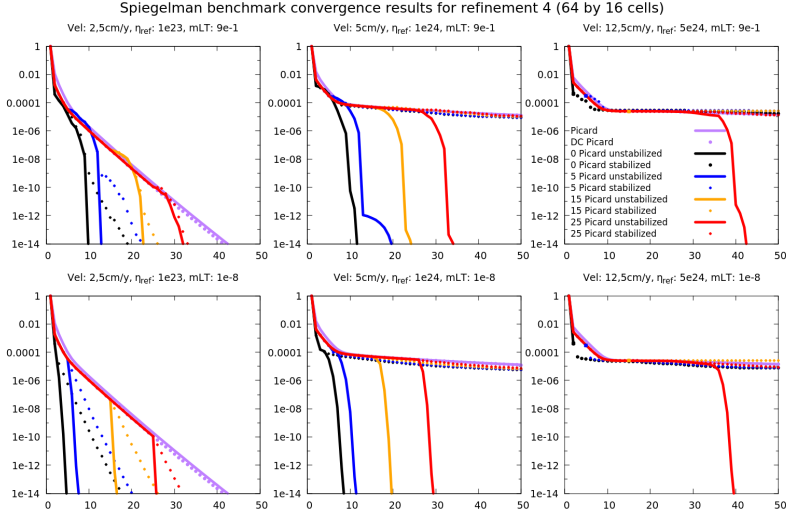


Figure 2.4: Spiegelman et al. benchmark: A reproduction of three of the nine pressure dependent Drucker-Prager cases with a resolution of 64×16 cells (substantially coarser than the resolution in the Spiegelman et al. (2016)). Top: Results for computations where linear systems are solved with a relative tolerance of 0.9. Bottom: With a tolerance of 10^{-8} . The initial Picard iteration is always solved to a linear tolerance of 10^{-16} . Left to right: Different prescribed velocities of $u_0 = 2.5, 5,$ and 12.5 cm/year and different reference viscosities of respectively $\eta_{\text{ref}} = 10^{23}, 10^{24}$ and 5×10^{24} Pa s. Horizontal axis: Number of the nonlinear (outer) iteration; vertical axis: Nonlinear residual. DC Picard refers to a Defect Correction Picard iteration, see Section 2.3.6.

that higher mesh resolution (more refinement steps) made the problem more difficult to solve. As expected, the benchmarks also become more difficult as the prescribed velocity u_0 at the boundary is increased, leading to a larger strain rate and more pronounced nonlinearity. This is visible in Fig. 2.3 where we also show the viscosity and the α factor necessary to keep the matrix positive definite. This factor drops to approximately one half in the vicinity of the shear band – a result consistent with the theoretical considerations discussed in Appendix 2.8.3.

Figures 2.4 and 2.5 show results obtained for 4 and 8 global mesh refinement steps, corresponding to meshes with 64×16 and 1024×256 cells, respectively. A comparison shows that the problem is indeed more difficult to solve on the finer meshes. Without enforcing the symmetry and positive definiteness of the top left block of the Jacobian, the linear solver converges quickly, but also crashes easily in a number of configurations because the matrix lacks the necessary structural properties; this is particularly the case for $u_0 = 12.5$ mm/year.

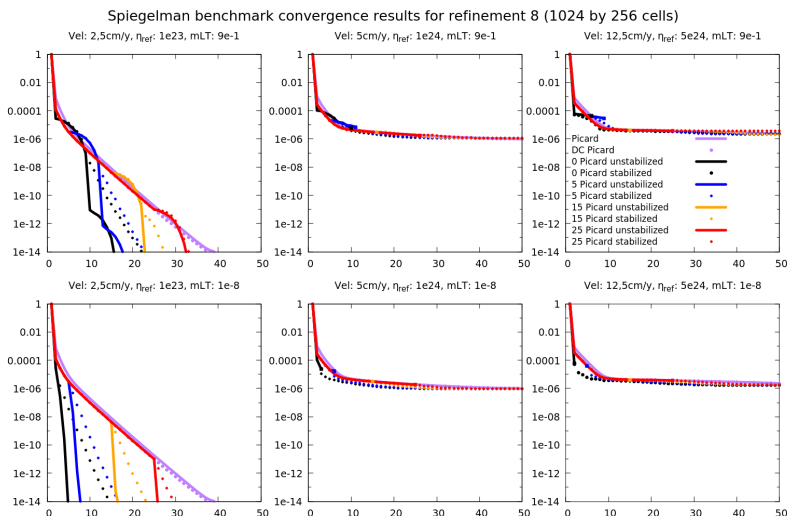


Figure 2.5: Spiegelman et al. benchmark: A reproduction of three of the nine pressure dependent Drucker-Prager cases with a resolution of 1024×256 cells (similar to the resolution in the Spiegelman et al. (2016)). Panels as in Fig. 2.4.

On the other hand, enforcing these properties on the matrix leads to some loss in speed of convergence of the nonlinear iterations (because the computed search direction is no longer second order accurate), though we then also no longer encountered any linear solver failures.

We noticed that the runs without the matrix modifications are very sensitive to changes in many of the solver parameters. In the following, let us discuss a number of settings that we have found useful when working with the unmodified matrix, though for many cases the linear solver still fails with these settings. In particular, using a few line search iterations may help reduce the amount of iterations; however, allowing the step length parameter to decrease too much generally leads to too small steps and slow overall convergence. A good default value for the maximum number of line search step length reductions appears to be 5 or 10. We also found that performing 5–10 Picard iterations before switching to Newton iterations is a good number. Setting the relative linear tolerance of the GMRES solver rather strict typically results in fewer outer iterations, but at the price of more inner iterations that then increase the run time per outer iteration. A good compromise for this parameter is to require that the linear residual is reduced to a factor of 0.1 or 0.01 of its initial value. Using the RSM, i.e., determining the linear solver tolerance automatically, usually requires one

or two more outer iterations, but greatly decreases the chance that the linear solver will fail.

On the contrary, enforcing the \mathbf{J}^{uu} to be symmetric and positive definite will yield a very different behavior. As stated above, the linear solver then *always* converges. Furthermore, the sensitivity of the convergence history to all the parameters above is greatly reduced. This has the advantage that a very loose linear tolerance can be used without a significant penalty in the number of iterations; that said, and as is apparent from the figures, modifying the matrix may require tens of nonlinear iterations more to converge.

We have obtained the best performance by combining the two methods: We do not enforce the symmetry and positive definiteness of the matrix in the first nonlinear iterations until (if) the linear solver fails; after this, we continue with the matrix modifications. This strategy combines the fast convergence of the unmodified method with the stability and robustness of the modified one.

2.4.3 Tosi et al. benchmark

The Tosi benchmark of Tosi et al. (2015) is designed as a community benchmark for mantle flow based on nonlinear rheologies featuring a temperature, pressure, and strain rate-dependent viscosity. Here, we specifically consider case 4 from Tosi et al. (2015), which seeks the steady state (at a large, unspecified end time) of a time dependent problem. Unlike the two previous benchmarks, this benchmark has a temperature field that is coupled to the viscosity and therefore evolves over time.

Setup. The benchmark is posed in a two-dimensional square unit box with all free slip boundaries and an initial temperature given by $T(x, z) = (1 - z) + A \cos(\pi x) \sin(\pi z)$. The viscosity is chosen as the harmonic mean

$$\eta(T, z, \varepsilon(\mathbf{u})) = 2 \left(\frac{1}{\eta_{\text{lin}}(T, z)} + \frac{1}{\eta_{\text{plast}}(\varepsilon(\mathbf{u}))} \right)^{-1}, \quad (2.21)$$

where the two components of the viscosity are defined as a linear but depth- and temperature-dependent viscosity as well as a plastic yield criterion respectively:

$$\eta_{\text{lin}}(T, z) = e^{-\gamma_T T + \gamma_z z}, \quad \eta_{\text{plast}}(\varepsilon(\mathbf{u})) = \eta^* + \frac{\sigma_Y}{\|\varepsilon(\mathbf{u})\|}.$$

Here, η^* is the constant effective viscosity at high strain rate, and σ_Y the yield stress. Numeric values for all of these constants can be found in the original paper.

Results. The original paper does not contain convergence plots. Consequently, we can only compare between the methods available in our reference implementation. Specifically, these are: (i) A method whereby we solve the advection equation, then the Stokes equation with frozen coefficients, and then iterate these two steps out until we have reached convergence for the current time step; we will refer to this scheme as “Picard” even though this stretches the term (as, strictly speaking, a “Picard” iteration for the coupled system would solve both the Stokes and advection problem linearized around the previous solution; in our implementation, the Stokes system is linearized around the already computed advection solution). In ASPECT, this scheme is called “iterated advection and Stokes”. (ii) A method where we first solve the advection equation and then do one Newton step on the nonlinear Stokes system; again, these two parts are iterated out in each time step. We will refer to this method as “Newton”; in ASPECT, it is called “iterated advection and Newton Stokes”.

Figure . 2.6 shows results for this benchmark, where the horizontal axis indicates the number of the nonlinear iteration performed. Each spike corresponds to a time step starting at a large residual that is gradually decreased. A method that converges quickly shows a steeper decrease, requires fewer nonlinear iterations, and can consequently fit more time steps (spikes) into the same number of nonlinear iterations. Because the computational effort is largely confined to building and solving the linear systems, the horizontal axis also corresponds closely to the elapsed wall time.

The different panels of the figure can be summarized as follows: (i) The Newton method converges much faster than Picard iterations. For example, after the initial few time steps, the Newton method (with and without matrix stabilization) only requires two nonlinear iterations per time step, whereas the Picard iteration requires six. This also translates to a speed up in wall time of around the same factor. (ii) Matrix stabilization is not necessary for this benchmark and in fact leads to a slight but not substantial degradation of performance. (iii) Convergence behavior can differ substantially between timesteps (both for Picard and Newton). (iv) For this experiment, the defect correction form of the Picard iteration is slightly, but not substantially faster than the original Picard iteration.

2.4.4 A 3d subduction test case

In order to verify that our implementation is not only applicable to academic benchmarks, but also to settings that occur in geodynamic modeling, let us consider a three-dimensional simulation of oceanic plate subduction. The

2.4. Numerical experiments using common benchmarks

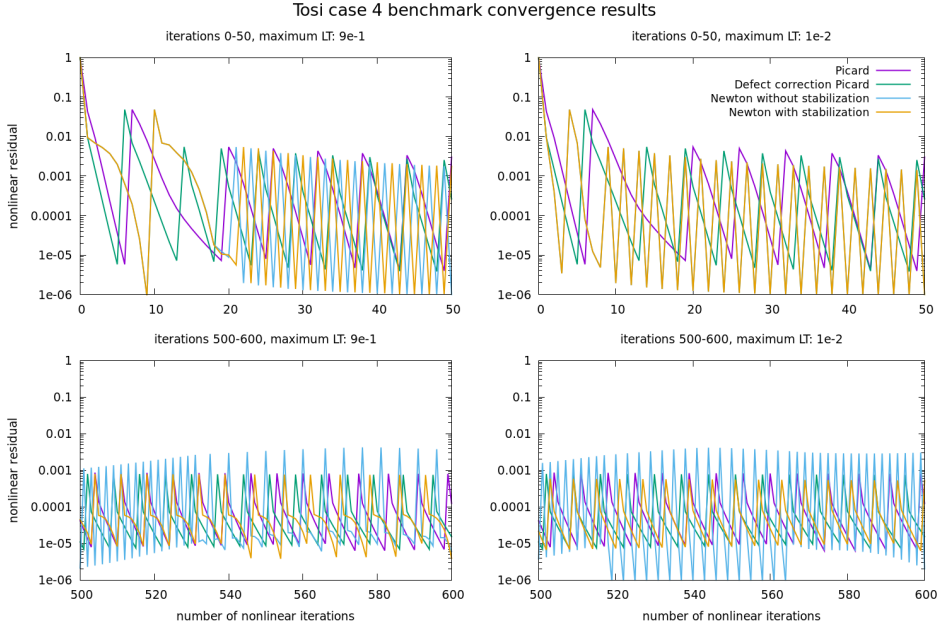


Figure 2.6: Tosi et al. benchmark: Nonlinear Stokes residual (vertical axis) as a function of the number of the nonlinear solves. Each spike corresponds to one time step during which iterations start with a large residual that is gradually decreased until it reaches the desired nonlinear tolerance of 10^{-5} . No line search is used here. Top: The first 50 nonlinear iterations. Bottom: Nonlinear iteration 500–600. Left: Linear systems are solved to a relative Linear Tolerance (LT) of 0.9. Right: With a linear tolerance of 0.01. In the top right panel, the results for the two Newton variations coincide.

model is inspired by the geodynamic setting of the Caribbean region, and simulates a slab that subducts while the subducting plate has a motion oblique to the trench which causes the slab to be dragged laterally through the mantle, a motion called slab dragging (Spakman et al., 2018).

Setup. We situate our test case in a Cartesian box with dimensions of 3000 km in width and length and 1000 km in depth (see Fig. 2.7). The viscoplastic rheology includes dislocation creep, diffusion creep, and plasticity. The viscosity is then given by

$$\eta_{\text{eff}} = \max \left(\min \left\{ \left(\frac{1}{\eta_{\text{diff}}} + \frac{1}{\eta_{\text{disl}}} \right)^{-1}, \eta_{\text{plastic}}, \eta_{\text{max}} \right\}, \eta_{\text{min}} \right) \quad (2.22)$$

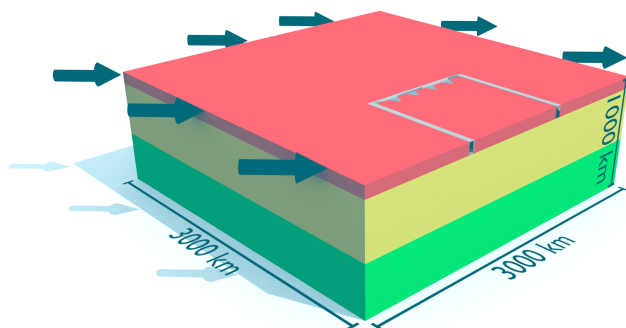


Figure 2.7: A conceptual visualization of the 3d subduction test case setup. The red layer represents the lithosphere, the yellow layer the upper mantle, and the green layer the lower mantle. The arrows indicate the lithospheric boundary velocity direction.

Time: 3.55256e6 yr

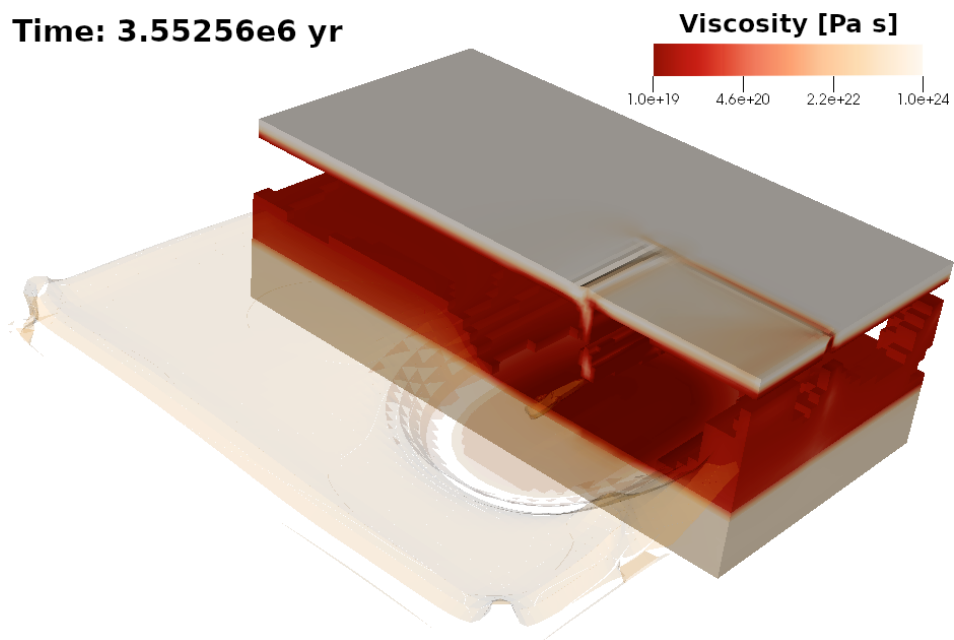


Figure 2.8: Right: The viscosity field of the 3d subduction test case after $3.55 \cdot 10^6$ years, just as the tip of the subducting slab starts to detach from the rest of the slab. In the left half, two isosurfaces show where in the bottom half of the model the viscosity equals $3.25 \cdot 10^{19}$ and $5.5 \cdot 10^{19}$ Pa s. In the right half, colors indicate the viscosity when restricted to areas where the viscosity is greater than $5 \cdot 10^{19}$ Pa s, i.e., to cold areas of the upper mantle, as well as the lower mantle.

where

$$\eta_x = \frac{1}{2} v_x A_x^{-\frac{1}{n_x}} \left[\frac{1}{\sqrt{2}} \|\varepsilon(\mathbf{u})\| \right]^{\frac{1}{n_x} - 1} \exp\left(\frac{E_x + P V_x}{n_x R T}\right),$$

$$\eta_{\text{plastic}} = \frac{6(C \cos(\phi) + P \sin(\phi))}{\sqrt{3}(3 - \sin(\phi))} \frac{1}{\sqrt{2} \|\varepsilon(\mathbf{u})\|}.$$

Here, a symbol of the form \square_x stands for the corresponding property of either diffusion (if $x = \text{diff}$) or dislocation creep (if $x = \text{disl}$), v is a constant factor which can be used to scale the rheology, R is the gas constant, T is temperature, and P is pressure. A_x are prefactors, E_x are the activation energies, and V_x are the activation volumes. The values of all of these parameters used for the simulations are listed in Appendix 2.9.

The model consists of two layers, the upper and lower mantle (above and below 660 km, see figure 2.7), that differ only in strength through a 100-fold increase in η_{diff} and η_{disl} by choosing v_{diff} and v_{disl} larger by a factor of 100 in the lower mantle. This means that the lithosphere is fully defined by temperature. This thermal lithosphere is divided into two regions: a U-shaped region representing an oceanic plate that surrounds a region representing a Large Igneous Province (LIP), both modeled by a plate (Fowler, 2005) of thickness 95 km and the ridge far outside the domain. The slab is also 95 km thick and is divided into three segments in which an analytic temperature field is prescribed following McKenzie (1970). The first segment is 200 km long with a dip angle relative to the surface starting with 20° that smoothly steepens to a dip angle of 30°. The second segment is 150 km long and the dip angle has at the end of the segment smoothly increased to 70°. The third, straight segment is 50 km long and has a constant dip angle of 70°. We describe the fault zones between the oceanic plate and the LIP as thin, vertical regions with an elevated initial temperature. The U-shaped lithosphere has a prescribed boundary velocity of 1 cm/year in each component of the horizontal directions (for direction see the arrows on Fig. 2.7), and zero velocity in the vertical direction. The LIP has a prescribed boundary velocity of zero in all directions. Below the lithosphere we use open boundary conditions. The top is a free surface (Rose et al., 2017) and the bottom has a zero velocity boundary condition.

This model is discretized on a mesh that has a total of 153,046 cells, resulting in 3,938,115 velocity and 172,097 pressure unknowns (in addition to another 1,312,705 unknowns each for the temperature and a compositional field). All results shown below were obtained on the Dutch national cluster Cartesius. Each node is equipped with Intel® Xeon E5-2690 v3 (“Haswell”) processors and

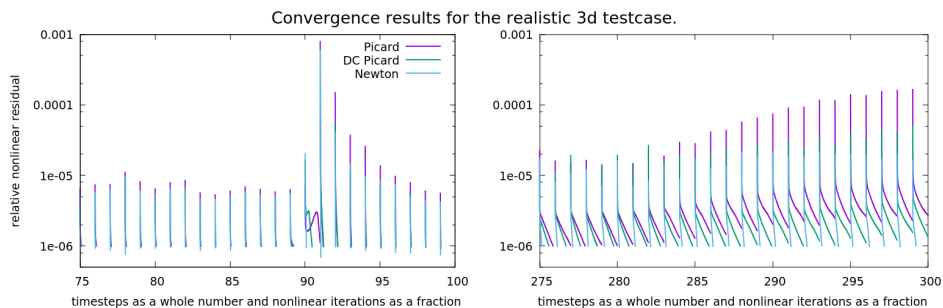


Figure 2.9: The 3d subduction test case: Nonlinear convergence results for the Picard, Defect Correction (DC) Picard, and Newton iterations. The horizontal axis shows time steps, with nonlinear iterations depicted at $\frac{1}{250} = 0.004$ increments given that we allow at most 250 nonlinear iterations per time step. The vertical axis represents the nonlinear relative residual.

has 24 cores; we used 10 nodes and 20 MPI processes per node. The model is run with a Courant-Friedrichs-Lewy (CFL) number of 0.1 and the time step size is limited to grow by a maximum of 25% from one time step to the next. The time step sizes computed by all nonlinear solver methods used below are essentially identical.

The experiments in previous sections show that it is in general not necessary to solve the linear systems in defect correction schemes (i.e., the defect correct version of Picard iterations, as well as Newton iterations) particularly accurately. As a consequence, we only use a linear solver tolerance for these methods that requires a reduction of the linear residual in the F-GMRES solve by a factor of 0.1. On the other hand, the initial Picard iteration solves for the solution, not an update, and consequently requires a substantially larger reduction of the linear residual; we use a factor of 10^{-6} (the default value of ASPECT).

Results. We have performed this experiment using our implementation of the Newton method, as well as the Picard iteration and its defect correction (DC) variation. In the following, let us provide two perspectives on this comparison.

First, Fig. 2.9 shows how all three methods reduce the nonlinear residual in each time step, for two selected periods of the simulation (from time steps 75 to 100, and 275 to 300). In the early phases of the simulation, all methods quickly converge the nonlinear residual to the desired tolerance of 10^{-6} , though even here, the Newton method requires fewer iterations. Interestingly, starting around time step 275 – corresponding to about 3.92 million years of model time

2.4. Numerical experiments using common benchmarks

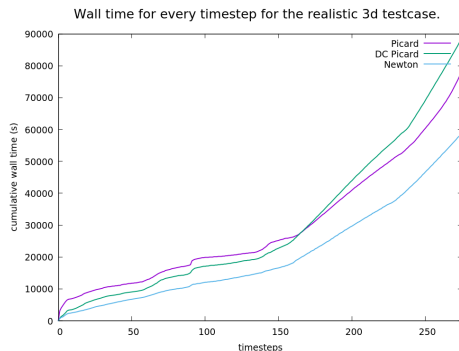


Figure 2.10: The 3d subduction test case: Wall clock time necessary to compute up to a certain time step for the Picard, Defect Correction (DC) Picard, and Newton iterations.

– the problem appears to become substantially more difficult to solve. This time corresponds to the break-off of the deeper part of the slab (the necking in Fig. 2.8), which then rapidly sinks, and the increased velocity implies a larger strain rate and consequently stronger nonlinearity. Indeed, as Fig. 2.9 shows, both variations of the Picard iteration are then no longer able to converge the residual to the desired tolerance, even though we allow up to 250 iterations per time step. Any results of these simulations must then necessarily be suspicious. On the other hand, the Newton iteration continues to rapidly converge. These data therefore underline the stability and robustness of the Newton method, and illustrate that it can solve problems that are otherwise not solvable with reasonable effort.

A second perspective is shown in Fig. 2.10, where we plot the wall time necessary to solve the problem up to a given time step. The figure demonstrates that our implementation of Newton’s method is approximately one third faster than the other two methods for the time steps shown. The difference becomes particularly notable after time step 170, where the subducting slab starts to thin under its own gravity. We did not include data beyond time step 275 because, as discussed above, the Picard variants do not converge any more after this point. Consequently, even though the curves would substantially diverge after this timestep due to the far larger number of nonlinear iterations taken by the Picard variants, the solutions would no longer be comparable.

These models mostly use ASPECT’s default parameter values. It is possible that we could further optimize any of the methods shown by changing these values. Furthermore, we have also so far not put much effort into optimizing the efficiency of the Newton solver code itself. That said, both of these issues



are outside of the scope of the paper.

2.5 Conclusions

Newton's method is generally considered the best method to solve nonlinear systems, but it is also known to be difficult to make work in practice. In this contribution, we have demonstrated that a naive application of Newton linearization may lead to a system for the updates δX_k that may be ill-posed even if the original nonlinear problem is well-posed. We have shown how one can modify the Newton system to guarantee a stable solution of the update equations. Specifically, we have modified the partial differential operators that give rise to the Newton matrix so that the matrix is symmetric and positive definite. Importantly, however, we do not modify the right hand side of the Newton update equation, and consequently the iterates of the modified Newton method still converge to the solution of the original nonlinear problem. We have also described the globalization strategies necessary to guarantee that the Newton method actually converges in nearly all cases, and demonstrated our methods using standard geodynamic benchmarks and a real application.

The results we have shown demonstrate that with this combination of methods, Newton's method (with globalization approaches and stabilization of the matrix) really is the better choice: It converges more rapidly, in fewer iterations, is robust, and takes less computational time. Furthermore, it can be applied to problems that are known to be very nonlinear and difficult to solve, such as the Spiegelman et al. (2016) benchmark. Finally, we have applied this method to a rheologically and geodynamically complex, three-dimensional case of oceanic subduction that is sufficiently nonlinear that the typical Picard iteration no longer converges in a reasonable number of steps. The Newton method we have discussed here not only converges, but does so in relatively few steps and with substantial savings in wall clock time.

There are, of course, cases where a simple Picard solver would have been sufficient. In those cases, however, it is worth pointing out that using a Newton method is not substantially more expensive than a Picard method: The assembly of the Newton matrix is marginally more complicated because of the terms involving the derivatives of the viscosity, but other than that the solution procedure is the same between the two methods because of the structurally very similar matrices involved. Consequently, it is reasonable to advocate for *always* using a Newton method instead of Picard iterations.

2.6 Acknowledgments

The authors greatly appreciate the support by all developers of the ASPECT code, and in particular by Juliane Dannberg, Rene Gassm oller, and Timo Heister. WB would like to thank Denis Davydov, Andrew McBride, and Jean-Paul Pelteret for helpful discussions about issues such as those discussed in Sections 2.3.4 and 2.3.5 in the context of strain-weakening solid mechanics.

This work is funded by the Netherlands Organization for Scientific Research (NWO), as part of the Caribbean Research program, grant 858.14.070, as well as by the NWO project ‘Large scale finite element models of the Caribbean region: Newton vs. Picard nonlinear iterations’ with project number 15820.

We acknowledge computational support by the Netherlands Research Centre for Integrated Solid Earth Science (ISES).

WB’s work was partially supported by the Computational Infrastructure in Geodynamics initiative (CIG), through the National Science Foundation under Award No. EAR-0949446 and The University of California – Davis; and by the National Science Foundation under awards OCI-1148116 and OAC-1835673 as part of the Software Infrastructure for Sustained Innovation (SI2) program (now the Cyberinfrastructure for Sustained Scientific Innovation, CSSI).

WS acknowledges support from the Research Council of Norway through its Centres of Excellence funding scheme, project number 223272.

2.7 Appendix A: The connection between elliptic operators, well-posedness of the Newton update equation, and eigenvalues of coefficients

The discussion in Section 2.3.5 made use of the fact that a positive definite coefficient H^{spd} in the definition of the J^{uu} block of the matrix implies that the underlying differential operator is elliptic, and that consequently the J^{uu} block is invertible. Since this connection is not obvious unless one works daily with partial differential equations, let us discuss this step in slightly more detail in this appendix.

To this end, let us first note that we call an operator \mathcal{A} acting on functions $u, v \in V$ (where V is a function space) bounded and coercive if there are constants $c > 0, C < \infty$ so that the following two conditions are satisfied:

$$\begin{aligned} \text{boundedness:} & \quad \langle u, \mathcal{A}v \rangle \leq C \|u\|_V \|v\|_V & \quad \forall u, v \in V, \\ \text{coercivity:} & \quad \langle u, \mathcal{A}u \rangle \geq c \|u\|_V^2 & \quad \forall u \in V. \end{aligned}$$

Here, $\langle \cdot, \cdot \rangle$ is the inner (or duality) product in V . It is well known from the theory of partial differential equations that such operators lead to unique solutions of the equation $\mathcal{A}u = f$, see Brenner and Scott (2002).

In the context of second order partial differential operator – such as the one that governs the top-left block of the Stokes problem –, an operator of the form $\mathcal{A}\mathbf{u} = -\text{div}[H\boldsymbol{\varepsilon}(\mathbf{u})]$ acting on a velocity field $\mathbf{u} \in V = H_0^1(\Omega)^d$ is said to be elliptic if there exists $c_1 > 0$ so that

$$\tau : (H\tau) \geq c_1 \|\tau\|^2 \tag{2.23}$$

for all symmetric tensors τ . Here, E is the rank-4 tensor that maps strain rates to stresses. For such operators, we have by integration by parts that $\langle \mathbf{u}, \mathcal{A}\mathbf{v} \rangle = (\boldsymbol{\varepsilon}(\mathbf{u}), H\boldsymbol{\varepsilon}(\mathbf{v})) = \int_{\Omega} \boldsymbol{\varepsilon}(\mathbf{u}) : [H\boldsymbol{\varepsilon}(\mathbf{v})]$. It is then easy to see that ellipticity implies coercivity by recognizing that $\|\mathbf{v}\|_V = \|\mathbf{v}\|_{H_0^1(\Omega)^d} = \left(\int_{\Omega} |\mathbf{v}|^2 + \int_{\Omega} |\nabla \mathbf{v}|^2 \right)^{1/2}$ and knowing that $c_2 \|\mathbf{v}\|_V \leq \|\boldsymbol{\varepsilon}(\mathbf{v})\| \leq c_3 \|\mathbf{v}\|_V$ for some constants $c_2 > 0$, $C_3 < \infty$.

In other words, if the coefficient H inside the differential operator satisfies condition (2.23), then the associated operator is well posed and invertible. This condition carries over to the case where we let V be a finite dimensional subspace of $H_0^1(\Omega)^d$ – for example, the set of all finite element functions defined on a mesh.

The important realization is now that the constant c in (2.23) equals the smallest eigenvalue of H where we consider H as an operator that maps a symmetric tensor to a symmetric tensor. To see this, assume that H had a negative or zero eigenvalue λ . Then, we can choose σ as the corresponding eigenvector and obtain that $\sigma : (H\sigma) = \sigma : (\lambda\sigma) = \lambda \|\sigma\|^2 \leq 0$, in violation of (2.23). Consequently, (2.23) can only be satisfied if all eigenvalues of H are positive.

This argument proves that if all eigenvalues of H are positive, then $\mathcal{A}\mathbf{u} = -\text{div}[H\boldsymbol{\varepsilon}(\mathbf{u})]$ is elliptic and consequently coercive, and as a result the top left block \mathbf{J}^{uu} of the matrix is positive definite and invertible.

It may be of interest to note that the operator \mathcal{A} may be invertible even if it is not elliptic, i.e., it is only *sufficient* but not *necessary* that H only have positive eigenvalues. However, it is much more difficult to specify the exact conditions that have to hold for H to ensure that \mathcal{A} is invertible, and we will not attempt to do so here. This is particularly true if $H = H(\mathbf{x})$ is spatially variable with eigenvalues that may also be different from one location to another.

2.8 Appendix B: A look at some common rheologies

The results of Section 2.3.4 and 2.3.5 were obtained for general rheologies in which the viscosity is a function of the strain rate $\boldsymbol{\varepsilon}(\mathbf{u})$ and possibly the pressure

p . However, if we know the specific form of this dependence, we can say more about whether or not it is necessary to symmetrize the matrix, and/or whether it is necessary to use a scaling factor α that is less than one. In the following, we will consider some common rheologies from this perspective.

2.8.1 Power law rheology

Some of the simplest rheology are of power law type where the viscosity is defined as

$$\eta(\varepsilon(\mathbf{u})) = \eta_0^{-\frac{1}{n}} \left(\sqrt{\frac{1}{2} \varepsilon(\mathbf{u}) : \varepsilon(\mathbf{u})} \right)^{\frac{1}{n}-1} = \eta_0^{-\frac{1}{n}} 2^{\frac{1}{2}-\frac{1}{2n}} (\|\varepsilon(\mathbf{u})\|^2)^{\frac{1}{2n}-\frac{1}{2}},$$

with $n \geq 1$. The form on the right shows that the viscosity is a function of the square of the norm of the strain rate. This implies that the matrix is automatically symmetric and does not have to be explicitly symmetrized with the procedure of Section 2.3.4.

Furthermore, we can compute the derivative of the viscosity with respect to the strain rate and obtain

$$\frac{\partial \eta(\varepsilon(\mathbf{u}))}{\partial \varepsilon} = \eta_0^{-\frac{1}{n}} 2^{\frac{1}{2}-\frac{1}{2n}} \left(\frac{1}{2n} - \frac{1}{2} \right) (\|\varepsilon(\mathbf{u})\|^2)^{\frac{1}{2n}-\frac{1}{2}-1} 2\varepsilon(\mathbf{u}) = \eta(\varepsilon(\mathbf{u})) \left(\frac{1}{n} - 1 \right) \frac{\varepsilon(\mathbf{u})}{\|\varepsilon(\mathbf{u})\|^2}.$$

Identifying $a = \varepsilon(\mathbf{u})$ and $b = \frac{\partial \eta(\varepsilon(\mathbf{u}))}{\partial \varepsilon}$ as in Section 2.3.5, we see that the important term in the scaling factor definition (2.18) is

$$\left[1 - \frac{b : a}{\|a\| \|b\|} \right]^2 \|a\| \|b\| = [1 - (-1)]^2 \left| \frac{1}{n} - 1 \right| \eta = 4 \left(1 - \frac{1}{n} \right) \eta.$$

It is clear that this term grows with n towards a value of 4η and will eventually exceed its limit $c_{\text{safety}} 2\eta$ as defined in (2.18). In particular, even if we choose $c_{\text{safety}} = 1$ (i.e., allow the smaller eigenvalue of $H^{\text{sp}d}$ to be equal to zero, then the condition will only be satisfied for $n \leq 2$, i.e., if the strain weakening is not too pronounced. For smaller values of c_{safety} , we can only choose $\alpha = 1$ if n is even less than that – for example, with $c_{\text{safety}} = 0.9$, the condition is only satisfied only if $n \lesssim 1.82$.

It is interesting to note that the condition is independent of the the flow field – what value α one has to choose is entirely decided by n and c_{safety} , and α will be the same at every quadrature point at which we integrate the bilinear form. Indeed, for this class of material model, we need to choose $\alpha = \min \left\{ \frac{1}{2} c_{\text{safety}} \frac{n}{n-1}, 1 \right\}$.

2.8.2 Drucker-Prager rheology

For Drucker-Prager rheologies, the viscosity in 2D is typically given by

$$\eta(\varepsilon(\mathbf{u})) = \frac{C \cos \phi + p \sin \phi}{2\sqrt{\frac{1}{2}\varepsilon(\mathbf{u}) : \varepsilon(\mathbf{u})}} = \frac{C \cos \phi + p \sin \phi}{2\sqrt{\frac{1}{2}}} (\|\varepsilon(\mathbf{u})\|^2)^{-1/2}.$$

Here, C is the cohesion and ϕ the friction angle. As for the power-law case, the viscosity only depends on $\|\varepsilon(\mathbf{u})\|^2$ and we know that the resulting matrix will always be symmetric.

By comparing the formula for η with the one from the power law rheology above, we see that up to a different (and possibly pressure dependent) pre-factor, the Prager-Drucker law corresponds to a power law with $n = \infty$. Thus, we expect that we will have to choose $\alpha < 1$ in (2.18). Indeed, we can compute that

$$\frac{\partial \eta(\varepsilon(\mathbf{u}))}{\partial \varepsilon} = -\eta(\varepsilon(\mathbf{u})) \frac{\varepsilon(\mathbf{u})}{\|\varepsilon(\mathbf{u})\|^2},$$

which matches the corresponding formula for the power law with $n = \infty$. Thus,

$$\left[1 - \frac{b : a}{\|a\| \|b\|} \right]^2 \|a\| \|b\| = [1 - (-1)]^2 |-1| \eta = 4\eta,$$

which is of course *never* less than 2η and consequently *always* violates the necessary condition in (2.18) to choose $\alpha = 1$. In other words, we can expect that the original Newton method will always lead to an ill-posed equation for the Newton update. On the other hand, (2.18) tells us that the choice

$$\alpha = \frac{1}{2} c_{\text{safety}}$$

will always lead to a well-posed equation with $c_{\text{safety}} < 1$.

It is interesting to note that for both the power law rheology with large n and the Prager-Drucker rheology, one *always* needs to choose $\alpha < 1$. This implies that the equations that define our stabilized Newton update are *never* the derivative of the residual, and we can consequently not expect quadratic convergence. As the calculations above show, this has, in fact, nothing to do with the concrete test case or setup: the choice of α does not depend on the value of the strain rate or other solution variables at a given point, but is the same throughout the entire domain.

2.8.3 The rheology of the Spiegelman et al. benchmark

To demonstrate that α does not need to be constant throughout the domain and may, in fact, depend on the flow field, we need to consider a rheology in which $\frac{\partial \eta}{\partial \boldsymbol{\varepsilon}} : \boldsymbol{\varepsilon}$ is not a fixed multiple of the viscosity as in the last two cases. This is indeed the case for the rheology of the benchmark by Spiegelman et al. discussed in Section 2.4.2. There, the viscosity is – up to a factor of 2 – given by the harmonic average of a linear rheology and the Drucker-Prager model considered above:

$$\eta(\boldsymbol{\varepsilon}(\mathbf{u})) = \frac{1}{\frac{1}{\eta_{\text{ref}}} + \frac{1}{\eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))}} = \frac{\eta_{\text{ref}} \eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))}{\eta_{\text{ref}} + \eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))}$$

where η_{ref} is a constant reference viscosity and $\eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))$ is the viscosity computed with the Drucker Prager rheology as shown above. The derivative of this equation is easily computed using the formulas from the previous section:

$$\begin{aligned} \frac{\partial \eta(\boldsymbol{\varepsilon}(\mathbf{u}))}{\partial \boldsymbol{\varepsilon}} &= \frac{\eta_{\text{ref}}^2}{(\eta_{\text{ref}} + \eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u})))^2} \frac{\partial \eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))}{\partial \boldsymbol{\varepsilon}} = -\frac{\eta_{\text{ref}}^2 \eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))}{(\eta_{\text{ref}} + \eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u})))^2} \frac{\boldsymbol{\varepsilon}(\mathbf{u})}{\|\boldsymbol{\varepsilon}(\mathbf{u})\|^2} \\ &= -\frac{\eta(\boldsymbol{\varepsilon}(\mathbf{u}))^2}{\eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))} \frac{\boldsymbol{\varepsilon}(\mathbf{u})}{\|\boldsymbol{\varepsilon}(\mathbf{u})\|^2}. \end{aligned}$$

Since this expression is again proportional to $\boldsymbol{\varepsilon}(\mathbf{u})$, the matrix \mathbf{J}^{uu} is again symmetric by construction, but not necessarily positive definite unless the expression

$$\left[1 - \frac{b : a}{\|a\| \|b\|} \right]^2 \|a\| \|b\| = [1 - (-1)]^2 \frac{\eta(\boldsymbol{\varepsilon}(\mathbf{u}))^2}{\eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))} = 4 \frac{\eta(\boldsymbol{\varepsilon}(\mathbf{u}))}{\eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))} \eta(\boldsymbol{\varepsilon}(\mathbf{u}))$$

is less than $c_{\text{safety}} 2\eta(\boldsymbol{\varepsilon}(\mathbf{u}))$. It is obvious that this is the case exactly if

$$\frac{\eta(\boldsymbol{\varepsilon}(\mathbf{u}))}{\eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))} \leq \frac{1}{2} c_{\text{safety}},$$

which is equivalent to the condition $\eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u})) \geq \left(1 + \frac{2}{c_{\text{safety}}}\right) \eta_{\text{ref}}$. This condition makes sense since large values of η_{DP} (compared to η_{ref}) yield a roughly constant viscosity $\eta(\boldsymbol{\varepsilon}(\mathbf{u})) \approx \eta_{\text{ref}}$ for which the negative contributions to \mathbf{J}^{uu} are small and we can consequently choose $\alpha = 1$.

On the other hand, if the condition above is not satisfied, then (2.18) tells us that we need to choose $\alpha = \frac{1}{2} c_{\text{safety}} \frac{\eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))}{\eta(\boldsymbol{\varepsilon}(\mathbf{u}))} = \frac{1}{2} c_{\text{safety}} \frac{\eta_{\text{ref}} + \eta_{\text{DP}}(\boldsymbol{\varepsilon}(\mathbf{u}))}{\eta_{\text{ref}}}$. Because this comparison does not simplify to anything that is independent of $\boldsymbol{\varepsilon}(\mathbf{u})$, the factor α that ensures positive definiteness will in general be spatially variable.

It is worth noting that the choice for α above is consistent with the results of the previous section. Namely, if one chooses $\eta_{\text{ref}} = \infty$, then the viscosity of this section equals the Prager-Drucker rheology. In that case, first, the condition $\eta_{\text{DP}} \geq \left(1 + \frac{2}{c_{\text{safety}}}\right) \eta_{\text{ref}}$ can never be satisfied; and second, the choice for α just derived simplifies to $\alpha = \frac{1}{2} c_{\text{safety}}$, which is exactly the value we have obtained for the Prager-Drucker rheology before. Fig. 2.3 also nicely shows that this is exactly the behavior we get in the benchmark: Where the strain rate is large, α drops to one half, whereas it is close to or exactly one in areas where the viscosity is dominated by the background viscosity.

2.8.4 The rheology of the Tosi et al. benchmark

Similar considerations also hold for the rheology used by the benchmark by Tosi et al. considered in Section 2.4.3. Indeed, based on the definition in (2.21) and repeated application of the chain rule, we have that

$$\begin{aligned} \frac{\partial \eta(\boldsymbol{\varepsilon}(\mathbf{u}))}{\partial \boldsymbol{\varepsilon}} &= 2 \left(\frac{1}{\eta_{\text{lin}}} + \frac{1}{\eta_{\text{plast}}(\boldsymbol{\varepsilon}(\mathbf{u}))} \right)^{-2} (\eta_{\text{plast}}(\boldsymbol{\varepsilon}(\mathbf{u})))^{-2} \frac{\partial \eta_{\text{plast}}(\boldsymbol{\varepsilon}(\mathbf{u}))}{\partial \boldsymbol{\varepsilon}} \\ &= -2 \left(\frac{1}{\eta_{\text{lin}}} + \frac{1}{\eta_{\text{plast}}(\boldsymbol{\varepsilon}(\mathbf{u}))} \right)^{-2} (\eta_{\text{plast}}(\boldsymbol{\varepsilon}(\mathbf{u})))^{-2} \frac{\sigma_Y}{\|\boldsymbol{\varepsilon}(\mathbf{u})\|} \frac{\boldsymbol{\varepsilon}(\mathbf{u})}{\|\boldsymbol{\varepsilon}(\mathbf{u})\|^2} \\ &= -\frac{1}{2} \left(\frac{\eta(\boldsymbol{\varepsilon}(\mathbf{u}))}{\eta_{\text{plast}}(\boldsymbol{\varepsilon}(\mathbf{u}))} \right)^2 (\eta_{\text{plast}}(\boldsymbol{\varepsilon}(\mathbf{u})) - \eta^*) \frac{\boldsymbol{\varepsilon}(\mathbf{u})}{\|\boldsymbol{\varepsilon}(\mathbf{u})\|^2}, \end{aligned}$$

where we have omitted the dependence of η on T, z for the moment because it is not relevant to our considerations. Since this expression is again proportional to $\boldsymbol{\varepsilon}(\mathbf{u})$, the matrix \mathbf{J}^{uu} is again symmetric by construction, but not necessarily positive definite unless the expression

$$\left[1 - \frac{\mathbf{b} : \mathbf{a}}{\|\mathbf{a}\| \|\mathbf{b}\|} \right]^2 \|\mathbf{a}\| \|\mathbf{b}\| = 2 \left(\frac{\eta(\boldsymbol{\varepsilon}(\mathbf{u}))}{\eta_{\text{plast}}(\boldsymbol{\varepsilon}(\mathbf{u}))} \right)^2 (\eta_{\text{plast}}(\boldsymbol{\varepsilon}(\mathbf{u})) - \eta^*)$$

is less than $c_{\text{safety}} 2\eta$. As in the previous section, the factor α that ensures positive definiteness will in general again be spatially variable.

2.9 Appendix C: Parameters for the 3d subduction test case

The following table provides the numeric values of all material parameters used in the subduction example shown in Section 2.4.4:

2.9. Appendix C: Parameters for the 3d subduction test case

thermal conductivity ($\text{W m}^{-1} \text{K}^{-1}$)	4
specific heat capacity ($\text{J kg}^{-1} \text{K}^{-1}$)	1250
reference temperature (K)	273.0
reference densities (kg)	3300
initial viscosity (Pa s)	10^{20}
cohesion C (Pa)	$20 \cdot 10^6$
angle of internal friction ϕ ($^\circ$)	30
dislocation stress exponent n_{disl}	3
dislocation prefactor A_{disl} ($\text{Pa}^{-n} \text{s}^{-1}$)	$3.12504 \cdot 10^{-14}$
dislocation activation energy E_{disl} (J mol^{-1})	$4.3 \cdot 10^5$
dislocation activation volume V_{disl} ($\text{m}^{-3} \text{mol}^{-1}$)	$25 \cdot 10^{-6}$
diffusion stress exponent n_{diff}	1
diffusion prefactor A_{diff} ($\text{Pa}^{-n} \text{s}^{-1}$)	$1.92 \cdot 10^{-11}$
diffusion activation energy E_{diff} (J mol^{-1})	$335 \cdot 10^3$
diffusion activation volume V_{diff} ($\text{m}^{-3} \text{mol}^{-1}$)	$4 \cdot 10^{-6}$
minimum viscosity η_{min} (Pa s)	10^{19}
maximum viscosity η_{max} (Pa s)	10^{24}

In addition, we choose dimensionless scaling factors $\nu_{\text{disl}} = \nu_{\text{diff}} = 1$ in the upper mantle, and $\nu_{\text{disl}} = \nu_{\text{diff}} = 100$ in the lower mantle.

3

The Geodynamic World Builder: a solution for complex initial conditions in numerical modelling

3.1 Abstract

The Geodynamic World Builder is an open source code library intended to set up initial conditions for computational geodynamic models in both Cartesian and Spherical geometries. The inputs for the JSON-style parameter file are not mathematical, but rather a structured nested list describing tectonic features, e.g. a continental, an oceanic or a subducting plate. Each of these tectonic features can be assigned a specific temperature profile (e.g. plate model) or composition label (e.g. uniform). For each point in space, the Geodynamic World Builder can return the composition and/or temperature. It is written in C++, but can be used in almost any language through its C and Fortran wrappers. Various examples of 2D and 3D subduction settings are presented. The World builder comes with an extensive online User Manual.

3.2 Introduction

Geodynamic modelling has been used in the past four decades to help us better understand the physical processes of Earth's interior including large-scale mantle convection and plate tectonics, or detailed processes of crustal deformation. Numerical modelling of geodynamic processes involves solving the pertinent partial differential equations (PDEs) of mass, momentum and energy conservation supplemented with rheological laws, material parameters and with an equation of thermodynamic state relating, e.g., density, temperature and pressure (e.g. Gerya, 2010; Schubert et al., 2001). In addition these PDEs must be constrained by boundary conditions, which can be time-dependent, and by initial conditions which describe the starting model for solving the geodynamic problem at hand. For example, 3D initial models of a geometrically simplified nature are often constructed for modelling of generic subduction evolution using plate boundaries and lithosphere domains that are parallel to the sides of the (rectangular) model domain (e.g. Yamato et al., 2009; Stegman et al., 2010b; Brune and Autin, 2013; Schellart and Moresi, 2013; Duretz et al., 2014; Holt et al., 2015; Leng and Gurnis, 2015; Naliboff and Buitert, 2015; Kiraly et al., 2016; Schellart, 2017). When numerically simulating (regions of) the Earth, geometrically more complex initial models are required, e.g., involving

This paper has been submitted to Solid Earth as *The Geodynamic World Builder: a solution for complex initial conditions in numerical modelling*, M. Fraters, C. Thieulot, A. van den Berg and W. Spakman.

the starting plate-tectonic layout, initial trench geometry and slab shape for use either instantaneous dynamics modelling or as initial model for modelling of subduction evolution (e.g. Alisic et al., 2012; Liu and Stegman, 2011; Jadamec and Billen, 2010, 2012; Chertova et al., 2014a; Billen and Arredondo, 2018; Zhou et al., 2018). Such initial model setups cannot be easily created, adapted, or shared with the community, nor easily transferred to another code. We present in this paper a solution to these problems in the form of an open source code library, the Geodynamic World Builder, which has been designed to be user-friendly, extensible, and portable across different platforms. We present the first stable version of the World Builder which focuses on creating geometrically complex 3D initial models (geometry, composition, and temperature) consisting of first-order plate tectonic features such as continental and oceanic plates, oceanic ridges and transform faults and 3D lithosphere subduction. These configured initial models are intended to help advance research into simulations of instantaneous dynamic modelling and of plate tectonic evolution with a wide range of geometric complexity.

3.3 Geodynamic World Builder Philosophy

3.3.1 User Philosophy

In this section we describe the philosophy of how tectonic features such as plates, ridges, faults and slabs can be parameterized by lines and areas that implicitly define volumes to which temperature and composition can be assigned. A composition is a part of the model that is assigned a particular identifying label and in addition an indicator which is given a value between 0 and 1. This indicator can be used by codes using the GWB output to ascribe physical properties to different model regions.

To minimize user effort, the Geodynamic World Builder (GWB) utilizes a parameterization of 3D structures by 2D coordinate input, by defining their (projected) location on the surface. The GWB can be used to create initial models in Cartesian and spherical geometries.

User input files should be specified in JSON (json.org), which is an internationally standardized language (ISO/IEC 21778). We use a relaxed form of JSON which allows comments, NaN's and trailing commas to improve usability through RapidJSON (<http://rapidjson.org/>). The user inputs coordinates and can assign particular properties to features such as 'linear' for a temperature profile, or 'uniform' for the compositional makeup of the plate. Note that only a subset of the options is mentioned in this paper. We refer to the online Geodynamic World Builder Manual (<https://geodynamicworldbuilder.github.io>) for the complete listing.

The GWB uses a hierarchical overlay of features. This means that features defined first are spatially overlain by features defined later in places where both overlap. The GWB recognizes two types of features: area features and line features, which will be explained in the following sections. A possible third type of features, point features, will be discussed in section 3.5.

Continental lithosphere plate

A continental plate is an 'area feature' in the GWB and is defined by its surface perimeter and its thickness. The perimeter is specified as a list of points which enclose the continental area. Within the defined volume of the continental plate, the GWB offers various options for defining temperature values and compositions. For example, a continental plate can be assigned multiple layers of different compositions and a linear geotherm that matches a predefined adiabatic mantle temperature at the base of the lithosphere. We note that continental lithosphere with a variable thickness is a development for future releases of the GWB, but can be mimicked in the present version by specifying contiguous continental areas with different thickness. Also, continental topography is currently not explicitly implemented, but it can be achieved through a sticky air approach, where air is a composition of varying thickness atop the model (Schmeling et al., 2008; Cramer et al., 2012).

Oceanic lithosphere plate

Like the continental plate, the oceanic plate is parameterized as an area feature with a flat surface. We have implemented the 'plate model' (e.g. Fowler, 2005) for assigning an age-dependent temperature to oceanic lithosphere. In section 3.4.1 we will show an example of a ridge-transform system with ridge jumps. The workaround for implementing oceanic bathymetry is the same as for the continental lithosphere plate.

The mantle

The upper and lower mantle can also be parameterized as an area feature that starts below the lithosphere or at the surface and is overlain by lithosphere in a later building stage. This allows for defining an upper and lower mantle and to insert specific volumetric structures such as Large Low Shear wave Velocity Provinces (LLSVPs) at the core-mantle boundary. In the present version these mantle features can be assigned a radially uniform, linear or adiabatic temperature profile. Future versions may include laterally varying temperature or compositions, e.g. scaled from seismic tomography models (e.g. Steinberger et al., 2015).

A subducting plate

A subducting plate is a 'line feature' in the GWB and is defined by the location of the trench and one or more depth segments each describing a part of the geometry of the subducting slab. They are defined by a length and by thicknesses and dip angles at beginning and end of the slab segment. In sequence, these segments can make up a smoothly varying slab geometry which can for example flatten in the upper mantle transition zone, or may prescribe a slab entering the lower mantle. Every point in the trench coordinate list defines a vertical section of the subducting plate that may consist of one or several slab segments. Both sections and segments can vary in length, dip angle or thickness. The length of a subducting slab is always computed as the length along the top of the slab so that this can straightforwardly represent the amount of relative plate convergence during a certain period. The dip angle is defined as the angle between the surface and the local plunge of the slab. The dip angle is specified at the start and end point of each depth segment along the vertical section. Dip angles are linearly interpolated along a segment. The overall direction of slab dip can be to either side of the trench and is selected by specifying for each subducting plate an additional point at the surface, the 'dip-point', at the slab dip-side of the trench segment. Slab dip is linearly interpolated between subsequent vertical slab segments. This parameterization allows for constructing smoothly varying 3D slab morphology. Note that it is also possible to give slabs a starting depth to configure detached slabs.

For each point at the surface of the slab the depth and the distance to the trench, as measured along the surface, are available and can be used to assign slab temperatures, e.g., by using the McKenzie (1970) slab temperature model.

A fault

To allow for complicated fault shapes (e.g. listric faults), faults are also parameterized as line features. An important difference between faults and subducting plates is that for subducting plates the trench defines the top of the plate at the plate boundary, while for faults the line feature defines the center of the fault with respect to which a fault thickness can be defined.

3.3.2 Code philosophy

The following design principles define the Geodynamic World Builder:

1. *A single text-based input file centered around plate tectonic terminology:* as explained in Section 3.3.1. The particular syntax is specified in the online manual and will be illustrated with examples below.

2. *Code-, language-, and platform-independence:* The GWB is designed to be integrated in the different geodynamic codes through a simple interface. The library is written in C++, has official interfaces (wrappers) to C and Fortran and it is possible to call the GWB from the command line. Note that the C wrapper enables calling the GWB from almost any other language like Python and Matlab. The code is continuously tested with every change on the Linux, OSX and Windows operating systems.
3. *Up-to-date user manual and code documentation.* Manual and doxygen <http://doxygen.nl/> code documentation provided through <https://geodynamicworldbuilder.github.io>.
4. *Safe use in parallel codes:* The GWB is split into two phases. The setup phase, encapsulated in the function `create_world`, is not thread safe but upon completion the generated "world object" is thread-safe and can be used to query temperature and compositions in parallel.
5. *Readable and extensible code:* Following ASPECT (Kronbichler et al., 2012; Heister et al., 2017) we use a plugin system for different parts of the code. Such plugins enable users to add functionalities such as plate tectonic features or coordinate systems without knowledge of the rest of the code.
6. *Version numbering:* using Semantic Versioning 2.0.0 (<https://semver.org>). The input file should specify the major version number that must match the version number of the used GWB. Before the release of major version 1, backwards incompatible changes may be made in minor versions, because they will be beta releases. This implies that the input files for major version 0 also must contain the minor version number. All these features help ensuring *reproducibility of results*.

3.4 Using the World Builder

To exemplify input files and to show the capabilities of the Geodynamic World Builder, we show here three 2D examples, and two 3D examples of the GWB visualized through the standalone visualization application. This application creates so-called `vtu` files which can be visualized by programs like Paraview (paraview.org). Furthermore, we show examples of GWB use with the SEPRAN (van den Berg et al., 2015), ELEFANT (Plunder et al., 2018) and ASPECT codes. The annotated input files to create these models are presented in appendixes 3.8 to 3.13 and are part of the GWB repository.

3.4.1 Standalone examples

The GWB has an option to create a Paraview file of the GWB input file. This can be useful for model creation or visualization support of presenting geodynamic hypotheses, or for checking the user-designed model prior to using it in a next step, e.g., for creating an initial model for geodynamic modelling.

2D subduction

Here we show two subduction models, one in Cartesian coordinates (Fig. 3.1) and the same model in spherical (effectively cylindrical) coordinates (Fig. 3.2), which were created through the input files in appendix 3.8. These input files only differ in the selected coordinate system and whether the supplied coordinates are in meters or in degrees. The model has a 95 km thick oceanic plate of which the top 10 km defines the crust and which turns into a 500 km long subducting slab in the center of the domain. The temperature in the oceanic plate follows the plate model (Fowler, 2005) with a bottom temperature of 1600 K. The slab temperature is computed using the McKenzie model for a particular slab history. The model also contains a 100 km thick continental plate of which the top 30 km is crust. Furthermore, the upper and lower mantle are given different compositions and follow a linear temperature profile in the upper mantle from 1600 K at 95 km depth to 1820 K at 660 km depth, and in the lower mantle from 1820 at 660 km depth to 2000 at 1160 km depth.

This example is created by placing the features in a particular order in the input file. The features overlay, and in this case overwrite, an adiabatic background temperature and all compositions set to zero. This example consists of five features: an oceanic plate, a continental plate, an upper mantle, a lower mantle and a subducting plate. The first four do not overlap in their input definition, so the order of definition in the Geodynamic World Builder input file does not make a difference in the result. The subducting plate overwrites parts of the oceanic plate, continental plate and the upper mantle, which is effectuated by defining the slab after these three features. For each feature temperature and composition models are selected.

3D ocean spreading

We show in figure 3.3 a 3D rifting model with two rift systems next to each other. The temperature is defined by the plate model. The mantle is given an adiabatic geotherm defined by $\theta_S \exp(\alpha g d / C_p)$, where θ_S is the potential surface temperature of the mantle, α is the thermal expansion coefficient, g is the gravitational acceleration, C_p is the specific heat and d is the depth. The input

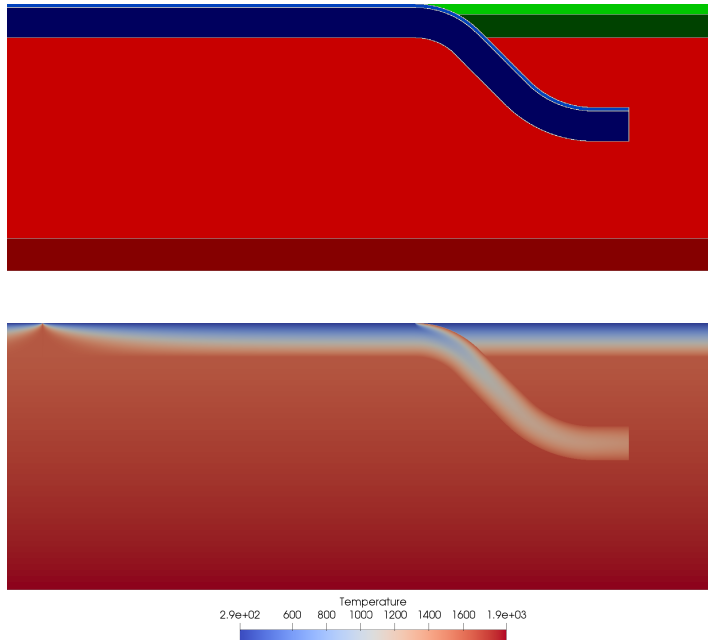


Figure 3.1: The top figure shows the distribution of different compositions through the model domain. The oceanic crust composition is light blue, the oceanic lithosphere is dark blue, the continental crust is light green, the continental lithosphere is dark green, the upper mantle is light red and the lower mantle dark red. The bottom figure shows the temperature distribution in the model (in Kelvin).

file of this example consists of the definition of the mantle domain followed by two oceanic plates, which form the two ridge-plate systems. The two oceanic plates are exactly the same, except for the shifted ridge location. The input file for this example can be found in Appendix 3.9.

3D subduction

Figure 3.4 shows a 3D example defining a subduction geometry similar to the one in Plunder et al. (2018). In this example the trench consists of three connected straight lines. To create a smooth transition between these sections, the user can choose to use a monotone spline interpolation between the coordinates given by the user. This example includes a linear temperature upper and lower mantle as described in the 2D subduction example. The 95 km thick oceanic plate and the 120 km thick continental plate features are both defined before the subducting plate feature, of which the trench is defined along the interface between the two. The slab itself is 95 km thick and consists of four

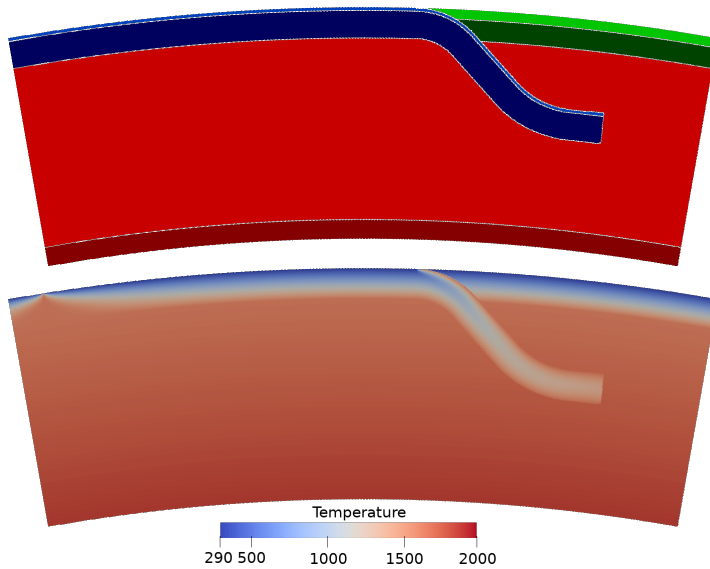


Figure 3.2: The same as setup as in figure 3.1, but now in spherical geometry. The top figure shows the composition, the bottom figure shows the temperature.

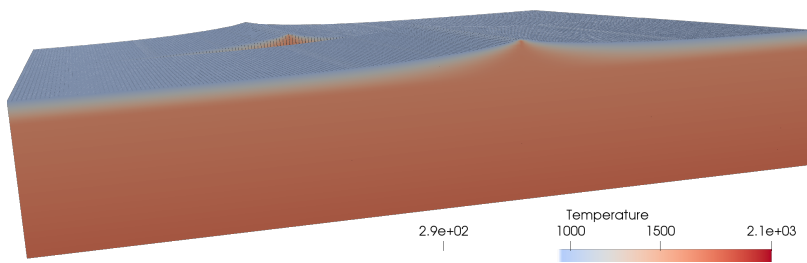


Figure 3.3: The temperature field of the 3D two rift systems example. Material with a temperature below 950 K has been omitted, in order to better show the rifts. Note the second rift system in the background.

3

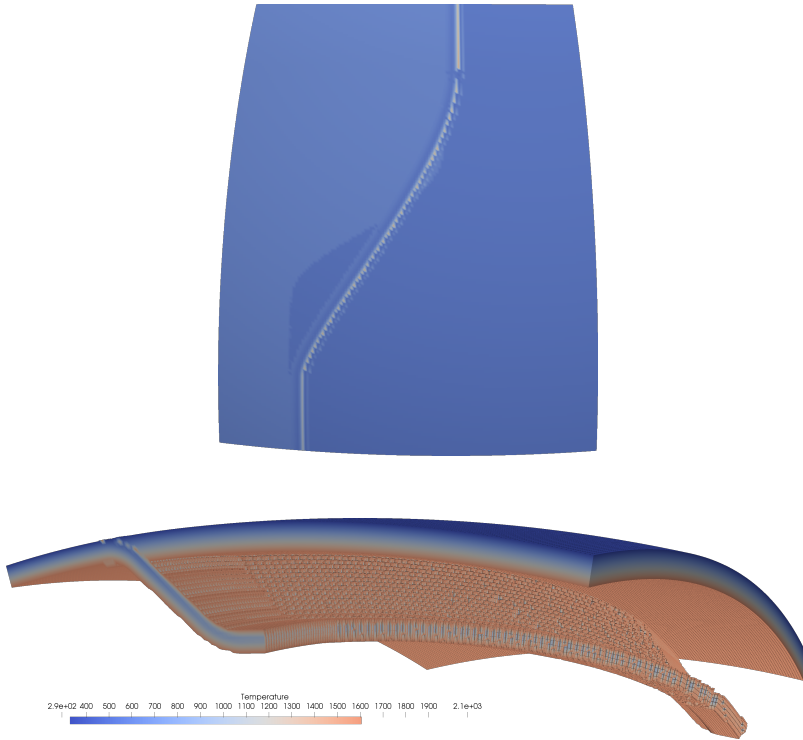


Figure 3.4: The temperature field of the 3D subduction example. Note the smooth transition between the upper and lower part of the subduction system in the top figure and the curved geometry of the slab in the lower figure. For visualization purposes we have omitted the top 25 km of the model in the top figure.

segments. One 200 km long segment which goes from a dip angle of 0° to 45° , and one 400 km long segment which has an angle of 45° , one 200 km long segment which goes from 45° to 0° and one 100 km long segment, with constant dip angle of 0° . The input file for this example can be found in Appendix 3.10.

3.4.2 Using the GWB with SEPRAN

SEPRAN is a general purpose finite element toolkit applied in engineering problems as well as in development of 2D and 3D numerical models in geodynamics and planetary science (Chertova et al., 2012, 2014a; Čížková et al., 2012; van den Berg et al., 2015, 2019; Zhao et al., 2019). The model contains a lithospheric slab subducting under an overriding plate as shown in Fig. 3.5. One sided subduction is obtained in a self consistent way by the presence of a weak crustal layer of uniform viscosity 10^{21} Pa·s as part of the top of the subducting

lithosphere. The mantle underlying the crust has a temperature and pressure dependent viscosity with an Arrhenius type parameterization representative of diffusion creep in olivine under upper mantle pressure and temperature conditions. Viscosity is modeled as a material property for the crustal layer material and the mantle material. Material transport is implemented using particle tracers that are advected by the convective flow. The medium is described as a mechanical mixture of materials with contrasting properties.

A 2D rectangular domain of 1000 km depth and 2000 km width is used. The initial thermal and composition state is created using the Fortran wrapper of the GWB library. The GWB tool is called in a loop over all nodal points of the FEM mesh to define the initial temperature field for the subsequent convection calculations. In a similar way the material distribution of the initial state is defined by calling the composition function of the GWB library in a program loop over particle tracers. The input file for this example can be found in Appendix 3.11.

3.4.3 Using the GWB with ELEFANT

ELEFANT is a 2D/3D Finite Element code for geodynamic problems (Maffione et al., 2015; Lavecchia et al., 2017; Thieulot, 2017; Plunder et al., 2018) written in Fortran. It principally relies on bi/tri-linear velocity-constant pressure elements and uses the Marker-in-Cell technique to track materials. In order to demonstrate the GWB flexibility of use a 3D double subduction setup was created with the Fortran wrapper of the GWB (see Fig. 3.6): a composition between 1 and 6 was then easily assigned to all markers (two different oceanic crusts and oceanic lithospheres, one upper mantle and one lower mantle) and a temperature based on the McKenzie model (McKenzie, 1970) was prescribed onto the FE mesh, as shown in Fig. 3.7.

The domain is a Cartesian box with dimensions $2000 \times 2000 \times 800$ km and the Finite Element mesh counts $120 \times 120 \times 50 = 720,000$ elements. Each element contains 64 randomly distributed markers. Free slip boundary conditions are imposed at the bottom ($z = 0$), top ($z = L_z$) and sides ($y = 0$ and $y = L_y$) of the domain. The other two sides, $x = 0$ and $x = L_x$, are a mix of free slip (for $z < 100$ km or $z > 690$ km) and open boundary conditions (for $100 < z < 690$ km) (Chertova et al., 2012). The input file for this example can be found in Appendix 3.12.

3.4.4 Using the GWB with ASPECT

ASPECT is an open source community FEM designed for geodynamic problems (Heister et al., 2017; Kronbichler et al., 2012). The model which was run

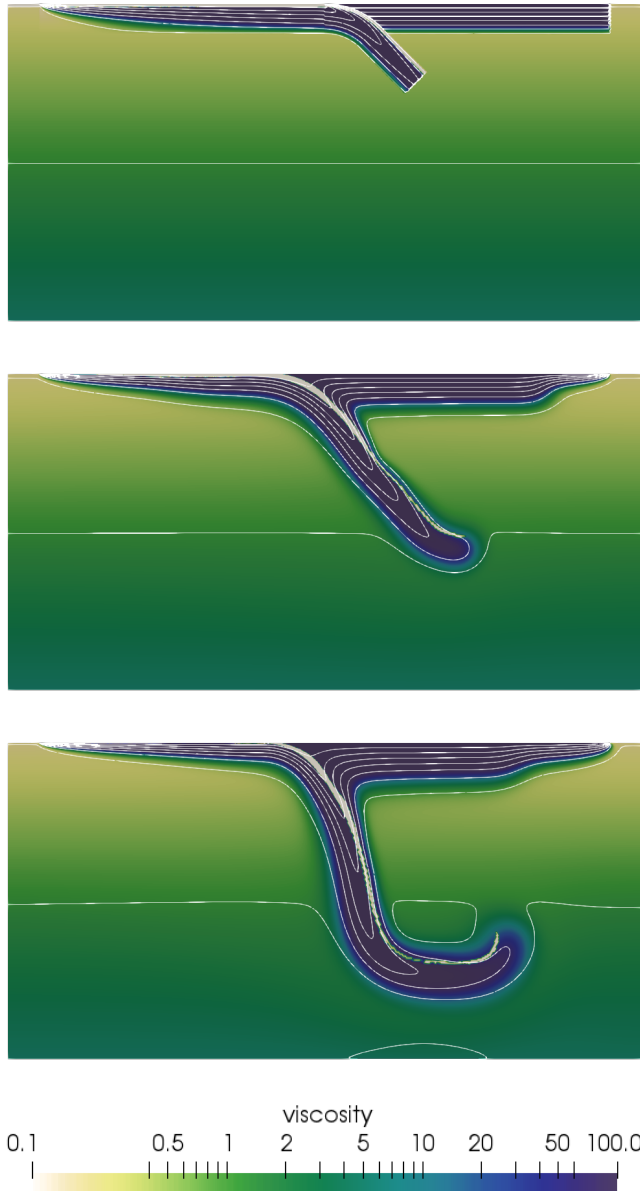


Figure 3.5: SEPRAN model. Dimensionless viscosity field in log scale superimposed with 10 (dimensionless) temperature (between 0 and 0.82) isocontours.

```
1 do im=1,nmarker
2   depth=Lz-zm(im)
3   do imat=1,nmat
4     call composition_3d(cworld,xm(im),&
5                        ym(im),zm(im),&
6                        depth,imat-1,flag)
7     if (flag==imat) then
8       mat(im)=imat
9     exit
10    end if
11  end do
12 end do
13
14 do ip=1,np
15   depth=Lz-z(ip)
16   call temperature_3d(cworld,x(ip),y(ip),&
17                     z(ip),depth,gz,T(ip))
18 end do
```

Figure 3.6: Example ELEFANT query routine using the GWB supplied Fortran wrappers `composition_3d()` and `temperature_3d()`: a) a loop runs over all markers and determines for each the composition at its location; b) a loop runs over all grid points and the GWB returns its temperature as a function of their spatial coordinates.

with ASPECT is a 3D Cartesian model of a curved subduction system similar to the plate-tectonic setting of the Lesser Antilles subduction of the eastern Caribbean region. The lithosphere consists of a strong zero velocity Caribbean upper plate, surrounded by an oceanic North American plate to the north and northeast and the oceanic-continental South American plate to the south and southeast. In the model the North American and South American plates move west at a average rate over the past 5 Ma of 1.4 cm/yr relative to the Caribbean plate (Boschman et al., 2014). The Lesser Antilles trench curves around the east and north of the Caribbean plate. To the south, the Caribbean plate is partially decoupled from the South American plate by a 50 km wide weak zone. To the northwest a 250 km wide weak zone, from the western end of the trench to the western edge of the model, partially decouples the North American plate from the Caribbean plate. Below the lithosphere the sidewalls are open (Chertova et al., 2012, 2014a) allowing for horizontal in/out flow of mantle material. From 660 km down a denser and more viscous material has been prescribed to delay sinking of the slab into the lower mantle. The top boundary is a free surface

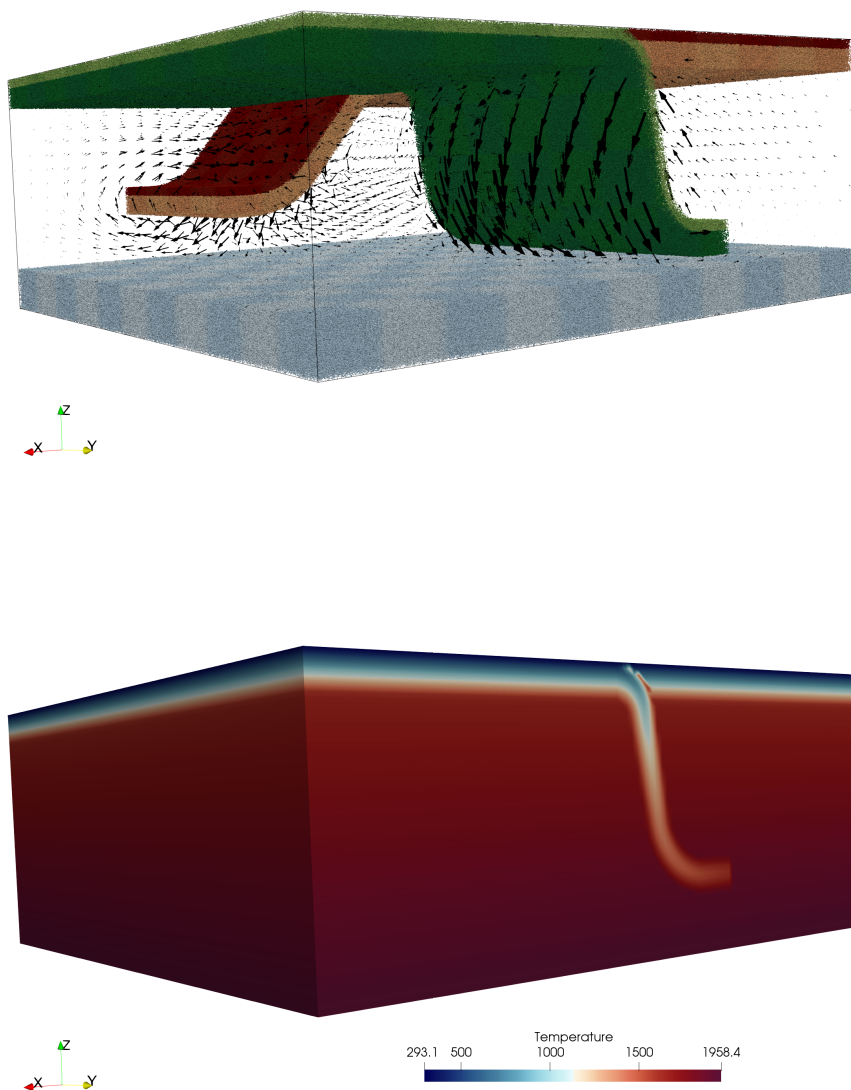


Figure 3.7: ELEFANT model. Top: Markers for 5 compositions (the mantle markers have been left out for ease of visualization) with the resulting velocity field; Bottom: Temperature field.

3.4. Using the World Builder

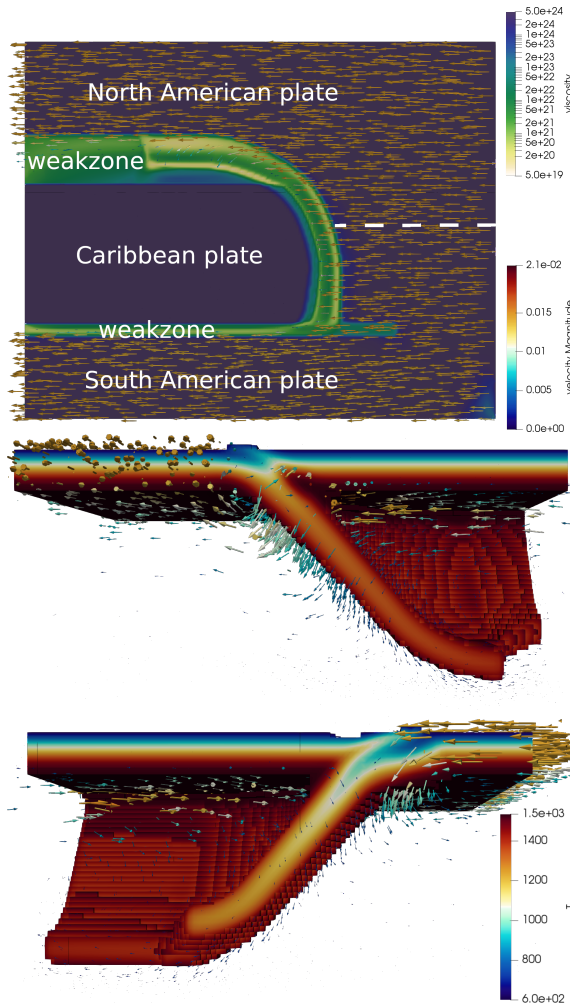


Figure 3.8: The 3D ASPECT Caribbean example after 2.5 million years of evolution. The top image is a top view of the model, where the top 50 km is removed, and where the viscosity field is shown with the velocity field indicated by the arrows. The bottom two figures are cut outs of the temperature field between 600 K and 1535 K, showing in colour the temperature (T) and with arrows the velocity fields, highlighting the velocity field in the slab and lithosphere.

(Rose et al., 2017) and the bottom boundary has a prescribed zero velocity. The result of about 2.5 million years of evolution is shown in figure 3.8.

The details of the setup are presented in Appendix 3.13.

3.4.5 Performance

The Finite Element mesh used in the example of section 3.4.4 is built in several steps by ASPECT: the code starts with a regular grid and allows adaptive mesh refinement to take place one level at the time. Each step of this process calls the GWB library. The first step generates a grid counting 28,000 elements and reports a total setup time for the initial conditions of 3.6 seconds on 480 MPI processes. The second step mesh counts 99,000 elements while the setup of the initial conditions took (cumulatively) 10 seconds. The third step sees the number of element jump to about 560,000 elements while its total (cumulative) time to setup the initial conditions remains low at about 36 seconds. This figure represents about 0.7% of the total wall time of the first time step, and a negligible portion of the total wall time of the 20Myr-long simulation.

3.5 Discussion

We presented the Geodynamic World Builder Version 0.2.0 as a tool for constructing 2D and 3D initial models of geodynamic settings involving crust/lithosphere, plate boundaries, and subduction. The interface of the GWB with a numerical modelling code is based on a query of the modelling code to supply temperature, density, or other information at a particular position. This paper discusses version 0.2.0 of the Geodynamic World Builder, which is considered to be a beta version of the code. Input format and/or functionality may change between minor versions and this will be documented on the website. From version 1.0.0, we will use Semantic Versioning 2.0.0 (<https://semver.org/spec/v2.0.0.html>), and backwards incompatible changes will only be made in every major version of the code. Future improvements may for example include extra temperature or composition modules, e.g. derived from tomographic models, new or improved features or even new output interfaces, e.g. velocity boundary conditions or initial topography. As an extension to area and line features, adding point features are another possible improvement to the Geodynamic World Builder. These can represent, for example, a spherical weak seed or a plume. Because of a simple query interface it is in principle possible to use the GWB in connection with existing numerical modelling codes used by the geodynamic community. The use of the GWB can also just be restricted to creating 2D or 3D geodynamic models/cartoons for, e.g., teaching purposes or for illustrating a complex geodynamic setting.

3.6 Code availability

The code is freely available at <https://geodynamicworldbuilder.github.io> under licence LGPLv2.1. All examples presented in this work are available as cookbooks in the code.

3.7 Acknowledgements

M.F. acknowledges constructive feedback from the ASPECT community, and especially from T. Heister, W. Bangerth and R. Gassmüller. The authors also acknowledge constructive proofreading by R. Myhill, H. Brett and L. van de Wiel. MF and CT are indebted to the Computational Infrastructure for Geodynamics (CIG) for their recurring participation to the ASPECT hackathons, during which the foundation of this work was laid out. This work is funded by the Netherlands Organization for Scientific Research (NWO), as part of the Caribbean Research program, grant 858.14.070 and partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 223272. Data visualization is carried out with ParaView software <https://paraview.org/>.

3.8 2D subduction examples

3.8.1 Cartesian input file

```
1 {
2   "version":"0.2",
3   "cross section":[[0,0],[100,0]],
4   "features":
5   [
6     // defining the oceanic plate
7     {
8       "model":"oceanic plate", "name":"oceanic plate",
9       "coordinates":[[-1e3,-1e3],[1150e3,-1e3],[1150e3,1e3],[-1e3,1e3]],
10      "temperature models":
11      [
12        {"model":"plate model", "max depth":95e3, "bottom temperature":1600,
13         "spreading velocity":0.005,
14         "ridge coordinates":[[100e3,-1e3],[100e3,1e3]]}
15      ],
16      "composition models":
17      [
18        {"model":"uniform", "compositions":[0], "max depth":10e3},
19        {"model":"uniform", "compositions":[1], "min depth":10e3,
20         "max depth":95e3}
21      ]
22    },
23    // defining a continental plate
24    {
25      "model":"continental plate", "name":"continental plate",
26      "coordinates":[[1150e3,-1e3],[2001e3,-1e3],[2001e3,1e3],[1150e3,1e3]],
27      "temperature models":
28      [
29        {"model":"linear", "max depth":95e3, "bottom temperature":1600}
30      ],
31      "composition models":
32      [
33        {"model":"uniform", "compositions":[2], "max depth":30e3},
34        {"model":"uniform", "compositions":[3], "min depth":30e3,
35         "max depth":65e3}
36      ]
37    },
38    // defining the upper mantle
```

```

39 {
40   "model": "mantle layer", "name": "upper mantle",
41   "min depth": 95e3, "max depth": 660e3,
42   "coordinates": [[-1e3, -1e3], [2001e3, -1e3], [2001e3, 1e3], [-1e3, 1e3]],
43   "temperature models":
44   [
45     {"model": "linear", "min depth": 95e3, "max depth": 660e3,
46      "top temperature": 1600, "bottom temperature": 1820}
47   ],
48   "composition models": [{"model": "uniform", "compositions": [4]}]
49 },
50 // defining the lower mantle
51 {
52   "model": "mantle layer", "name": "lower mantle",
53   "min depth": 660e3, "max depth": 1160e3,
54   "coordinates": [[-1e3, -1e3], [2001e3, -1e3], [2001e3, 1e3], [-1e3, 1e3]],
55   "temperature models":
56   [
57     {"model": "linear", "min depth": 660e3, "max depth": 1160e3,
58      "top temperature": 1820, "bottom temperature": 2000}
59   ],
60   "composition models": [{"model": "uniform", "compositions": [5]}]
61 },
62 // defining the subducting plate dipping towards the continental plate
63 {
64   "model": "subducting plate", "name": "Subducting plate",
65   "coordinates": [[1150e3, -1e3], [1150e3, 1e3]], "dip point": [2000e3, 0],
66   "segments": [{"length": 200e3, "thickness": [95e3], "angle": [0, 45]},
67                {"length": 200e3, "thickness": [95e3], "angle": [45]},
68                {"length": 200e3, "thickness": [95e3], "angle": [45, 0]},
69                {"length": 100e3, "thickness": [95e3], "angle": [0]}],
70   "temperature models":
71   [
72     {"model": "plate model", "density": 3300, "plate velocity": 0.01 }
73   ],
74   "composition models":
75   [
76     {"model": "uniform", "compositions": [0], "max distance slab top": 10e3},
77     {"model": "uniform", "compositions": [1], "min distance slab top": 10e3,
78      "max distance slab top": 95e3 }
79   ]
80 }
81 ]
82 }

```

Listing 3.1: 2D Cartesian subduction example. The lines of green text (preceded by the double forward slashes) are comments and have no effect on the result.

3.8.2 Spherical input file

```

1 {
2   "version": "0.2",
3   "cross section": [[0, 0], [10, 0]],
4   "features":
5   [
6     // defining the oceanic plate
7     {
8       "model": "oceanic plate", "name": "oceanic plate",
9       "coordinates": [[-1, -1], [11.5, -1], [11.5, 1], [-1, 1]],
10      "temperature models":
11      [
12        {"model": "plate model", "max depth": 95e3, "bottom temperature": 1600,
13         "spreading velocity": 0.005,
14         "ridge coordinates": [[1, -1], [1, 1]]}

```

3.8. 2D subduction examples

```
15 ],
16   "composition models":
17   [
18     {"model":"uniform", "compositions":[0], "max depth":10e3},
19     {"model":"uniform", "compositions":[1], "min depth":10e3,
20      "max depth":95e3}
21   ],
22 },
23 // defining a continental plate
24 {
25   "model":"continental plate", "name":"continental plate",
26   "coordinates":[[[11.5,-1],[21,-1],[21,1],[11.5,1]],
27   "temperature models":
28   [
29     {"model":"linear", "max depth":95e3, "bottom temperature":1600}
30   ],
31   "composition models":
32   [
33     {"model":"uniform", "compositions":[2], "max depth":30e3},
34     {"model":"uniform", "compositions":[3], "min depth":30e3,
35      "max depth":65e3}
36   ]
37 },
38 // defining the upper mantle
39 {
40   "model":"mantle layer", "name":"upper mantle",
41   "min depth":95e3, "max depth":660e3,
42   "coordinates":[[[-1,-1],[21,-1],[21,1],[-1,1]],
43   "temperature models":
44   [
45     {"model":"linear", "min depth":95e3, "max depth":660e3,
46      "top temperature":1600, "bottom temperature":1820}
47   ],
48   "composition models":[{"model":"uniform", "compositions":[4]}]
49 },
50 // defining the lower mantle
51 {
52   "model":"mantle layer", "name":"lower mantle",
53   "min depth":660e3, "max depth":1160e3,
54   "coordinates":[[[-1,-1],[21,-1],[21,1],[-1,1]],
55   "temperature models":
56   [
57     {"model":"linear", "min depth":660e3, "max depth":1160e3,
58      "top temperature":1820, "bottom temperature":2000}
59   ],
60   "composition models":[{"model":"uniform", "compositions":[5]}]
61 },
62 // defining the subducting plate dipping towards the continental plate
63 {
64   "model":"subducting plate", "name":"Subducting plate",
65   "coordinates":[[[11.5,-1],[11.5,1]], "dip point":[20,0],
66   "segments":[{"length":200e3, "thickness":[95e3], "angle":[0,45]},
67     {"length":200e3, "thickness":[95e3], "angle":[45,0]},
68     {"length":200e3, "thickness":[95e3], "angle":[45,0]},
69     {"length":100e3, "thickness":[95e3], "angle":[0]}],
70   "temperature models":
71   [
72     {"model":"plate model", "density":3300, "plate velocity":0.01 }
73   ],
74   "composition models":
75   [
76     {"model":"uniform", "compositions":[0], "max distance slab top":10e3},
77     {"model":"uniform", "compositions":[1], "min distance slab top":10e3,
78      "max distance slab top":95e3 }
79   ]

```

```
80 }
81 ]
82 }
```

Listing 3.2: 2D Spherical subduction example. The lines of green text (preceded by the double forward slashes) are comments and have no effect on the result.

3.9 3D ocean spreading example input file

```
1 {
2   "version": "0.2",
3   "features":
4   [
5     // defining one of the oceanic plates with a ridge
6     {
7       "model": "oceanic plate", "name": "oceanic plate A",
8       "coordinates": [[-1e3, -1e3], [2001e3, -1e3], [2001e3, 1000e3], [-1e3, 1000e3]],
9       "temperature models":
10      [
11        {
12          "model": "plate model", "max depth": 95e3, "spreading velocity": 0.005,
13          "ridge coordinates": [[1200e3, -1e3], [1200e3, 1000e3]]
14        }
15      ],
16      "composition models":
17      [
18        {"model": "uniform", "compositions": [0], "max depth": 10e3},
19        {"model": "uniform", "compositions": [1], "min depth": 10e3,
20         "max depth": 95e3}]
21      ],
22      // defining the other oceanic plate with a ridge
23      {
24        "model": "oceanic plate", "name": "oceanic plate B",
25        "coordinates": [[-1e3, 1000e3], [2001e3, 1000e3], [2001e3, 2001e3], [-1e3, 2001e3
26         ]],
27        "temperature models":
28        [
29          {
30            "model": "plate model", "max depth": 95e3, "spreading velocity": 0.005,
31            "ridge coordinates": [[800e3, 1000e3], [800e3, 2000e3]]
32          }
33        ],
34        "composition models":
35        [
36          {"model": "uniform", "compositions": [0], "max depth": 10e3},
37          {"model": "uniform", "compositions": [1], "min depth": 10e3,
38           "max depth": 95e3}]
39        ]
40      }
41    ]
42  }
```

Listing 3.3: 3d ocean spreading example input file. The lines of green text (preceded by the double forward slashes) are comments and have no effect on the result.

3.10 3D subduction example input file

```
1 {
2   "version": "0.2",
3   "coordinate system": {"model": "spherical", "depth method": "begin segment"},
4   "cross section": [[0, 0], [10, 0]],
5   "maximum distance between coordinates": 0.01,
6   "interpolation": "monotone spline",
```

3.10. 3D subduction example input file

```
7 "features":
8 [
9   // defining the upper mantle
10  {
11    "model": "mantle layer", "name": "upper mantle",
12    "min depth": 95e3, "max depth": 660e3,
13    "coordinates": [[-1, -1], [41, -1], [41, -1], [-1, -1]],
14    "temperature models":
15    [
16      {
17        "model": "linear", "min depth": 95e3, "max depth": 660e3,
18        "top temperature": 1600, "bottom temperature": 1820
19      }
20    ],
21    "composition models": [{"model": "uniform", "compositions": [4]}]
22  },
23  // defining the lower mantle layer
24  {
25    "model": "mantle layer", "name": "lower mantle",
26    "min depth": 660e3, "max depth": 1160e3,
27    "coordinates": [[-1, -1], [41, -1], [41, -1], [-1, -1]],
28    "temperature models":
29    [
30      {
31        "model": "linear", "min depth": 660e3, "max depth": 1160e3,
32        "top temperature": 1820, "bottom temperature": 2000
33      }
34    ],
35    "composition models": [{"model": "uniform", "compositions": [5]}]
36  },
37  // defining the oceanic plate
38  {
39    "model": "oceanic plate", "name": "oceanic plate",
40    "coordinates": [[-1, -1], [-1, 41], [15, 41], [15, 20], [5, 10], [5, -1]],
41    "temperature models":
42    [{"model": "linear", "max depth": 95e3, "bottom temperature": 1600}],
43    "composition models":
44    [
45      {"model": "uniform", "compositions": [0], "max depth": 10e3},
46      {"model": "uniform", "compositions": [1], "min depth": 10e3,
47       "max depth": 95e3}
48    ]
49  },
50  // defining the continental plate
51  {
52    "model": "continental plate", "name": "continental plate",
53    "coordinates": [[41, 41], [15, 41], [15, 20], [5, 10], [5, -1], [41, -1]],
54    "temperature models": [{"model": "linear", "max depth": 120e3,
55     "bottom temperature": 1600}],
56    "composition models":
57    [
58      {"model": "uniform", "compositions": [2], "max depth": 30e3},
59      {"model": "uniform", "compositions": [3], "min depth": 30e3,
60       "max depth": 120e3}
61    ]
62  },
63  // defining the subducting plate
64  {
65    "model": "subducting plate", "name": "Subducting plate",
66    "coordinates": [[15, 41], [15, 25], [5, 5], [5, -1]], "dip point": [20, 0],
67    "segments": [{"length": 200e3, "thickness": [95e3], "angle": [0, 45]},
68     {"length": 400e3, "thickness": [95e3], "angle": [45]},
69     {"length": 200e3, "thickness": [95e3], "angle": [45, 0]},
70     {"length": 100e3, "thickness": [95e3], "angle": [0]}],
71    "temperature models":
```

```
72 [{"model": "plate model", "density": 3300, "plate velocity": 0.05 }],
73   "composition models":
74   [
75     {"model": "uniform", "compositions": [0], "max distance slab top": 10e3},
76     {"model": "uniform", "compositions": [1], "min distance slab top": 10e3}
77   ]
78 }
79 ]
80 }
```

Listing 3.4: 3d subduction spreading example input file. The lines of green text (preceded by the double forward slashes) are comments and have no effect on the result.

3.11 SEPRAN 2D subduction

```
1 {
2   "version": "0.2",
3   "cross section": [[0,0],[100,0]],
4   "features":
5   [
6     // defining an oceanic plate on the left side of the model
7     {
8       "model": "oceanic plate", "name": "oceanic plate", "max depth": 95e3,
9       "coordinates": [[-1e3, -1e3], [1000e3, -1e3], [1000e3, 1e3], [-1e3, 1e3]],
10      "temperature models":
11      [
12        {
13          "model": "plate model", "max depth": 95e3, "bottom temperature": 1600,
14          "spreading velocity": 0.01,
15          "ridge coordinates": [[100e3, -1e3], [0e3, 1e3]]
16        }
17      ],
18      "composition models":
19      [
20        {"model": "uniform", "compositions": [0], "max depth": 10e3}
21      ]
22    },
23    // defining a weakzone oceanic plate at the first 100 km
24    {
25      "model": "oceanic plate", "name": "weak zone left", "max depth": 95e3,
26      "coordinates": [[-1e3, -1e3], [100e3, -1e3], [100e3, 1e3], [-1e3, 1e3]],
27      "temperature models":
28      [
29        {
30          "model": "linear", "max depth": 95e3, "bottom temperature": 1600,
31          "top temperature": 1573
32        }
33      ]
34    },
35    // defining a continental plate at the right side of the model
36    {
37      "model": "continental plate", "name": "continental plate", "max depth": 95e3
38      ,
39      "coordinates": [[1000e3, -1e3], [2001e3, -1e3], [2001e3, 1e3], [1000e3, 1e3]],
40      "temperature models":
41      [
42        {"model": "linear", "max depth": 95e3, "bottom temperature": 1600}
43      ]
44    },
45    // defining an oceanic plate as weakzone at the rightmost side of the model
46    {
47      "model": "oceanic plate", "name": "weak zone right", "max depth": 95e3,
48      "coordinates": [[1900e3, -1e3], [2000e3, -1e3], [2000e3, 1e3], [1900e3, 1e3]],
49      "temperature models":
```


3.11. SEPRAN 2D subduction

```
49 [
50   {
51     "model": "linear", "max depth": 95e3, "bottom temperature": 1600,
52     "top temperature": 1573
53   }
54 ]
55 },
56 // defining the upper mantle
57 {
58   "model": "mantle layer", "name": "upper mantle",
59   "min depth": 95e3, "max depth": 660e3,
60   "coordinates": [[-1e3, -1e3], [2001e3, -1e3], [2001e3, 1e3], [-1e3, 1e3]],
61   "temperature models":
62   [
63     {"model": "linear", "max depth": 660e3,
64      "top temperature": 1600, "bottom temperature": 1820}
65   ]
66 },
67 // defining the lower mantle
68 {
69   "model": "mantle layer", "name": "lower mantle",
70   "min depth": 660e3, "max depth": 1160e3,
71   "coordinates": [[-1e3, -1e3], [2001e3, -1e3], [2001e3, 1e3], [-1e3, 1e3]],
72   "temperature models":
73   [
74     {"model": "linear", "max depth": 1160e3,
75      "top temperature": 1820, "bottom temperature": 2000}
76   ]
77 },
78 // defining the subducting plate
79 {
80   "model": "subducting plate", "name": "Subducting plate",
81   "coordinates": [[1000e3, -1e3], [1000e3, 1e3]], "dip point": [2000e3, 0],
82   "segments":
83   [
84     {"length": 200e3, "thickness": [95e3], "angle": [0, 45]},
85     {"length": 200e3, "thickness": [95e3], "angle": [45]}
86   ],
87   "temperature models":
88   [
89     {"model": "plate model", "density": 3300, "plate velocity": 0.01 }
90   ],
91   "composition models":
92   [
93     {"model": "uniform", "compositions": [0], "max distance slab top": 10e3}
94   ]
95 },
96 // defining a continental plate on top of the slab to force 293.15 K at
97 // the surface near the slab
98 {
99   "model": "continental plate", "name": "top on slab", "max depth": 1,
100  "coordinates": [[900e3, -1e3], [1100e3, -1e3], [1100e3, 1e3], [900e3, 1e3]],
101  "temperature models": [{"model": "uniform", "temperature": 293.15}]
102 }
103 ]
104 }
```

Listing 3.5: 2d SEPRAN subduction example input file. The lines of green text (preceded by the double forward slashes) are comments and have no effect on the result.

3.12 ELEFANT 3D Double subduction setup

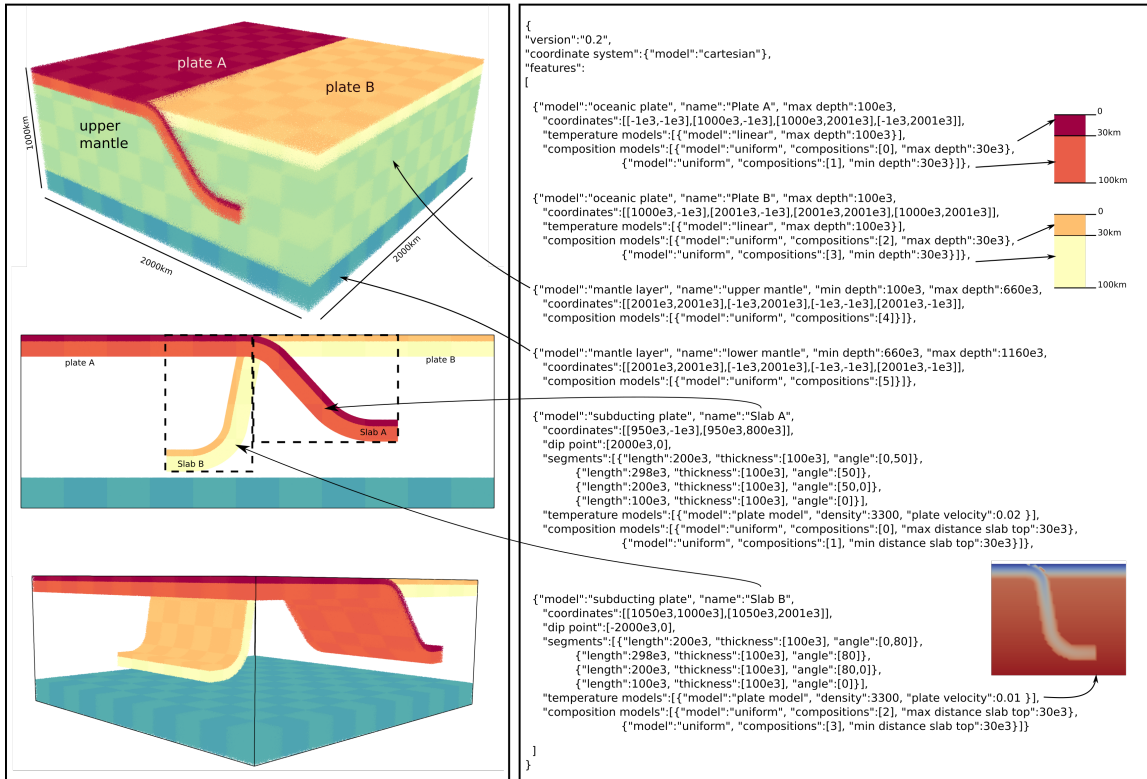


Figure 3.9: Connection between the GWB input file (right panel) and the resulting marker fields (left panel). Small upper inserts in the right panel show each plate layering while the bottom insert shows the temperature field zoomed in on slab B.

3.13 ASPECT 3d curved subduction

```

1  {
2  "version":"0.2",
3  "potential_mantle_temperature":1500,
4  "thermal_expansion_coefficient":2.0e-5,
5  "maximum_distance_between_coordinates":100000,
6  "interpolation":"monotone_spline",
7  "surface_temperature":293.15,
8  "force_surface_temperature":true,
9  "coordinate_system":{"model":"cartesian"},
10 "features":
11 [
12   // defining an oceanic plate for the North and South American plate
13   {"model":"oceanic_plate", "name":"NS American plate",
14    "coordinates":[[1700e3,0],[1700e3,300e3],[1606e3,650e3],
15                  [1350e3,906e3],[1000e3,1000e3],[-1e3,1000e3],
16                  [-1e3,1501e3],[2501e3,1501e3],[2501e3,-501e3],
17                  [-1e3,-501e3],[-1e3,-50e3],[2000e3,-50e3],
18                  [2000e3,0e3]],
19    "temperature_models":[{"model":"linear", "max_depth":100e3}],
20    "composition_models":[{"model":"uniform", "compositions":[0],
21                          "max_depth":30e3}]},
22
23   // Defining an oceanic plate for the Caribbean plate
24   {"model":"oceanic_plate", "name":"Caribbean plate",
25    "coordinates":[[1700e3,300e3],[1689e3,422e3],[1658e3,539e3],
26                  [1606e3,650e3],[1536e3,749e3],[1450e3,836e3],
27                  [1350e3,906e3],[1239e3,958e3],[1122e3,989e3],
28                  [1000e3,1000e3],[650e3,1000e3],[-1e3,1000e3],
29                  [-1e3,0e3],[1700e3,0e3]],
30    "temperature_models":[{"model":"linear", "max_depth":100e3}],
31    "composition_models":[{"model":"uniform", "compositions":[1],
32                          "max_depth":30e3}]},
33
34   // Defining a continental plate for the weak zone
35   {"model":"continental_plate", "name":"Caribbean weak zone",
36    "coordinates":[[-1e3,1000e3],[-1e3,750e3],[1536e3,749e3],
37                  [1450e3,836e3],[1350e3,906e3],[1239e3,958e3],
38                  [1122e3,989e3],[1000e3,1000e3],[650e3,1000e3]],
39    "temperature_models":[{"model":"linear", "max_depth":100e3}],
40    "composition_models":[{"model":"uniform", "compositions":[2],
41                          "max_depth":30e3},
42                          {"model":"uniform", "compositions":[3],
43                          "min_depth":30e3}]},
44
45   // Defining a mantle layer for the lower mantle
46   {"model":"mantle_layer", "name":"660", "min_depth":660e3,
47    "coordinates":[[-1e3,-500e3],[-501e3,2500e3],[2501e3,2500e3],
48                  [2501e3,-501e3]],
49    "composition_models":[{"model":"uniform", "compositions":[4]}}],
50
51   // Defining a subducting plate for the Lesser Antilles slab
52   {"model":"subducting_plate", "name":"Lesser Antilles slab",
53    "coordinates":[[1700e3,0],[1700e3,300e3],[1606e3,650e3],
54                  [1350e3,906e3],[1000e3,1000e3],[650e3,1000e3]],
55    "dip_point":[-1,-1],
56    "min_depth":0, "max_depth":660e3,
57    "segments":
58    [
59      {"length":300e3, "thickness":[100e3], "angle":[0,50]},
60      {"length":371e3, "thickness":[100e3], "angle":[50]},
61      {"length":275e3, "thickness":[100e3], "angle":[50,0]},
62      {"length":0e3, "thickness":[100e3], "angle":[0]}
63    ],

```

```
64     "sections":
65     [
66     { "coorindate": "0",
67       "segments":
68       [
69         { "length": 300e3, "thickness": [100e3], "angle": [0,25]},
70         { "length": 371e3, "thickness": [100e3], "angle": [50]},
71         { "length": 300e3, "thickness": [100e3], "angle": [50,0]},
72         { "length": 50, "thickness": [100e3], "angle": [0]}
73       ]
74     },
75     { "coorindate": "5",
76       "segments":
77       [
78         { "length": 300e3, "thickness": [100e3], "angle": [0,25]},
79         { "length": 371e3, "thickness": [100e3], "angle": [50]},
80         { "length": 50e3, "thickness": [100e3], "angle": [50,0]},
81         { "length": 0, "thickness": [100e3], "angle": [0]}
82       ]
83     }
84   ],
85   "temperature models":
86   [
87     { "model": "plate model", "density": 3300, "plate velocity": 0.0144,
88       "thermal conductivity": 2.5, "thermal expansion coefficient": 2e-5 }
89   ],
90   "composition models":
91   [
92     { "model": "uniform", "compositions": [0], "min distance slab top": 30e3}
93   ]
94 },
95 // Defining a continental plate for the weakzone between the Caribbean and
96 // South America
97 {"model": "continental plate", "name": "South Weakzone",
98   "coordinates": [[-1e3, 0e3], [-1e3, -50e3], [2000e3, -50e3], [2000e3, 0e3]],
99   "temperature models": [{"model": "linear", "max depth": 100e3}],
100  "composition models":
101  [
102    { "model": "uniform", "compositions": [2], "max depth": 30e3},
103    { "model": "uniform", "compositions": [3], "min depth": 30e3}}
104  ]
105 ]
106 }
107 }
```

Listing 3.6: Input for the ASPECT example. The lines of green text (preceded by the double forward slashes) are comments and have no effect on the result.

4

Assessing the geodynamics of strongly arcuate subduction zones: the eastern Caribbean subduction setting.

4.1 Summary

In this chapter I build on the accomplishments of Chapters 2 and 3 to create a next step in numerical simulating the geodynamic evolution of natural subduction. The target natural system is that of strongly arcuate subduction which is common on Earth. The eastern Caribbean Lesser-Antilles subduction system is taken as the example system for 3D numerical simulation. I demonstrate for this arcuate subduction system the geodynamical feasibility of westward directed trench-parallel slab transport through the mantle, i.e. slab dragging by North-American plate motion, of the northern slab segment while the eastern slab is subducting at a mantle-stationary trench. The resistance of the ambient mantle against slab dragging as well as the lateral deformation of the arcuate slab create a complex 3-D stress field in the slab that deviates strongly from the classical view of slab-dip aligned orientation of slab stress. Slab dragging thus may reveal itself in the focal mechanisms of intermediate and deep earthquakes. These characteristics of arcuate subduction are generic and may as well apply to other arcuate subduction systems, such as Izu-Bonin-Marianas or the Aleutians-Alaska systems, where anomalous focal mechanisms of slab events are observed. This chapter demonstrates the developed capabilities and computational feasibility towards assessing the 3D-complexity and geodynamics of natural subduction systems.

4.2 Introduction

Numerical three-dimensional (3D) thermo-mechanical modelling provides a crucial means for investigating the geodynamic evolution of plate convergence involving lithosphere subduction. Particularly, subduction modelling facilitates a way for creating quantitative links between crust-mantle processes and the various surface observations of plate convergence zones obtained from geological, seismological, or geodetic methods. Clearly, this requires the capability of numerical simulation of natural subduction systems that are characterized by complex tectonic histories and 3D slab evolution. Advances in global mantle convection modelling have shown that plate motions and lithosphere subduction can be sustained in a dynamically self-consistent way (e.g.

Assessing the geodynamics of strongly arcuate subduction zones: the eastern Caribbean subduction setting., M. Fraters, W. Spakman, C. Thieulot and D. van Hinsbergen, in preparation for publication.

Cramer and Tackley, 2015; Coltice and Shephard, 2017), but creating useful tectonic predictions for natural subduction is still difficult (Coltice and Shephard, 2017). When subduction modelling is restricted to a regional modelling domain, it is easier to control governing plate motions through boundary conditions and impose initial model conditions derived from (paleo-)tectonic settings. Most investigations based on a regional model domain define an initial 3D model in which plate boundaries are either parallel or perpendicular to the rectangular boundaries (e.g. Stegman et al., 2006; Yamato et al., 2009; Burkett and Billen, 2010; Stegman et al., 2010a; van Hunen and Allen, 2011; Capitanio et al., 2011; Moresi et al., 2014; Schellart and Moresi, 2013; Duretz et al., 2014; Sternai et al., 2014; Magni et al., 2014; Pusok and Kaus, 2015; Capitanio et al., 2015; Kiraly et al., 2016; Sternai et al., 2016; Schellart, 2017; Pusok et al., 2018; Chertova et al., 2018; Gülcher et al., 2019). In this setup, subduction develops from an initially straight trench with trench-perpendicular plate convergence and is controlled by the trench-perpendicular forcing of slab pull and convergent plate motions. Much less studies have attempted to simulate natural subduction by implementing constraints from the first-order inferences of, e.g., governing trench-oblique plate motions, lithosphere structure and trench geometry, or from plate-tectonic reconstructions (e.g. Billen and Gurnis, 2003; Liu and Stegman, 2011; Jadamec and Billen, 2010; Jadamec et al., 2013; Malatesta et al., 2013; Chertova et al., 2014b, 2018; Hu and Liu, 2016). This class of subduction modelling, however, entails resolving the practical difficulty of constructing geometrically complex 3D initial models for the paleogeographic setting of incipient subduction or for an advanced stage of subduction evolution.

The research we report here is in the realm of the latter modelling approach and assesses the geodynamics of natural subduction zones that are characterized by strongly trench-oblique plate convergence. We focus our investigation on the geodynamical processes involved in strongly arcuate subduction. Prominent examples of such subduction systems are the major Aleutian-Alaska, Izu-Bonin-Marianas, or Andaman-Sumatra subduction zones. These are characterized by large absolute plate motions of the subducting plate at the trench, ranging between 70-110 mm/yr, and by directions of plate convergence that vary from near trench-perpendicular (at the Alaska, Marianas, and east-Sumatra trenches, respectively) to strongly trench-oblique with departures from trench-perpendicular convergence by 60° to even 90° (e.g. Fig. S4 of Doubrovine et al. (2012)). Much smaller arcuate subduction systems with significant trench-parallel components of subducting plate motion occur in the Mediterranean region. These systems involve low absolute plate motion of the

subducting African plate to the NNE that varies between ~ 8 mm/yr for the east-dipping Gibraltar subduction in the western Mediterranean, to ~ 10 mm/yr for the northwest-dipping Calabria subduction of the Central Mediterranean, to ~ 12 mm/yr for NNE-dipping Aegean subduction of the Eastern Mediterranean.

The Lesser Antilles subduction of the eastern Caribbean region (Fig. 4.1A) is a strongly arcuate subduction system that is similar in geometry with the three major systems while it involves relatively small absolute motion of the subducting North and South American plates of at present ~ 20 mm/yr in westward direction. The plate-tectonic setting, geometry, and evolution of the Lesser Antilles subduction system are used here as a guide for our assessment of the geodynamics of arcuate subduction that involves low subducting-plate motion. Because it is unknown how arcuate subduction initializes, we focus our research here on taking a more advanced evolution stage of established arcuate subduction as the starting point for modeling for which we create an initial 3D model using our recently developed 'Geodynamic World Builder' (GWB; version 0.1.0) (Fraters et al. 2019b, submitted to *Solid Earth*; Chapter 3). The GWB is a versatile open-source software tool for the relatively easy construction of geometrically intricate initial temperature-composition models for geodynamic modelling. It allows construction of complex 3D subduction geometries as a starting point for generic modelling, for testing plate-tectonic evolution scenarios, for investigating the evolution of natural subduction, or for the instantaneous dynamics of present-day subduction.

We employ the GWB to construct an approximation of the 3D geometry of the 3D Lesser Antilles subduction zone at ~ 10 Ma including overriding and subducting plates. Our aim is two-fold. First, we experiment with various rheology and temperature settings of the initial model enabling local exploration of the model space for the style of subduction evolution. Particularly, we aim to obtain parameter settings for crust-mantle rheology such that a more or less mantle-stationary eastern Lesser Antilles trench is obtained as is suggested by plate reconstructions in several mantle reference frames (Müller et al., 1999; Boschman et al., 2014). Second, we investigate the hypothesis that, while the eastern Lesser Antilles slab is mantle-stationary, the northern limb of the Lesser Antilles subduction is subject to strong near trench-parallel lateral transport of the slab by the absolute motion of the North American plate, which is an example of slab dragging (Spakman et al., 2018). Our main purpose is to investigate how such an arcuate slab system develops from the interplay between 1) slab dragging forced by the absolute motion of the North and South American plates, 2) the almost not moving overriding Caribbean plate, 3) the vertically-

4.3. Three-dimensional initial model of arcuate subduction based on the tectonic setting of the Lesser Antilles slab

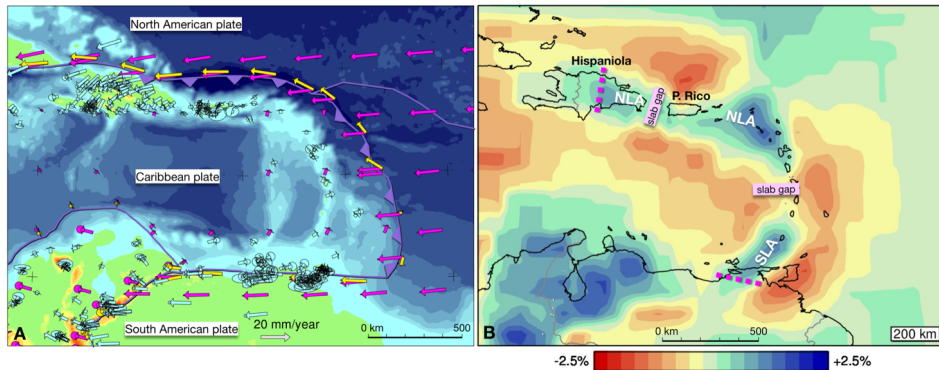


Figure 4.1: **Plate tectonic setting of the Eastern Caribbean region.** **A:** Purple lines: plate boundaries from MORVEL56 (Argus et al., 2011) delineating the Caribbean plate, the LA trench (barbed), and the North and South American plate boundary; magenta vectors: present-day plate motions in the mantle reference frame of Doubrovine et al. (2012); yellow vectors: trench-parallel component of North and South American plate motion; blue vectors: GPS motions (Kreemer et al., 2014) plotted in the same mantle reference frame. Background colors: ETOPO1 bathymetry (Amante and Eakins, 2009). **B:** Tomographic section through model UU-P07 (Hall and Spakman, 2015; van der Meer et al., 2018) at a depth of 200 km. NLA and SLA denote the southern and northern Lesser Antilles slab. Two slab gaps are indicated. Dashed lines: tomography-based interpreted slab edges (van Benthem et al., 2014; Harris et al., 2018)

directed slab pull, and 4) the viscous coupling of the slab to the ambient mantle. Particularly, we explore the role of slab dragging in arcuate subduction involving slowly moving subducting and overriding plates. Our work focuses in a generic way on the acting processes and serves as a stepping stone towards more elaborate numerical simulations of arcuate subduction, e.g., involving lateral lithosphere heterogeneity and (inherited) lithosphere weakness, such that future models can come closer to prediction of surface observations.

4.3 Three-dimensional initial model of arcuate subduction based on the tectonic setting of the Lesser Antilles slab

The Lesser Antilles (LA) subduction zone comprises a west-dipping slab under the eastern Lesser Antilles island arc that curves laterally by about 90° into a south dipping slab along the northern Hispaniola-Puerto Rico trench (Fig. 4.1A). Under the eastern LA arc both the North-American (NAM) and South-American (SAM) lithosphere subducts. The plate boundary between SAM and NAM is poorly defined. Fig. 4.1A shows the boundary from Argus et al. (2011)

while interpretations of seismic tomography suggest it coincides with a gap in the slab that is imaged to depths of 200-300 km separating the southern Lesser Antilles slab (SLA) from the northern Lesser Antilles (NLA) slab (Fig. 4.1B; van Benthem et al., 2013; Harris et al., 2018). Another slab gap between Hispaniola and Puerto Rico (Fig. 4.1B) was proposed by Harris et al. (2018), while Meighan et al. (2013) suggest vertical slab segmentation to the east of Puerto Rico. The arcuate slab geometry is well delineated by slab seismicity to depths of ~200 km (Harris et al., 2018) and in the upper mantle by seismic tomography van Benthem et al. (2013); Harris et al. (2018).

In a plate motion frame relative to the Caribbean plate, the NAM plate converges at present highly oblique at an angle of ~20° with the strike of the northern trench. This suggests a component of westward trench-parallel motion of the NLA slab with respect to the Caribbean plate (van Benthem et al., 2014). Investigating actual slab motion through the mantle requires a mantle frame of reference (e.g. Doubrovine et al. (2012)) to define the absolute plate motion (APM) of the NAM and SAM subducting plates and that of the overriding Caribbean plate. Along the eastern trench, the APM of the NAM and SAM plates is at present largely trench-perpendicular in the south and increasingly trench-oblique when going to the north, while the NAM plate moves near trench-parallel along the northern trench (Fig. 4.1A). The Caribbean plate shows at present little motion in the mantle frame (Fig. 4.1A) while the small GPS motions of the LA islands above the eastern LA slab imply that the eastern LA trench may also be about mantle-stationary. This plate-tectonic setting in the mantle reference frame suggests that the NLA slab is at present being transported almost trench-parallel through the upper mantle by the APM of the NAM plate whereas the eastern LA slab is subducting at a mantle-stationary trench.

The trench-parallel slab motion is an example of slab dragging, i.e., 'the process of lateral slab transport through the mantle that is forced by the absolute surface motion of the subducting plate' (Spakman et al., 2018). Slab dragging at a rate of 6-7 mm/yr has recently been documented for the Gibraltar slab of the western Mediterranean (Spakman et al. 2018), and at a trench-parallel rate of about 30 mm/yr for the Tonga-Kermadec slab (van de Lagemaat et al., 2018). The actual occurrence and effects of slab dragging on slab deformation during arcuate subduction are here important targets of investigation.

We employ the GWB to reconstruct the 3D subduction zone geometry for setting up an initial geometrical model that simulates 3D slab morphology at 10 Ma and which implements overriding and subducting plates. By doing so

4.3. Three-dimensional initial model of arcuate subduction based on the tectonic setting of the Lesser Antilles slab

we avoid defining the poorly constrained paleogeographic and plate tectonic setting of subduction initiation in the Paleogene (e.g. Pindell and Kennan, 2009; Boschman et al., 2014). Another advantage is that we can reduce the computation time for such complex subduction models. The Caribbean plate has been more or less mantle-stationary since the Paleogene (Müller et al., 1999; Boschman et al., 2014). Assuming that the eastern LA trench has also been mantle-stationary, the simplest reconstruction for 10 Myr ago requires shifting the NAM plate, and thus the northern slab edge, to the east. Given the large uncertainty of absolute plate motions (Doubrovine et al., 2012), we assume a mantle-stationary Caribbean plate and use the more accurate relative plate motion between the Caribbean and NAM for NAM of ~ 15 mm/yr during the past 10 Myr (Boschman et al., 2014). We adopt this value for the west-directed APM of both NAM and SAM. This choice keeps our approximation of the Lesser Antilles subduction close enough to the actual system while our generic modelling can focus on assessing the dynamics of arcuate subduction at low APM of the subducting plate and close to the African APM involved in arcuate Mediterranean subduction systems.

Effectively, for the construction of the slab geometry at 10 Ma we estimate the first-order geometrical characteristics (dip, length, curvature) of the tomographically imaged slab of Harris et al. (2018) and remove 150 km from the western part of the slab. This leads to the slab geometry of Fig. 4.2. Importantly, the 150 km eastward shift of the slab at 10 Ma is just an assumption and provides a starting geometry. Modelling forward in time, the slab evolution may show if the resulting slab geometry at 0 Ma conforms to the tomographic observation. For instance, this would provide a primary test of the hypothesis that the NLA slab is being entirely dragged westward by the APM of the NAM plate, instead of only partially, or not significantly.

A Cartesian model box is adopted with X-Y-Z dimensions of $2500 \text{ km} \times 2000 \text{ km} \times 800 \text{ km}$. The initial model is built from six different GWB-features which are assigned GWB-compositions for identification. The background temperature is defined by the adiabatic geotherm $\theta_S \exp(\alpha g d / C_p)$, where θ_S is the potential temperature of the mantle (1500 K), α is the thermal expansion coefficient ($2 \times 10^{-5} \text{ K}^{-1}$), g is the gravitational acceleration (m/s^2), C_p is the specific heat ($1250 \text{ J} \cdot \text{kg}^{-1} \text{ K}^{-1}$) and d is the depth. When no composition label is defined, the background composition is assumed, i.e. the mantle. This holds for instance for the lithospheric mantle of the tectonic plates. The initial model consists of the following parts:

- The North and South American plate, which are combined here into one

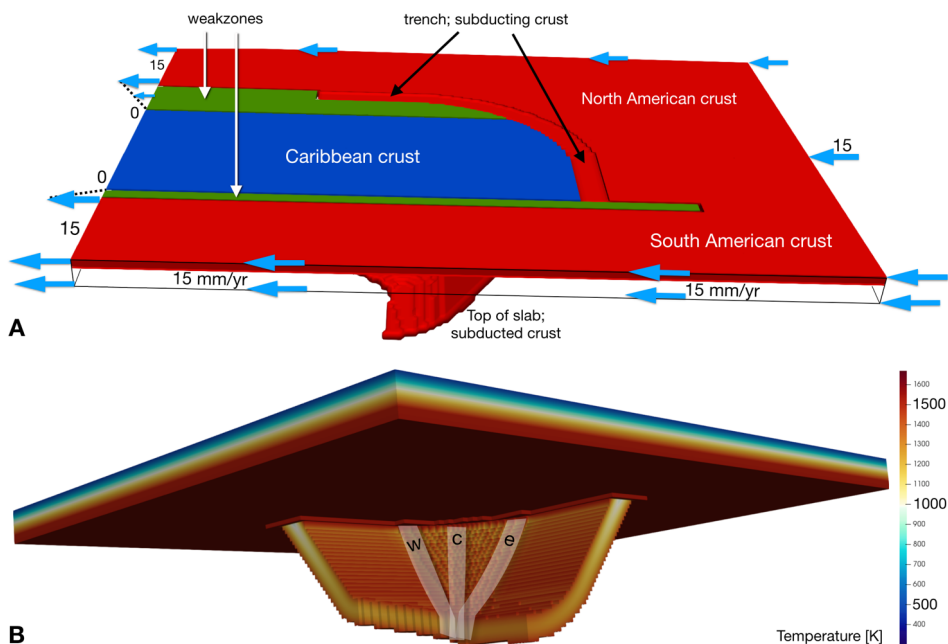


Figure 4.2: **Initial subduction model geometry and velocity boundary conditions.** **A:** Regions composing the crust. The thickness of the Caribbean, NAM and SAM crust is 30 km and its rheology is tunable separately from the lithospheric mantle. Weakness zones of tunable rheology provide visco-plastic decoupling between the Caribbean and NAM and SAM plates. The lithospheric mantle is configured as part of the mantle and shown in B. The NAM and SAM lithosphere obtain uniform west pointing plate velocity which is imposed on the vertical model edges and across the lithosphere thickness of 100 km. The top of the model is a dynamic free surface. **B:** Temperature view of the 3D lithosphere starting model. Temperature in the lithosphere increases linearly. In the slab, a temperature field is shown computed from the analytical solution provided by McKenzie (1970). Slabs with different initial temperature are used in experiments. The three transparent zones plotted on the slab and labeled w(est),c(entral), and e(ast), denote locations of *a priori* inserted vertical slab weakness zones used for experimenting.

4.3. Three-dimensional initial model of arcuate subduction based on the tectonic setting of the Lesser Antilles slab

plate, North-South American (NSAM) plate, is given a thickness of 100 km and a linear temperature profile from the surface temperature (293 K) to the adiabatic temperature at the depth of 100 km. The crustal layer is represented by compositional layer 0 and is 30 km in thickness.

- The Caribbean plate only differs from the North-South American plate by its compositional label 1 for the crust.
- The LA trench together with the subducting slab is the next GWB-feature. The trench is defined by 12 trench points, interpolated by a monotone spline, that effectively describe 3 main lateral trench sectors: the eastern trench, the curved trench, and the more or less EW-striking northern trench (Fig. 4.2A). In each trench point the slab is parameterized by 4 depth segments along a vertical section. The first 3 segments obtain the same dip variation while the 4th, deepest segment, varies in geometrical properties along the trench from SE to NW. The thickness of all slab segments is equal to the thickness of the subducting North-South American plate, i.e. 100 km. Laterally, the geometrical properties are linearly interpolated between the 12 vertical trench-slices. The first slab segment has a length of 150 km and starting-dip angle of 0° , i.e. parallel with the surface, which increases linearly with depth to 25° . The second segment, also of 150 km, starts with a dip angle of 25° which linearly increases along the segment to 50° . The third segment has a length of 171 km and dips uniformly at 50° . The last segment is of variable length but the dip angle in each trench slice changes linearly from 50° to 0° . In the eastern main slab sector, from the SE slab edge to the beginning of the slab bend, the length of the fourth depth segment varies from 300 km to 275 km. In the bend-sector, the length is 275 km and after the bend, i.e. the northern slab sector, the length varies from 275 to 50 km. The temperature in the plate is defined by the McKenzie (1970) temperature model and the top 30 km of the slab has compositional label 0. The southern sector has the largest slab length of 875 km and reaches the depth of 660 km at the transition to the lower mantle. In 10 Myr of subduction evolution the length would increase to ~ 1025 km which underestimates the ~ 1100 km of subduction since ~ 45 Ma (Boschman et al., 2014). This is done to avoid too strong coupling between slab and lower mantle in all numerical models. Effects of possible deep upper-mantle anchoring of the slab (e.g. Uyeda and Kanamori, 1979; Heuret and Lallemand, 2005) will be tested in experiments with an increased viscosity in the deeper upper mantle.

- To the west of the northern slab edge under Hispaniola, a 100 km thick and 250 km wide lithosphere plate is defined with the same linear geotherm as in the adjacent plates. The zone is extended to the east where it partly replaces the Caribbean plate (top green zone of Fig. 4.2A). To west of the slab edge, this continental zone will serve as a weak zone decoupling the North American and Caribbean plates whereas it defines a weak upper plate above the NLA slab. The continental segment is given a separate composition label 2 for the crust and composition label 3 for the deeper lithosphere, which enables tuning the lithosphere rheology separately from the adjacent plates.
- A similar lithospheric weak zone of thickness 100 km and width 50 km decouples the South American plate from the Caribbean plate and the slab. Composition labels are the same as for the northern weak zone.
- The lower mantle below 660 km is separately defined with composition label 4.

To test the effects of lateral rheological slab heterogeneity, e.g. subducted inherited weak zones, vertical segments of slab have been assigned a separate compositional label 2 for the crustal part and label 3 for the lithospheric part of the slab, such that the rheology can be assigned to be weaker than the surrounding slab. Weak zones can be placed in three locations (Fig. 4.2B).

Depending on the compositional label, all model features can be assigned specific material rheology. Rheology and slab temperature will be defined in each numerical experiment leading to a complete definition of the initial conditions.

The details of how the model is setup through the GWB can be found in appendix 4.7.

4.4 Model setup

4.4.1 Numerical model setup

Modelling the dynamic evolution of subduction requires numerically solving the governing equations of continuum mechanics for which modern, open source, parallel geodynamics codes are available, such as pTatin (May et al., 2015), LaMem (Kaus et al., 2016), Citcom(S) Zhong et al. (2000), Underworld 2 (Mansour et al., 2019), or ASPECT (Kronbichler et al., 2012; Heister et al., 2017). In this thesis we use version 2.1.0-pre of ASPECT with some unmerged but proposed changes at the time of writing. The most significant changes are three

added plugins. The first plugin is a boundary temperature model which prescribes the initial temperature to all boundaries, except for the surface, where a fixed temperature is set. Secondly a mesh refinement model which allows to set minimum and maximum refinement levels between a pre-set minimum and maximum temperature. Thirdly, a postprocessor which computes the maximum shear stress. This is computed by first computing the eigenvalues and eigenvectors of the stress tensor. The direction of the maximum shear is then computed by subtracting the smallest eigenvector from the largest eigenvector. The magnitude is computed by subtracting smallest eigenvalue by the largest eigenvalue and dividing it by two. The exact setup can be found in appendix 4.8, which shows the reference ASPECT input file.

Governing equations

The mass, momentum and energy conservation equations are those of a fluid under the Boussinesq approximation (Schubert et al., 2001):

$$-\nabla \cdot (2\eta\boldsymbol{\varepsilon}(\mathbf{u})) + \nabla p = \rho\mathbf{g} \quad \text{in } \Omega, \quad (4.1)$$

$$\nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega, \quad (4.2)$$

$$\rho_0 C_p \left(\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T \right) - \nabla \cdot k \nabla T = \rho H \quad \text{in } \Omega, \quad (4.3)$$

$$\frac{\partial c_i}{\partial t} + \mathbf{u} \cdot \nabla c_i = q_i \quad \text{in } \Omega, i = 1 \dots C \quad (4.4)$$

where \mathbf{u} is the velocity, p the pressure and T the temperature. The set of advected quantities c_i are called *compositional fields* and are used to track materials. The coefficient are as follows: η is the viscosity, ρ the density, \mathbf{g} the gravity vector and $\boldsymbol{\varepsilon}(\cdot)$ denotes the (deviatoric) symmetric gradient operator defined by $\boldsymbol{\varepsilon}(\mathbf{v}) = \frac{1}{2}(\nabla\mathbf{v} + \nabla\mathbf{v}^\top)$, C_p is the specific heat, k the thermal conductivity and H is the internal heat production. Note that all coefficients that appear in these equations can depend on the pressure, temperature, chemical composition, and the strain rate, thereby making these equations potentially nonlinear.

In our modelling we have used the default Finite Element type of ASPECT, i.e. the $Q_2 \times Q_1$ element (Gresho and Sani, 2000). Details of the Finite Element discretization of the above equations can be found in Chapter 2 as well as in Kronbichler et al. (2012); Heister et al. (2017). All simulations were performed with the stabilized Newton solver, as described in Chapter 2, and run in parallel on 480 threads on five nodes made from AMD EPYC 7451 24-Core Processors. Depending on the chosen set of rheological parameters simulations took between 3 and 6 days to simulate 10 million years. In the reference model,

the number of elements varies through time between 560000 and 660000. The number of Stokes degrees of freedom (dofs) vary between 15 and 18 million, the number of temperature dofs between 4.7 and 5.7 million and the total number of composition dofs between 23 and 28 million over a total simulation time of 20 million years.

4.4.2 Boundary conditions

The 6 geometrical boundaries of the regional model box comprise ten different boundary domains for defining boundary conditions: a top and bottom boundary, four side boundaries above 100 km depth and four side boundaries below 100 km depths. The top boundary is a (stabilized) free surface (Rose et al., 2017; Kaus et al., 2010; Quinquis et al., 2011). On the bottom boundary a zero velocity boundary condition (no-slip) is defined. The four bottom side boundaries are open boundaries (Chertova et al., 2012) which were implemented in ASPECT by Anne Glerum (Glerum et al., 2018). The eastern and western top side boundaries are assigned velocity boundary condition (Dirichlet type) of 1.5 cm/yr (or 3.0 cm/yr in some experiments), uniformly pointing to the west in most of the domain. At the Caribbean plate the boundary velocity is set to zero, and in the weak zones the velocity linearly increases from zero at the side of the Caribbean plate to 1.5 (or 3.0 cm/yr) in the NSAM plate. To compensate for the lower outflow in the southern weak zone, there is slightly less inflow on the south-east of the model. The north and south boundaries have for the x and y component an open boundary condition, where the z component is fixed to zero velocity. Temperature boundary conditions are set along the sides and bottom according to the initial model. For details on the boundary conditions, see the ASPECT input file in section 4.8.

4.4.3 Rheological model

For the rheology we use a composite viscosity model consisting of three effective viscosities computed for each composition i as follows:

- the effective dislocation creep viscosity:

$$\eta_i^{\text{dsl}} = \frac{1}{2} A^{-\frac{1}{n_i}} \dot{\epsilon}_i^{\frac{1-n_i}{n_i}} \exp\left(\frac{E_i^* + PV_i^*}{n_i RT}\right)$$

- the effective diffusion creep viscosity:

$$\eta_i^{\text{dff}} = \frac{1}{2} A^{-1} d^{\frac{m_i}{n_i}} \exp\left(\frac{E_i^* + PV_i^*}{RT}\right)$$

- the effective plastic viscosity:

$$\eta_i^{\text{pl}} = \frac{p \sin \phi_i + c_i \cos \phi_i}{\sqrt{3}(3 + \sin \phi_i) 2\dot{\epsilon}_i}$$

which is a commonly used formulation in geodynamics, based on the Drucker-Prager yield criterion (Glerum et al., 2018; Spiegelman et al., 2016).

In these equations $\dot{\epsilon}_i$ is the square root of the second invariant of the strain rate tensor, R is the gas constant, and P and T are the pressure and temperature. The full pressure is used in these equations, including in the plasticity, and not the lithostatic pressure as in Chertova et al. (2014a). Although using the full pressure in the plasticity has been shown to be potentially problematic for the solvers (Spiegelman et al., 2016), we obtained convergence to a nonlinear normalized tolerance level of 10^{-6} within the first few time steps.

We make use of the pre-existing "visco plastic" material model of ASPECT, which first separately computes the three viscosities and then combines them as follows:

1. a composite viscous viscosity is first computed through

$$\eta^{\text{visc}} = \left(\frac{1}{\eta_i^{\text{dff}}} + \frac{1}{\eta_i^{\text{dsl}}} \right)^{-1}$$

2. if the associated viscous stress $2\eta^{\text{visc}}\dot{\epsilon}_i$ exceeds the plastic yield strength then the plastic viscosity η^{pl} is used instead of η^{visc} ;
3. if more than one compositional field is present at a given point, viscosities are averaged with a harmonic scheme (Schmelting et al. (2008)).

Finally, in order to avoid too poorly conditioned matrices, the resulting effective viscosities η_{eff} are bounded such that:

$$\eta_{\text{min}} \leq \eta_{\text{eff}} \leq \eta_{\text{max}} \tag{4.5}$$

where η_{min} and η_{max} are user-chosen parameters that typically define a range of 5-6 orders of magnitude.

There are 5 materials in our model, each associated with density, thermal and rheological parameters. The rheological parameters of these materials are listed in Table 4.1.

Symbol	Meaning	Ref. value	Units
η_{\min}	Minimum viscosity	5e19	Pa·s
η_{\max}	Maximum viscosity	5e24	Pa·s
T_0	Ref. temp for whole model	293	K
α	Thermal expansion	2e-5	K ⁻¹
ρ	Density	3300	kg·m ⁻³
d	Grain size	0.01	m
m^{diff}	Diffusion creep grain size exponent	3	m
m^{dis}	Dislocation creep grain size exponent	0	m
C_p	Heat capacity	1.25e3	J/kg/K
k	Thermal diffusivity	0.8e-6	m ² /s
Background (upper mantle)			
A_b^{diff}	Diffusion creep prefactor	8.88e-15	Pa ^{-n_{dis}} s ⁻¹
Q_b^{diff}	Diffusion creep activation energy	335e3	J / mol
V_b^{diff}	Diffusion creep activation volume	5.0e-6	m ³ / mol
n_b^{dis}	Dislocation creep exponent	3.5	
A_b^{dis}	Dislocation creep prefactor	6.51e-15	Pa ^{-n_{dis}} s ⁻¹
Q_b^{dis}	Dislocation creep activation energy	530.e3	m ³ / mol
Q_b^{dis}	Dislocation creep activation volume	18e-6	m ³ / mol
c_b	Cohesion	20e6	MPa
ϕ_b	Angle of friction	15	°
composition 0 (NSA crust)			
A_0^{diff}	Diffusion creep prefactor	8.88e-15	Pa ^{-n_{dis}} s ⁻¹
Q_0^{diff}	Diffusion creep activation energy	375e3	J / mol
V_0^{diff}	Diffusion creep activation volume	6.0e-6	m ³ / mol
n_0^{dis}	Dislocation creep exponent	4.0	
A_0^{dis}	Dislocation creep prefactor	8.57e-28	Pa ^{-n_{dis}} s ⁻¹
Q_0^{dis}	Dislocation creep activation energy	167e3	m ³ / mol
Q_0^{dis}	Dislocation creep activation volume	36e-6	m ³ / mol
c_0	Cohesion	10e4	MPa
ϕ_0	Angle of friction	5	°
composition 1 (Caribbean crust)			
A_1^{diff}	Diffusion creep prefactor	8.88e-15	Pa ^{-n_{dis}} s ⁻¹
Q_1^{diff}	Diffusion creep activation energy	375e3	J / mol
V_1^{diff}	Diffusion creep activation volume	6.0e-6	m ³ / mol
n_1^{dis}	Dislocation creep exponent	4.0	
A_1^{dis}	Dislocation creep prefactor	8.57e-28	Pa ^{-n_{dis}} s ⁻¹

4.4. Model setup

Q_1^{dis}	Dislocation creep activation energy	223e3	m^3 / mol
Q_1^{dis}	Dislocation creep activation volume	18e-6	m^3 / mol
c_1	Cohesion	10e6	MPa
ϕ_1	Angle of friction	10	$^\circ$
composition 2 (weakzone crust)			
A_2^{diff}	Diffusion creep prefactor	8.88e-15	$\text{Pa}^{-n_{dis}} \text{s}^{-1}$
Q_2^{diff}	Diffusion creep activation energy	375e3	J / mol
V_2^{diff}	Diffusion creep activation volume	6.0e-6	m^3 / mol
n_2^{dis}	Dislocation creep exponent	4.0	
A_2^{dis}	Dislocation creep prefactor	8.57e-28	$\text{Pa}^{-n_{dis}} \text{s}^{-1}$
Q_2^{dis}	Dislocation creep activation energy	167e3	m^3 / mol
Q_2^{dis}	Dislocation creep activation volume	18e-6	m^3 / mol
c_2	Cohesion	10e4	MPa
ϕ_2	Angle of friction	5	$^\circ$
composition 3 (weakzone lithosphere)			
A_3^{diff}	Diffusion creep prefactor	8.88e-15	$\text{Pa}^{-n_{dis}} \text{s}^{-1}$
Q_3^{diff}	Diffusion creep activation energy	335e3	J / mol
V_3^{diff}	Diffusion creep activation volume	6.0e-6	m^3 / mol
n_3^{dis}	Dislocation creep exponent	3.8	
A_3^{dis}	Dislocation creep prefactor	6.51e-15	$\text{Pa}^{-n_{dis}} \text{s}^{-1}$
Q_3^{dis}	Dislocation creep activation energy	440e3	m^3 / mol
Q_3^{dis}	Dislocation creep activation volume	18e-6	m^3 / mol
c_3	Cohesion	10e6	MPa
ϕ_3	Angle of friction	10	$^\circ$
composition 4 (lower mantle)			
A_4^{diff}	Diffusion creep prefactor	8.88e-15	$\text{Pa}^{-n_{dis}} \text{s}^{-1}$
Q_4^{diff}	Diffusion creep activation energy	335e3	J / mol
V_4^{diff}	Diffusion creep activation volume	6.0e-6	m^3 / mol
n_4^{dis}	Dislocation creep exponent	3.5	
A_4^{dis}	Dislocation creep prefactor	6.51e-15	$\text{Pa}^{-n_{dis}} \text{s}^{-1}$
Q_4^{dis}	Dislocation creep activation energy	530e3	m^3 / mol
Q_4^{dis}	Dislocation creep activation volume	18e-6	m^3 / mol
c_4	Cohesion	20e6	MPa
ϕ_4	Angle of friction	15	$^\circ$

Table 4.1: List of Parameters for the reference model

4.5 Experiments

Many dozens of experiments were needed in search for acceptable rheological settings. The initial model, although geometrically constrained by present-day mantle structure, plate motions and tectonic evolution, requires definition of the initial rheology and temperature field. Particularly, we searched for mantle rheology, within acceptable bounds, that would allow for a smooth continuation of subduction from the starting model avoiding complications such as, e.g., immediate detachment/necking of the slab, or other strong changes in slab morphology. Such 'sudden' slab morphology variation would be an abrupt change from the gradual 30-35 Myr of slab evolution leading to the arcuate slab geometry we reconstructed for 10 Ma. Moreover, strong slab morphology changes would also not be consistent with the present-day tomographic observation which provides a strong constraint on our modelling. Also, many model experiments were conducted to find rheology settings that would create a more or less mantle-stationary trench during 10 Myr of model evolution while still allowing for gradual slab morphology change in the upper mantle depending on (weak, intermediate, or strong) rheology of the deeper upper mantle.

We present a selection of models that assess the geodynamics of arcuate subduction in various ways. We have adopted one model as the reference model, relative to which we will discuss the effects of variations in initial model geometry, plate velocities, and model rheology. The selection of models presented here is based on the parameters which were most influential on model evolution, particularly, for obtaining ongoing subduction at a more or less stationary eastern LA trench in this setup of arcuate subduction.

The variations of the reference model concern using a weaker and stronger crust of the NSAM plate and of plate boundary weak zones, variation of the mantle strength and of slab temperature, addition of compositional weak zones in the slab and the effect of doubling the NSAM plate velocity. The full set of discussed models is shown in table 4.2.

4.5.1 The reference model

The starting geometry of the reference model is discussed in section 4.2. After ample experimentation we arrived at the rheology and temperature settings, as described in table 4.1. Starting point for the rheological parameters come from Gleason and Tullis (1995); Karato (2008); Hirth and Kohlstedt (2003). The reference model rheology is tuned to provide a more or less stationary trench. Figure 4.3 shows the reference model after 10 Myr of simulation, presented with similar view angles as in figure 4.2 giving an overall impression of the flow

4.5. Experiments

Model name	specificity
CFn_ppv15_pv15_nsaqds167e3_wzqds167e3_mvdf5.0e-6	Reference model A
CFn_ppv15_pv15_nsaqds223e3_wzqds223e3_mvdf5.0e-6	Stronger NSA plate crust and weakzone w.r.t. ref. model A
CFn_ppv15_pv15_nsaqds111e3_wzqds111e3_mvdf5.0e-6	Weaker NSA plate crust and weakzone w.r.t. ref. model A
CFn_ppv15_pv15_nsaqds167e3_wzqds167e3_mvdf6.0e-6	Stronger mantle w.r.t. ref. model A
CFn_ppv15_pv15_nsaqds167e3_wzqds167e3_mvdf4.0e-6	Weaker mantle w.r.t. ref. model A
CFn_ppv30_pv15_nsaqds167e3_wzqds167e3_mvdf5.0e-6	Reference model B: Colder initial slab by doubling the past slab velocity
CFn_ppv30_pv15_nsaqds167e3_wzqds167e3_mvdf6.0e-6	Stronger mantle w.r.t. ref. model B
CFn_ppv30_pv15_nsaqds167e3_wzqds167e3_mvdf4.0e-6	Weaker mantle w.r.t. ref. model B
CFn_ppv30_pv30_nsaqds167e3_wzqds167e3_mvdf5.0e-6	Double the NSA plate velocity w.r.t. ref. model B
CFe_ppv15_pv15_nsaqds167e3_wcqds167e3_mvdf5.0e-6	Eastern slab weakzone w.r.t. ref. model A
CFc_ppv15_pv15_nsaqds167e3_wcqds167e3_mvdf5.0e-6	Central slab weakzone w.r.t. ref. model A
CFw_ppv15_pv15_nsaqds167e3_wcqds167e3_mvdf5.0e-6	Western slab weakzone w.r.t. ref. model A

Table 4.2: List of model runs used in this chapter. CF indicates where a slab weakness zone is located (n for none, e for east, c for central and w for west), ppv is the past plate velocity used in the McKenzie model in mm/yr, pv is the plate velocity for the NSAm plate in mm/yr, nsaqds is the dislocation activation energy of the crustal part of the NSA plate and slab, wcqds is the dislocation activation energy of the crustal part of the weak zones and mvdf is the diffusion activation volume of the upper mantle.

field. Following the kinematic boundary condition on the model edges, the Caribbean plate remains stationary and without significant internal deformation while the NSAM plate moves coherently west.

Figure 4.4 shows a top view of the reference model, where the colors indicate the values of the strain rate at the surface. The strain rate is used for display as it allows to delineate the curved trench where high subduction-related strain rates occur. The black lines in the figure are for reference and show that the eastern trench moves slightly west. The observation of a near stationary eastern trench is supported by figure 4.5, in which the red line shows that the location of subduction channel at the base of the lithosphere does not significantly move over 10 Myr. The tip of the slab, however, does move a significant amount, as indicated by the black line, showing that subduction is active. Besides westward movement of the slab tip, part of the 150 km of plate convergence is taken up by the vertical sinking of the slab, increasing its curvature with depth, due to the vertically directed gravitational pull. The stationary nature of the eastern trench can also be seen from the red line in figure 4.6 delineating the slab core at the depth of 150 km and at 1 Myr of the simulation. In this figure we notice a strong difference with the NLA slab which exposes strong slab steepening, while the NLA trench does not appreciably move during 10 Myr of evolution

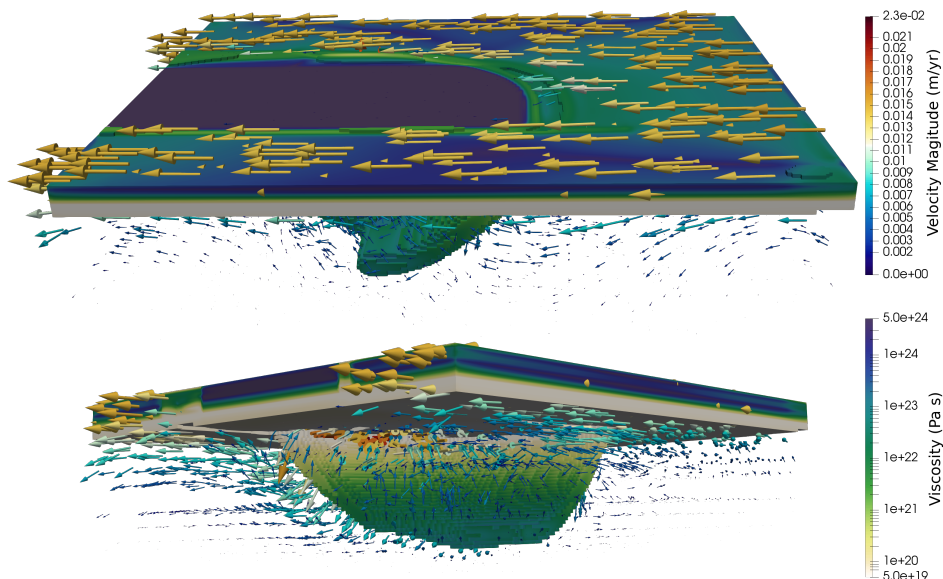


Figure 4.3: **The reference model at 10 Ma with similar viewing angles as in figure 4.2.** The volume rendering is based on temperature while the colors indicate viscosity. In the top box, all temperatures below 600 K and above 1535 K have been made transparent, and in the bottom figure only the temperatures above 1535 have been made transparent. The arrows and arrow colors indicate velocity direction and magnitude, respectively.

(Fig. 4.4).

We also note the high strain rate deformation that is visible at the surface involving the upper-plate weak zone above the NLA slab (green in Fig. 4.2, top) which is broadening towards the south during modelled 10 Myr of evolution (Fig. 4.4). This may indicate an effect of arcuate subduction on upper plate deformation which concurs in location with the actually observed tectonic deformation of upper plate, e.g. Calais et al. (2016). Investigating this further requires more detailed modelling of upper plate structure and rheology which is beyond our scope here.

The reference model exposes clearly the process of slab dragging. The NLA slab is being dragged to the west by the trench-parallel motion of the NAM plate. This can already be seen in figure 4.6 but is more clearly exposed in figure 4.7. The movement of the western slab edge supports the model of "slab edge push" van Benthem et al. (2014). In 10 Myr, the slab edge has been transported westward by about 100 km at the deep tip of the slab and by about 150 km at the top, the latter conforming with the imposed plate velocity. This implies that

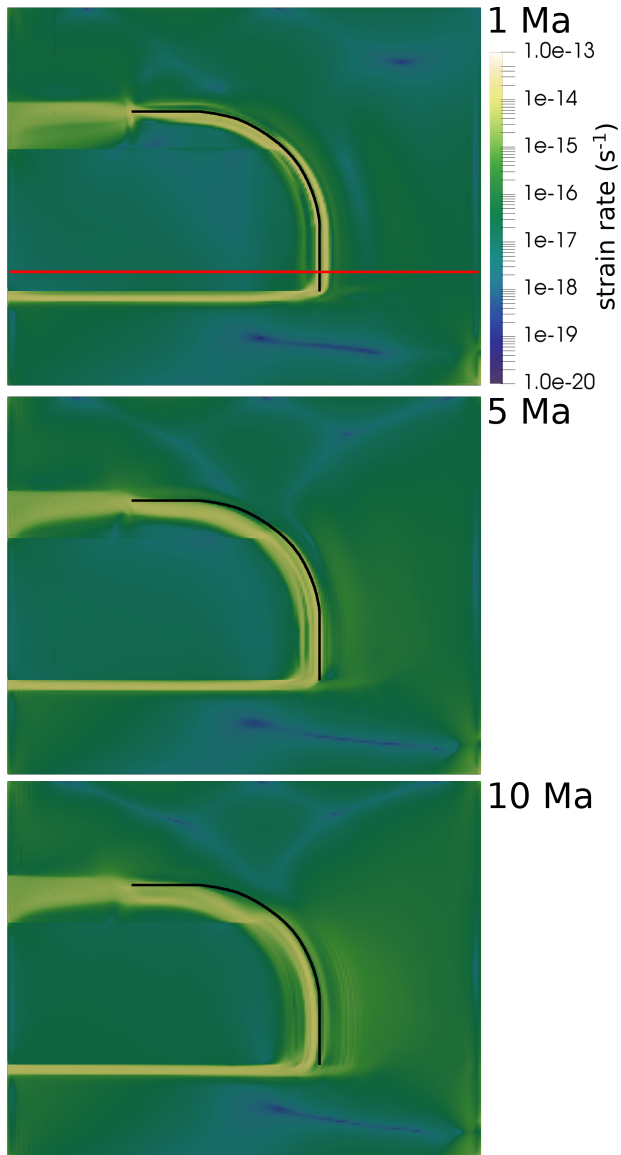


Figure 4.4: A top view of the model with the colors showing the magnitude of the strain rate for 1 Myr, 5 Myr and 10 Myr. The black lines mark the location of the trench at 1 Ma, at the surface. The red line delineates the location where side-views of the 3D model start by making the 3D model transparent to the south of this line as, e.g, in for example figure 4.5.

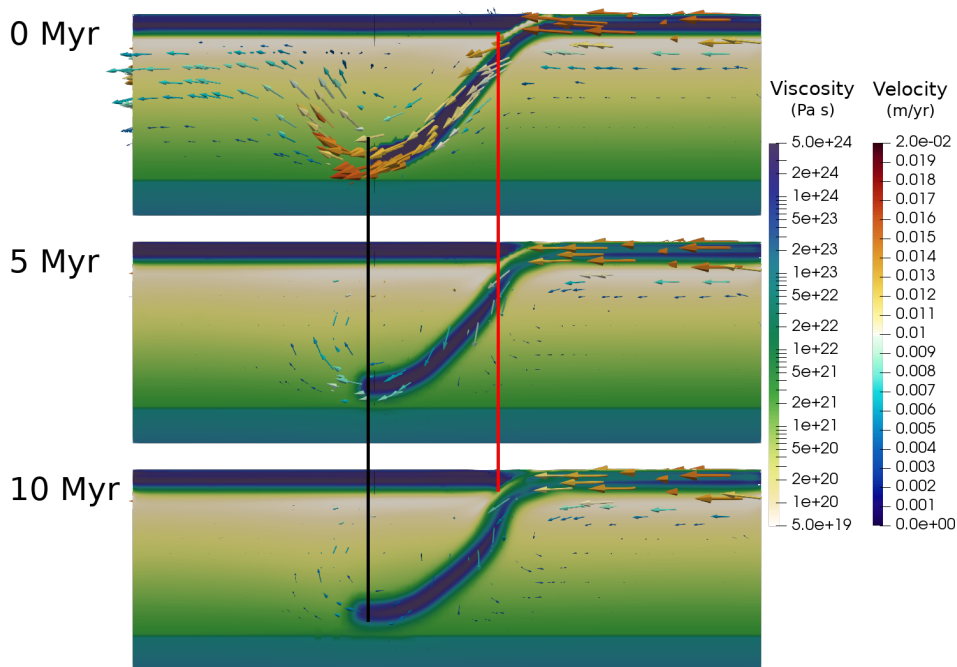


Figure 4.5: **Viscosity and velocity evolution of the reference model through time.** A view from the south to the north, where the southernmost 100 km of the model has been made transparent (see red line in figure 4.4) for location. The color scale indicates the viscosity, the arrow direction show the the velocity direction and magnitude. Note that the arrows still point in their original full 3d direction and may therefore go in and out the plane of the section. The black line shows the initial position of the slab at the base of the upper mantle and the red line shows the initial slab interface location at the base of the lithosphere.

much of the NLA slab is being dragged through the mantle, while this does not happen for the eastern LA slab. The differential motion must be reflected in slab stress and deformation.

To investigate slab stress, we visualize the maximum shear stress magnitude and direction in the slab in figures 4.8 and 4.9. In figure 4.8 the model is viewed towards the north. At 1 Myr, the shear stress pattern in the NLA slab still show much high-amplitude variation which we attribute to a redistribution of stress relative to the initial stage. Maximum shear stress patterns at 5 Myr and 10 Myr show large similarity and show horizontal directions in the deep western portion of the slab and away from the slab edge. This slab segment exposes overall slab-strike parallel horizontal shear which may be explained by the observation that the deep slab has moved less to the west than the top part of the

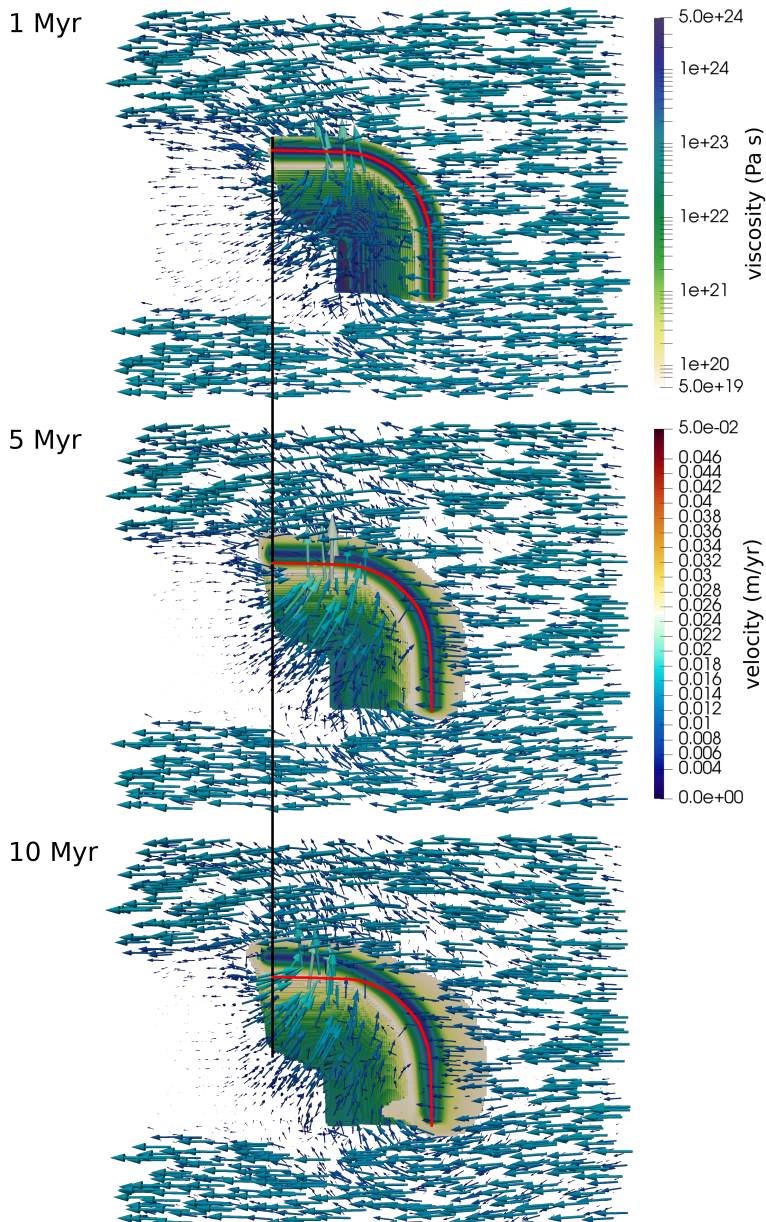


Figure 4.6: **Slab evolution through time for the reference model.** A top view of the model where the surface colors indicate viscosity and where all of the model above 150 km and the temperatures beyond the range 600 to 1535 K have been made transparent. The arrows show the velocity field. The red line shows the location of the slab in the reference model at the depth of 150 km and at 1 Myr, and the black line show the location of the slab edge at 1 Myr.

slab. Another contribution to slab deformation may be the resistance of the mantle against slab dragging (Spakman et al., 2018) which here increases with depth in the upper mantle. We will investigate the effect of mantle viscosity further below. In Fig. 4.8, the stress pattern in the eastern slab is seen in cross sectional view and exposes in the top half of the slab near vertical directions which concur with the observed vertical slab sinking of this slab segment. In figure 4.9 the model is viewed in eastward direction and exposes the maximum shear stress pattern along strike of the eastern LA slab. Near the southern free slab edge, we again observe the vertical shear directions of the upper part of the slab. No slab dragging occurs here and the slab sinks vertically between the stationary trench and the deep portion of the slab that experiences the resistance of increased mantle viscosity. Going to the north along the eastern LA slab, the direction of the maximum shear stress turns quickly along slab-strike to a 45° angle with the dip direction which we attribute to the transition between the relatively stationary eastern LA slab and the NLA slab being dragged towards the west.

Our reference model generally shows maximum shear stress directions that have a large slab-strike parallel component. This is in contrast with the classical '2D' view of dip-aligned directions of slab stress (e.g. Houston, 2015). Importantly, slab dragging in the geometrical framework of arcuate subduction offers an explanation for the observed but, as yet poorly explained, slab-strike parallel horizontal components of earthquake slip (e.g. Myhill, 2012; Houston, 2015; Christova, 2015), which deserves further exploration by numerical modelling.

4.5.2 The influence of the crustal rheology of the subducting plate

The crust of the NSAM plate is subducted along the LA trench. Its rheological strength is tuned to lubricate the subduction channel but if stronger/weaker it may also cause strong shear coupling/decoupling with the overriding Caribbean plate. In this experiment we change the strength of the subducting crust by changing the activation energy of the dislocation creep relative to that of the reference model. This change is equally applied to crust of the 'green' weak zones that serve to (de-)couple lateral plate motion (see figure 4.2, top). In figure 4.10 the effects on plate coupling and slab evolution are shown for the eastern LA slab. Between the strongest and weakest crust, the difference in position of the base of the subduction channel is about 35 km. A stronger subducting crust increases plate coupling and leads to trench advance accommodated by internal EW-deformation of the overriding plate. A too weak crustal strength leads to some slab rollback and increases the tendency of the slab to sink vertically and in the extreme case may lead to slab detachment. The tendency to

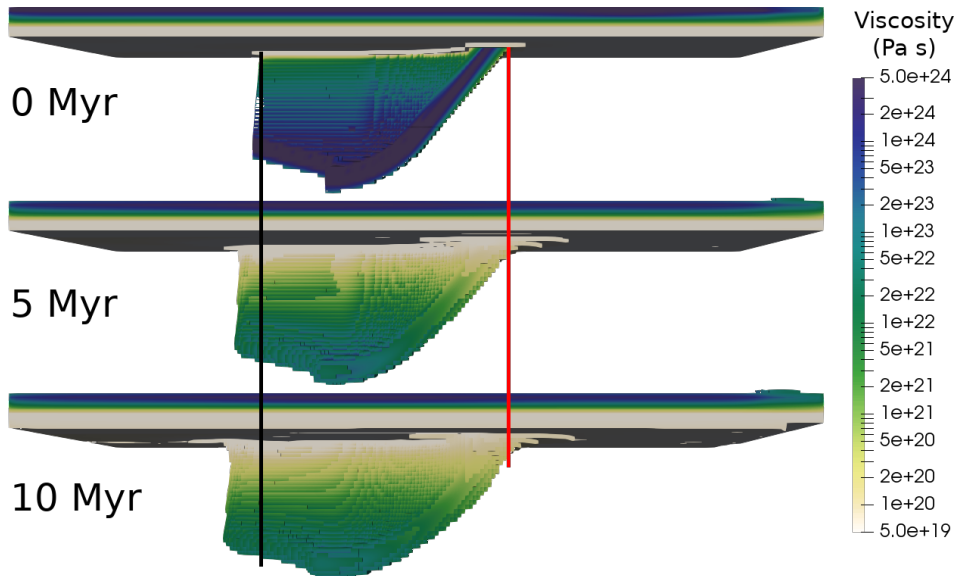


Figure 4.7: A view of the reference model at 0, 5 and 10 Myr. The model is transparent for temperatures below 600 K and above 1535 K. The colors indicate the viscosity. The black line shows the initial location of the western edge of the slab and the red line shows the initial location of the eastern trench.

buckling of the top part of the slab accommodates part of the enforced plate convergence leading to a more easterly position of the slab tip. A subtle interplay between deep-upper mantle viscosity, lithosphere rheology, slab pull, and advancing plate motion is exposed here.

Figure 4.11 presents a top view that also incorporates the response of the NLA slab to varying the strength of the subducting crust. In the model with the strongest crust (top panel), the plate coupling is large enough along the entire trench to cause the Caribbean plate to deform internally and assume local westward motion, which does not occur for the reference model. In case of the weakest crust (bottom panel), the top part of the slab is able to sink vertically while generating a stronger toroidal flow around the slab edge towards the top part of the slab as compared to the reference model (central panel). The toroidal flow around the southern slab edge is less affected.

4.5.3 The influence of mantle viscosity and the temperature of the slab

Slab temperature in the initial model is determined from the McKenzie (1970) analytical slab temperature model. This involves an assumed subduct-

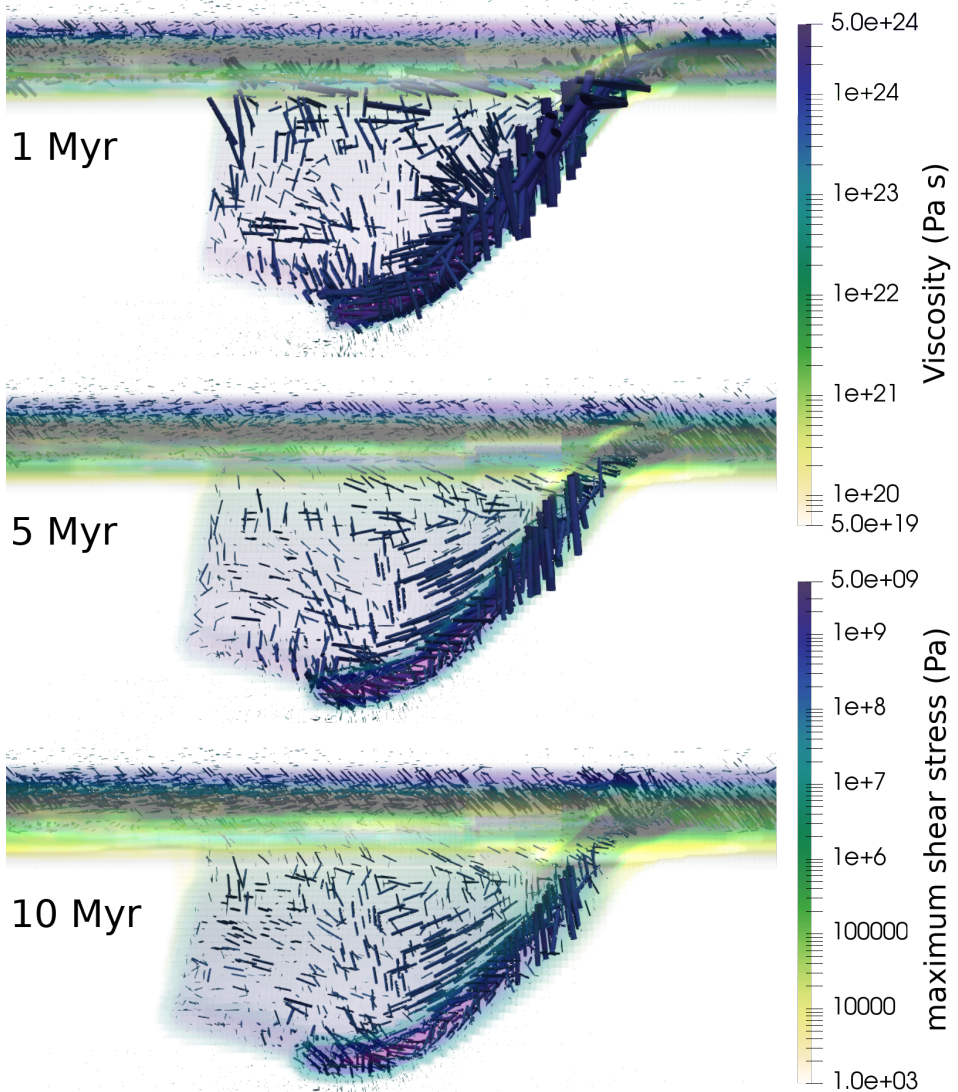


Figure 4.8: **Maximum shear stress through time.** The direction and magnitude of maximum shear stress visualized by the cylinders as viewed from the south. For reference, a partially transparent colored mesh of figure 4.7 has been added.

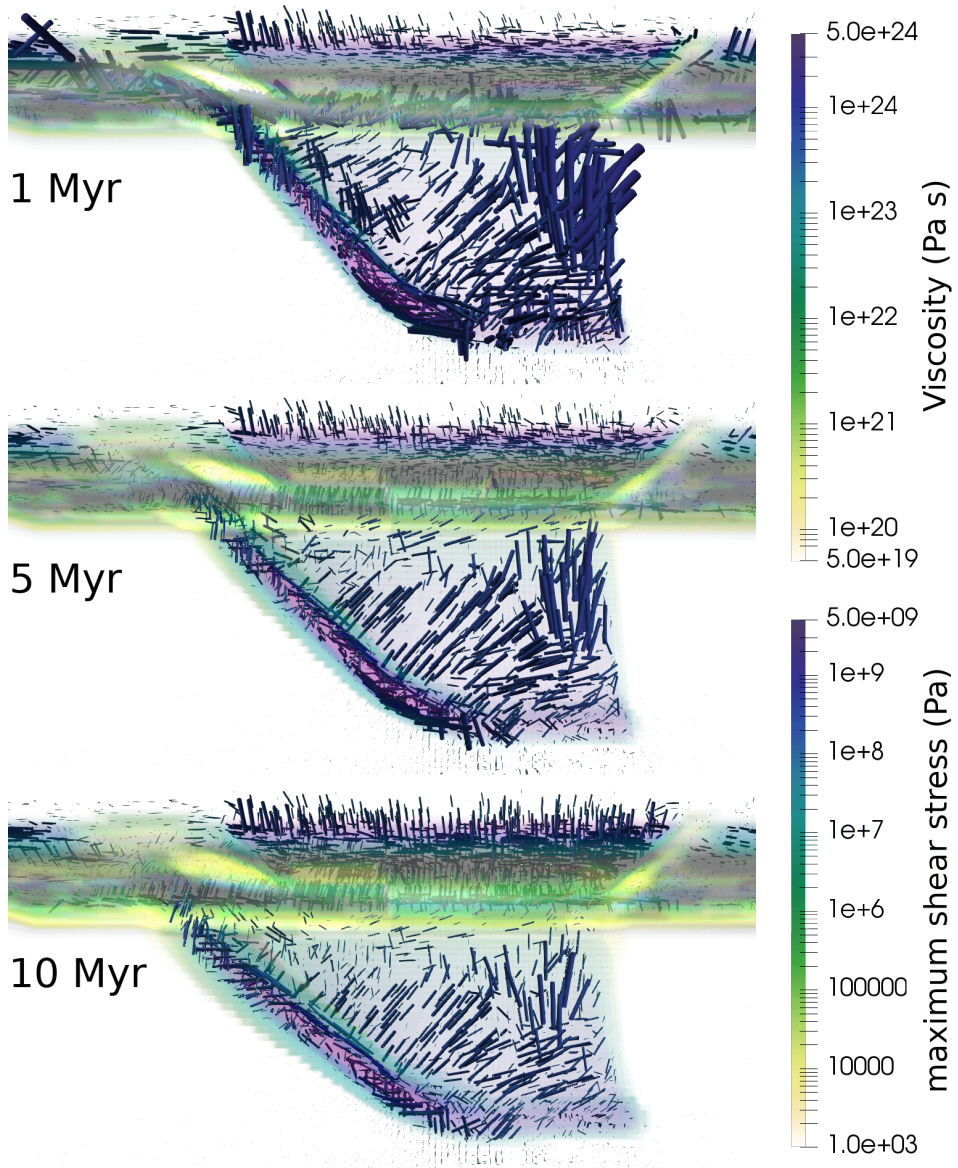


Figure 4.9: **Maximum shear stress through time.** The direction and magnitude of maximum shear stress visualized by the cylinders as viewed from the west. For reference, a partially transparent colored mesh of figure 4.7 has been added.

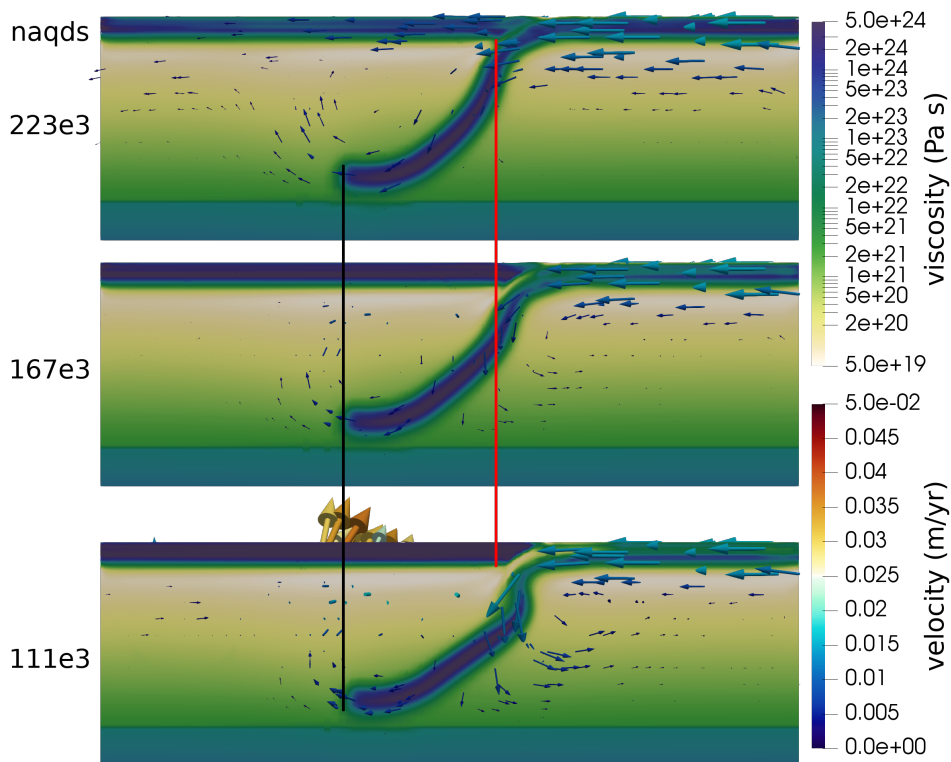


Figure 4.10: **Effects of variation of crustal strength.** Three model results are shown for the eastern LA slab that have been obtained for three different crustal strengths, as parameterized by the dislocation activation energy (naqds; m^3/mol). The black line lines up with the location of the tip of the slab and the red line lines up with the base of the subduction channel, both in the model with strongest (223e3) crust.

ing plate velocity, called here the past plate velocity (ppv), that controls the down-dip advection of slab temperature in the McKenzie model. Here we change the reference model ppv value of 1.5 cm/yr to 3 cm/yr leading to a colder slab with an overall stronger slab rheology as can be seen in figure 4.12. The stronger slab counteracts the tendency for slab buckling leading to a slightly straighter slab geometry as compared to the reference model slab 4.12.

In 6 modelling experiments, we combined these two initial slab-temperature models with a mantle rheology that is weaker, or stronger than in the reference model. The overall strength of composite mantle viscosity was determined by only changing the diffusion creep activation volume from the reference model value of $5\text{e-}6 \text{ m}^3/\text{mol}$ by a value of $1\text{e-}6 \text{ m}^3/\text{mol}$, keeping all other rheologi-

4.5. Experiments

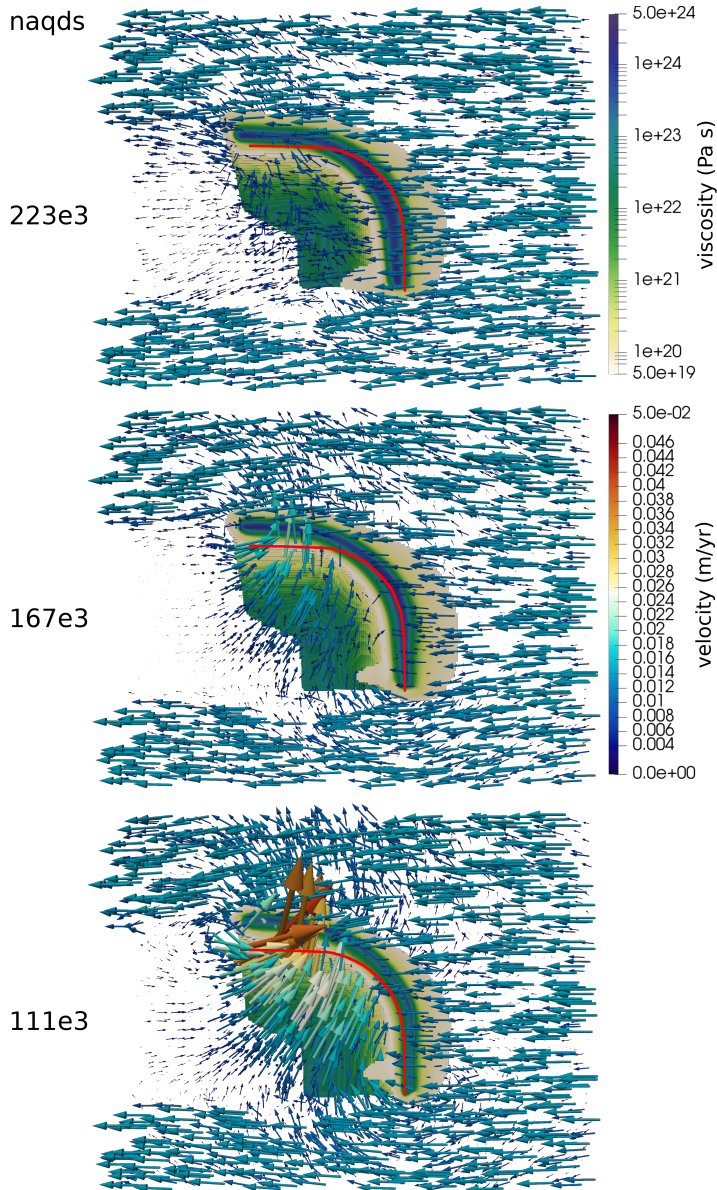


Figure 4.11: **Effects of variation of crustal strength.** A top view of the reference model (middle), a model with a weaker crust (bottom), and a model with a stronger crust (top). The colors indicate viscosity and the model above 150 km and the model volumes with temperatures outside the range 600-1535 K have been made transparent. The arrows show the velocity field. The red line shows the location of the slab in the reference model at the depth of 150 km and at 1 Myr, as in figure 4.6.

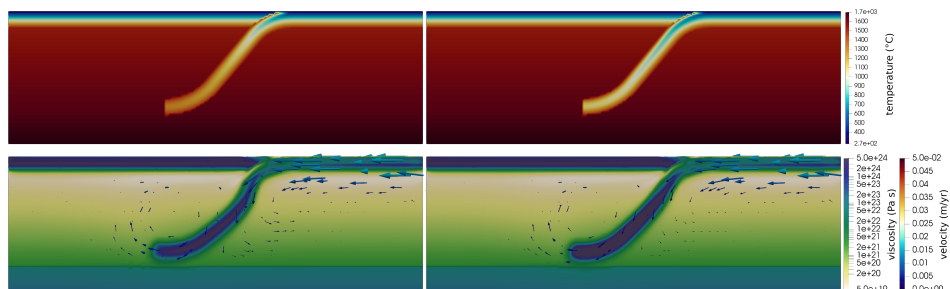


Figure 4.12: **Effect of slab temperature.** Figure layout as in figure 4.5. The left panels show the reference model, and the right panels show the 'cold' model with a ppv of 3 cm/yr. The top panels shows the temperature of the initial model, and the bottom panels shows the viscosity in colors and velocity in arrows after 10 Myr. See text for details.

cal settings of the reference model. In figure 4.13 six models are shown for the two levels of slab strength and 3 levels of mantle strength. The first notable effect is that a stronger mantle provides more viscous resistance to westward slab dragging as shown by the more eastward position of the slab near the 660 as compared to the weaker mantle cases. The second effect is that the entire slab sinks to larger depth in weaker mantle which results from the decreased viscous slab-mantle coupling. The difference in slab depth is 75-100 km between the strongest and weakest mantle. The third effect is that the southern edge of the slab just above 660 km has moved 60 km more to the west in the case of the weak mantle, as compared to the strong mantle. On the north side, the slab tip (of the south-dipping slab) is more or less at the same location. Lastly, the overall patterns of directions of maximum shear stress are comparable, but not the magnitudes. The colder slab can carry larger-amplitude stress, while a weaker mantle allows for more slab deformation associated with higher levels of slab stress.

These experiments demonstrate that slab dragging and the associated slab stress field are robust features with respect to the variations in slab temperature and in overall mantle viscosity. In addition, slab morphology (seismological observations) and slab stress (earthquake mechanism observations) may be important diagnostics for the dynamic state of arcuate subduction systems that involve slab dragging.

4.5.4 The influence of vertical weak zones in the slab

The Geodynamic World Builder facilitates making changes to the 3D initial model. Here we replace a vertical slab segment by much weaker material (as

4.5. Experiments

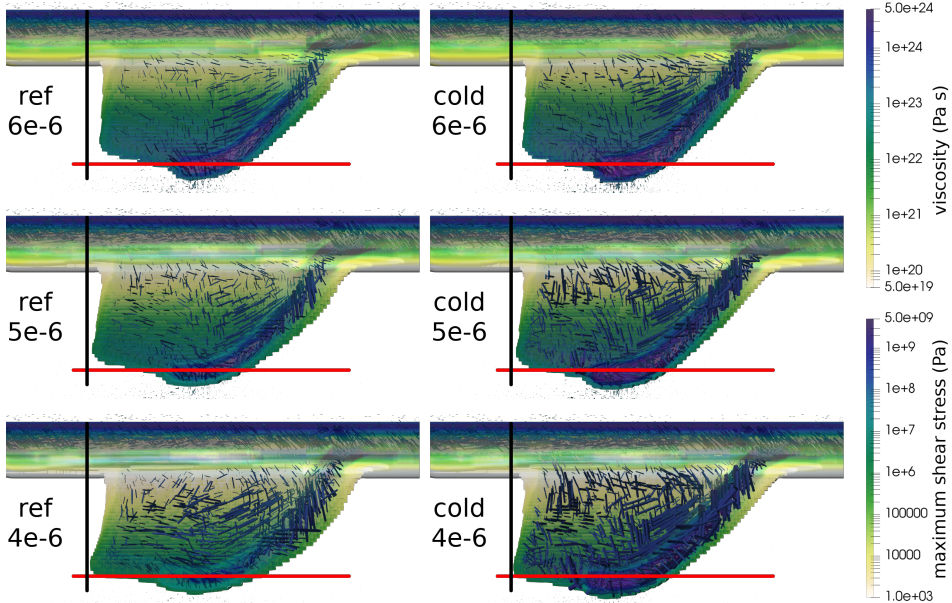


Figure 4.13: **Effect of mantle and slab strength.** A north-directed view, as in figure 4.7, of 6 slab models overlain by maximum shear stress directions as in figure 4.8. The left panels show models starting from the initial slab temperature of the reference model. The panels on the right show models computed with the colder initial slab of Fig. 4.12. The top row of panels shows models with a stronger overall upper mantle rheology defined by only changing the reference model diffusion creep activation volume to $6e-6 \text{ m}^3/\text{mol}$, the center panels show models based on the reference model rheology with a diffusion creep activation volume value of $5e-6 \text{ m}^3/\text{mol}$, and the bottom panels show models having the weakest mantle resulting involving the diffusion creep activation volume of $4e-6 \text{ m}^3/\text{mol}$. The black and red lines are plotted at exactly the same location in each model for cross-reference.

indicated in Fig. 4.2) to test for slab deformation during subduction evolution. The vertical weak zone is implemented by replacing the crust and mantle rheology of a specific GWB part of the slab by a much weaker rheology, decreasing the local slab strength by about 2 orders of magnitude. We inserted such slab weakness at three different places. The results (3D viscosity and velocity field) after 10 Myr of model evolution are shown in figure 4.14. The presence of a weak zone in the slab does not have a large influence on the overall slab evolution. The weak zone is, to first order, advected with the stronger ambient slab. One cause for this is that at the surface, the lithosphere of the subducting NSAM plate is defined to be rather strong (typically 10^{24} - 10^{25} Pa s) which does not allow for much lateral intra-plate deformation, implying that the lat-

eral width of the weak zone in the top of the slab is largely preserved. Similarly, with increasing depth the strength of the weak zone slowly increases, following a similar trend in the stronger adjacent slab, which causes that also in the deep upper mantle lateral slab deformation is hampered. More generally, this indicates that the lateral rheological strength of the subducting plate may strongly control the lateral deformation of the slab during arcuate subduction.

Although the lateral slab deformation is restricted, we can observe variable slab sinking, i.e. associated with vertical slab deformation, that correlates with the position of the inserted vertical weakness. This is best exposed by the lateral mantle velocity field under the Caribbean plate as can be inspected from figure 4.14 which exposes for each model different patterns of strong mantle flow towards the NLA slab. This flow is forced by the slab geometry change due to differential vertical slab sinking as was observed in our reference model for the NLA (Fig. 4.6) and SLA slabs (Fig. 4.5) and is amplified when the strength of the subducted crust is reduced (Fig. 4.10). A vertical weak zone in the slab leads to sufficient decoupling of the adjacent slab segments to facilitate differential slab sinking. Associated small slab geometry changes relative to the reference model slab (top panel; Fig. 4.6) can be identified in figure 4.14 where the red reference line follows the core of the reference slab. The closer the weak zone is located towards the western slab, the more slab sinking is accommodated by a laterally smaller segment of the slab. More detailed models and longer model evolution are required to investigate if inherited slab weakness may lead to vertical slab segmentation resulting from differential vertical and/or horizontal slab motion, as proposed for the NLA slab (Harris et al., 2018). In contrast, the eastern trench is not sensitive to the presence of a weak zone in the slab.

4.5.5 The influence of the subducting plate velocity

In this section we investigate the effect of doubling the NSAM plate velocity to 3 cm/yr to investigate its role in the geodynamic evolution of the arcuate slab. The initial temperature model of the slab is based on a ppv of 3 cm/yr as in Fig. 4.12, while the overall rheological model is that of the reference model. The resulting model, called model B, is compared to a model A based on the NSAM velocity of 1.5 cm/yr and also on a ppv of 3 cm/yr. We compare the slab in model A at stages 10 Myr and 20 Myr with the slab in model B at 5 Myr and 10 Myr of evolution, respectively, such that an equal amount of plate convergence has occurred at each of the two stages of slab evolution.

Figure 4.15 shows the slab at these two comparable stages of evolution. The main inference is that slab geometry at these two stages is highly comparable

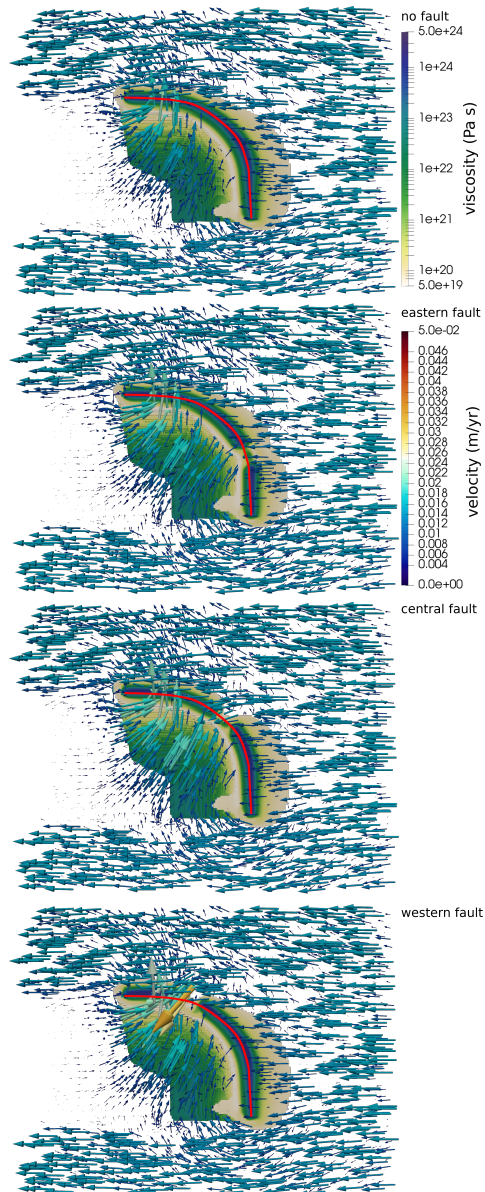


Figure 4.14: **Effects of vertical weak zones in the slab.** A top view of 4 models at 10 Myr, which differ by the presence and location of a weak zone in the slab. The surface colors indicate viscosity and where all of models above 150 km and the temperatures outside the range 600 to 1535 K have been made transparent. The arrows show the velocity field of the model. The curved red line shows the location of the slab without a weak zone.

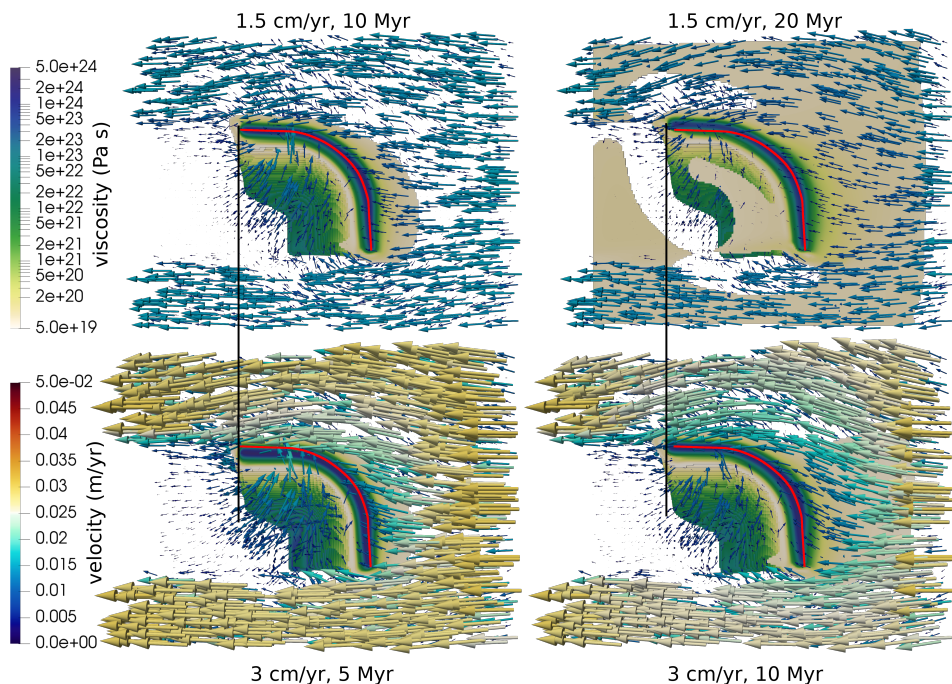


Figure 4.15: **Effect of doubling the subducting slab velocity.** Varying the subducting plate velocity from 1.5 cm/yr to 3 cm/yr for an initial slab with a ppv of 3cm/yr. The top two panels show model A with a plate velocity of 1.5 cm/yr at 10 Myr and 20 Myr while the bottom panels show model B, with subducting plate velocity of 3 cm/yr, at the comparable plate convergence stages of 5 Myr and 10 Myr, respectively. Model volumes with temperature below 600 K have been made transparent. The model viscosity is color coded. Arrows indicate the 3D velocity field with velocity magnitude shown in color. The red line is plotted for reference and delineates the core of the slab of model A at 10 Myr (top-left panel)

showing only morphology differences on the scale of 10-20 km. One cause for these small differences is the earlier inference that, due to a vertical component of slab sinking (e.g. Fig. 4.5 and Fig. 4.6), the slab morphology is also dependent on the viscosity of the ambient mantle (illustrated in Fig. 4.13) and thus on the slab model-residence time in the mantle. The latter is different between both models at the temporal stages of comparison. The most notable difference in slab curvature between model A and B may possibly be attributed to that effect. The western slab edge in both models is almost at the same position in both stages. This is to be expected near the top where the stiff NSAM plate rheology dominates slab dragging. Close comparison of the position of the entire western slab edge between models A and B shows that this is highly

comparable with differences in position at depth not exceeding 10 km. As we observed earlier (Fig. 4.6 and Fig. 4.7), the deep slab tip of the western edge in the reference model has been dragged by 50 km less to the west during 10 Myr of slab evolution. This is due to the mantle resistance against slab dragging and resistance of the slab against lateral deformation and the effect is similar for the stiffer slab of model A as can be inferred from Fig. 4.13 where the end-stages of the reference slab and the stiffer slab are highly comparable. The observation that the western slab edge in model B, at 5 and 10 My, is almost at the same position as in model A at 10 Myr and 20 Myr, respectively, leads to the important implication that doubling the absolute plate (dragging) motion of the subducting plate effectively increases the mantle resistance against slab dragging in the deeper upper mantle. We did not investigate the precise nature of the mantle resistance against slab dragging which can be a combination of the interaction of the base of the slab with the more viscous lower mantle and the viscous coupling between the slab and ambient upper mantle.

4.6 Discussion and conclusions

We assessed first-order geodynamic characteristics of strongly arcuate subduction systems for which we took the plate tectonic setting of the Lesser Antilles subduction as the natural example. We took an advanced subduction stage as starting point which was the relatively well-constrained plate tectonic setting of the Lesser Antilles trench at 10 Ma (e.g. van Benthem et al., 2014; Boschman et al., 2014). With the Geodynamic World Builder (Fraters et al., 2019b)(Chapter 3), we constructed the 3D slab following the arcuate trench-geometry and with a lateral slab curvature based on present-day tomography (Harris et al., 2018). The slab geometry is not necessarily correct but defines a starting point for modelling the evolution of arcuate subduction. Our numerical subduction experiments are conducted with the open source ASPECT code and are particularly based on the implementation of composite nonlinear visco-plastic rheologies (Glerum et al., 2018) and on the faster and stable convergence provided by the Newton method (Fraters et al., 2019a)(Chapter 2).

We particularly designed our experiments for a "90°"-arcuate subduction setting that combines trench-perpendicular subducting-plate convergence at a mantle-stationary trench along one side of the system with trench-parallel subducting-plate motion occurring at a mantle-stationary trench along the other side. This geodynamic setting defined the reference subduction model for which we showed in various experiments that the main dynamic features of modelled arcuate subduction evolution are robust with respect to realistic variations of the kinematic settings and of the rheological crust-mantle properties.

Foremost, our simulations of 10 Myr, in one case 20 Myr, of subduction evolution demonstrated that slab dragging, i.e. lateral transport of the lithosphere slab by the surface motion of the subducting plate, is a geodynamically feasible process that leads to internal slab deformation but not necessarily to immediate slab tearing or destruction of the arcuate nature in other ways. Slab dragging is demonstrated for a large subduction system with an arcuate slab across the entire upper mantle involving a more than ~ 2000 km long trench and for trench-parallel dragging velocities of 15 and 30 mm/yr during 10-20 Myr, involving trench-parallel slab transport by 150-300 km for the top of the slab. Earlier assessments of the geodynamical feasibility of slab dragging concerned relative short (< 200 km) and laterally small (~ 100 km) slab 'stumps' being dragged by the Pacific plate under the western NAM plate boundary (Pikser et al., 2012) and the isolated, about 200-300 km wide, Gibraltar slab across the upper mantle of the westernmost Mediterranean (Spakman et al., 2018). Here, the application to eastern Caribbean subduction demonstrates that the overall slab geometry obtained after 10 My of evolution conforms with the tomographic images of the present-day LA slab, which agrees with the hypothesis that westward slab dragging, involving the slab across the entire upper-mantle, occurs along the northern LA trench.

A next key observation is that slab dragging is resisted by the slab and mantle involving two dynamic effects. The first effect is the viscous mantle resistance against slab dragging, which was demonstrated in experiments using a weaker or stronger mantle rheology relative to the reference model. The second effect involves resistance of the slab itself against slab dragging. This is more of a geometrical nature and is here caused by the geodynamic setting in which the northern slab is moving away from the eastern slab. This implies laterally distributed deformation of the slab which is associated with a complex 3D stress field. We visualized the stress field by showing the magnitudes and directions of the maximum shear stress. For rheologically isotropic media, as we model here, such directions indicate candidate directions of earthquake fault-slip. We conclude from our experiments that slab dragging sets up a complex slab stress field characterized by strong deviations in the slab stress orientation from the more classical view of maximum stress directions parallel to the trench-perpendicular slab-dip plane. Particularly, we noticed horizontal slab-strike parallel maximum directions of shear stress in the northern slab which we relate to the bulk horizontal-shear deformation by 50 km across the entire upper-mantle slab that occurred during 10 Myr of slab dragging with 15 mm/yr. The generally off-dip directions of maximum shear stress explain observations

of slab-strike parallel slip components inferred from focal mechanisms of intermediate and deep earthquakes (e.g. Giardini and Woodhouse, 1986; Myhill and Warren, 2012; Houston, 2015; Christova, 2015; Meighan et al., 2013).

Slab stress and deformation due to slab dragging are transmitted into the upper plate where it leads to crustal deformation (Spakman et al., 2018). Modelling the tectonic response to slab dragging was not our goal, but in our reference model we did infer a first-order effect of upper plate deformation directly south of the northern trench, which concurs in location with the observed upper plate deformation along the NLA plate boundary (e.g. Calais et al., 2016).

Our work realizes an important step towards the numerical simulation of natural subduction by demonstrating the (open source) capabilities and computational feasibility of assessing the first-order geodynamic features associated with natural (arcuate) subduction such that the next step of detailed modelling of the tectonic evolution of the overlying crust is now within reach.

4.7 Appendix A: GWB input

The reference world builder input file is given in listing 4.1. Colder slabs are made through higher velocity in the temperature model of the slab. Note that the notation of version 0.2.0 of the GWB would be a lot shorter. See listing 3.13 for an example which is written for version 0.2.0 and which has a very similar initial setting. The weak zones are created by adding one of the three listings between the subducting slab and South Weakzone features: listing 4.2 for the western fault, listing 4.3 for the central fault and listing 4.4 for the eastern fault.

```

1 {
2   "version": "0.1",
3   "potential mantle temperature": 1500,
4   "thermal expansion coefficient": 2.0e-5,
5   "minimum points per distance": 100000,
6   "interpolation": "monotone spline",
7   "coordinate system": {"cartesian": {}},
8   "features":
9   {
10    "oceanic plate": {"name": "NS American plate",
11     "coordinates":
12     [[1700e3, 0], [1700e3, 300e3], [1689e3, 422e3], [1658e3, 539e3],
13     [1606e3, 650e3], [1536e3, 749e3], [1450e3, 836e3], [1350e3, 906e3],
14     [1239e3, 958e3], [1122e3, 989e3], [1000e3, 1000e3], [-1e3, 1000e3],
15     [-1e3, 1501e3], [2501e3, 1501e3], [2501e3, -501e3], [-1e3, -501e3],
16     [-1e3, -50e3], [2000e3, -50e3], [2000e3, 0e3]],
17     "temperature model": {"name": "linear", "depth": 100e3},
18     "composition model":
19     {"name": "constant layers", "depth": 100e3,
20      "layers": [{"compositions": [0], "thickness": 30e3}]},
21
22    "oceanic plate": {"name": "Caribbean plate",
23     "coordinates":
24     [[1700e3, 300e3], [1689e3, 422e3], [1658e3, 539e3], [1606e3, 650e3],
25     [1536e3, 749e3], [1450e3, 836e3], [1350e3, 906e3], [1239e3, 958e3],
26     [1122e3, 989e3], [1000e3, 1000e3], [650e3, 1000e3], [-1e3, 1000e3],
27     [-1e3, 0e3], [1700e3, 0e3]],
28     "temperature model": {"name": "linear", "depth": 100e3},
29     "composition model":
30     {"name": "constant layers", "depth": 100e3,
31      "layers": [{"compositions": [1], "thickness": 30e3}]},
32
33    "continental plate": {"name": "Caribbean weak zone",
34     "coordinates":
35     [[-1e3, 1000e3], [-1e3, 750e3], [1536e3, 749e3], [1450e3, 836e3],
36     [1350e3, 906e3], [1239e3, 958e3], [1122e3, 989e3], [1000e3, 1000e3],
37     [650e3, 1000e3]],
38     "temperature model": {"name": "linear", "depth": 100e3},
39     "composition model":
40     {"name": "constant layers", "depth": 100e3,
41      "layers": [{"compositions": [2], "thickness": 30e3},
42                {"compositions": [3], "thickness": 70e3}]},
43
44    "mantle layer": {"name": "upper mantle", "top depth": 100e3,
45     "coordinates":
46     [[-1e3, -500e3], [-501e3, 2500e3], [2501e3, 2500e3], [2501e3, -501e3]],
47     "temperature model": {"name": "none"},
48     "composition model": {"name": "none"}},
49
50    "mantle layer": {"name": "660", "top depth": 660e3,
51     "coordinates":

```


4.7. Appendix A: GWB input

```
52     [[-1e3,-500e3],[-501e3,2500e3],[2501e3,2500e3],[2501e3,-501e3]],
53     "temperature model":{"name":"none"},
54     "composition model":
55     {"name":"constant", "depth":200e3, "compositions":[4]}},
56
57     "subducting plate":{"name":"Lesser Antilles slab",
58     "coordinates":
59     [[1700e3,0],[1700e3,300e3],[1689e3,422e3],[1658e3,539e3],
60     [1606e3,650e3],[1536e3,749e3],[1450e3,836e3],[1350e3,906e3],
61     [1239e3,958e3],[1122e3,989e3],[1000e3,1000e3],[650e3,1000e3]],
62     "reference point":[-1,-1],
63     "starting depth":0, "maximum depth":660e3,
64     "segments":
65     {
66     "all":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
67     {"length":150e3, "thickness":100e3, "angle":[25,50]},
68     {"length":371e3, "thickness":100e3, "angle":[50]},
69     {"length":300e3, "thickness":100e3, "angle":[50,0]},
70     {"length":0, "thickness":100e3, "angle":[0]}],
71     "1":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
72     {"length":150e3, "thickness":100e3, "angle":[25,50]},
73     {"length":371e3, "thickness":100e3, "angle":[50]},
74     {"length":275e3, "thickness":100e3, "angle":[50,0]},
75     {"length":0, "thickness":100e3, "angle":[0]}],
76     "2":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
77     {"length":150e3, "thickness":100e3, "angle":[25,50]},
78     {"length":371e3, "thickness":100e3, "angle":[50]},
79     {"length":275e3, "thickness":100e3, "angle":[50,0]},
80     {"length":0, "thickness":100e3, "angle":[0]}],
81     "3":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
82     {"length":150e3, "thickness":100e3, "angle":[25,50]},
83     {"length":371e3, "thickness":100e3, "angle":[50]},
84     {"length":275e3, "thickness":100e3, "angle":[50,0]},
85     {"length":0, "thickness":100e3, "angle":[0]}],
86     "4":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
87     {"length":150e3, "thickness":100e3, "angle":[25,50]},
88     {"length":371e3, "thickness":100e3, "angle":[50]},
89     {"length":275e3, "thickness":100e3, "angle":[50,0]},
90     {"length":0e3, "thickness":100e3, "angle":[0]}],
91     "5":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
92     {"length":150e3, "thickness":100e3, "angle":[25,50]},
93     {"length":371e3, "thickness":100e3, "angle":[50]},
94     {"length":275e3, "thickness":100e3, "angle":[50,0]},
95     {"length":0, "thickness":100e3, "angle":[0]}],
96     "6":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
97     {"length":150e3, "thickness":100e3, "angle":[25,50]},
98     {"length":371e3, "thickness":100e3, "angle":[50]},
99     {"length":275e3, "thickness":100e3, "angle":[50,0]},
100    {"length":0, "thickness":100e3, "angle":[0]}],
101    "7":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
102    {"length":150e3, "thickness":100e3, "angle":[25,50]},
103    {"length":371e3, "thickness":100e3, "angle":[50]},
104    {"length":275e3, "thickness":100e3, "angle":[50,0]},
105    {"length":0, "thickness":100e3, "angle":[0]}],
106    "8":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
107    {"length":150e3, "thickness":100e3, "angle":[25,50]},
108    {"length":371e3, "thickness":100e3, "angle":[50]},
109    {"length":275e3, "thickness":100e3, "angle":[50,0]},
110    {"length":0, "thickness":100e3, "angle":[0]}],
111    "9":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
112    {"length":150e3, "thickness":100e3, "angle":[25,50]},
113    {"length":371e3, "thickness":100e3, "angle":[50]},
114    {"length":275e3, "thickness":100e3, "angle":[50,0]},
115    {"length":0, "thickness":100e3, "angle":[0]}],
116    "10":[{"length":150e3, "thickness":100e3, "angle":[0,25]},
```

```

117     {"length":150e3, "thickness":[100e3], "angle":[25,50]},
118     {"length":371e3, "thickness":[100e3], "angle":[50]},
119     {"length":275e3, "thickness":[100e3], "angle":[50,0]},
120     {"length":0, "thickness":[100e3], "angle":[0]},
121     "11": [{"length":150e3, "thickness":[100e3], "angle":[0,25]},
122            {"length":150e3, "thickness":[100e3], "angle":[25,50]},
123            {"length":371e3, "thickness":[100e3], "angle":[50]},
124            {"length":50e3, "thickness":[100e3], "angle":[50,0]},
125            {"length":0, "thickness":[100e3], "angle":[0]}
126   ],
127   "temperature model":
128   {"name":"plate model", "density":3300, "plate velocity":0.0150,
129    "thermal conductivity":2.5, "thermal expansion coefficient":2e-5 },
130   "composition model":
131   {"name":"constant layers",
132    "layers":[{"compositions":[0], "thickness":30e3},
133             {"compositions":[0], "fractions":[0], "thickness":70e3}]}},
134
135   "continental plate":{"name":"South Weakzone",
136    "coordinates":[[-1e3,0e3],[-1e3,-50e3],[2000e3,-50e3],[2000e3,0e3]],
137    "temperature model":{"name":"linear", "depth":100e3},
138    "composition model":{"name":"constant layers", "depth":100e3,
139     "layers":[{"compositions":[2], "thickness":30e3},
140              {"compositions":[3], "thickness":70e3}]}},
141   }
142 }

```

Listing 4.1: The reference GWB input file

```

1   "oceanic plate": {"name":"Compositional Western plate Fault",
2    "coordinates":[[1239e3,958e3],[1122e3,989e3],[1232e3,1167e3]],
3    "temperature model":{"name":"none"},
4    "composition model":{"name":"constant layers",
5     "layers":[{"compositions":[2], "thickness":30e3},
6              {"compositions":[3], "thickness":70e3}]}},
7
8   "subducting plate":{"name":"Compositional Western slab Fault",
9    "coordinates":[[1239e3,958e3],[1122e3,989e3]], "reference point":[-1,-1],
10   "segments":{"
11     "all":[{"length":150e3, "thickness":[100e3], "angle":[0,25]},
12            {"length":150e3, "thickness":[100e3], "angle":[25,50]},
13            {"length":371e3, "thickness":[100e3], "angle":[50]},
14            {"length":275e3, "thickness":[100e3], "angle":[50,0]}]},
15   "temperature model":{"name":"none"},
16   "composition model":{"name":"constant layers",
17    "layers":[{"compositions":[2], "thickness":30e3},
18             {"compositions":[3], "thickness":70e3}]}},

```

Listing 4.2: The addition to the reference GWB input file needed to create the western slab weakzone.

```

1   "oceanic plate": {"name":"Compositional Western plate Fault",
2    "coordinates":[[1536e3,749e3],[1450e3,836e3],[1635e3,932e3]],
3    "temperature model":{"name":"none"},
4    "composition model":{"name":"constant layers",
5     "layers":[{"compositions":[2], "thickness":30e3},
6              {"compositions":[3], "thickness":70e3}]}},
7
8   "subducting plate":{"name":"Compositional Western slab Fault",
9    "coordinates":[[1536e3,749e3],[1450e3,836e3]], "reference point":[-1,-1],
10   "segments":{"
11     "all":[{"length":150e3, "thickness":[100e3], "angle":[0,25]},
12            {"length":150e3, "thickness":[100e3], "angle":[25,50]},

```

4.8. Appendix B: ASPECT input

```
13         {"length":371e3, "thickness":[100e3], "angle":[50]},
14         {"length":275e3, "thickness":[100e3], "angle":[50,0]}}},
15     "temperature model":{"name":"none"},
16     "composition model":{"name":"constant layers",
17       "layers":[{"compositions":[2], "thickness":30e3},
18       {"compositions":[3], "thickness":70e3}]}}},
```

Listing 4.3: The addition to the reference GWB input file needed to create the central slab weakzone.

```
1     "oceanic plate": {"name":"Compositional Western plate Fault",
2       "coordinates":[[1689e3,422e3],[1658e3,539e3],[1867e3,532e3]],
3       "temperature model":{"name":"none"},
4       "composition model":{"name":"constant layers",
5         "layers":[{"compositions":[2], "thickness":30e3},
6         {"compositions":[3], "thickness":70e3}]}}},
7
8     "subducting plate":{"name":"Compositional Western slab Fault",
9       "coordinates":[[1689e3,422e3],[1658e3,539e3]], "reference point":[-1,-1],
10      "segments":{"
11        "all":[{"length":150e3, "thickness":[100e3], "angle":[0,25]},
12        {"length":150e3, "thickness":[100e3], "angle":[25,50]},
13        {"length":371e3, "thickness":[100e3], "angle":[50]},
14        {"length":275e3, "thickness":[100e3], "angle":[50,0]}}},
15      "temperature model":{"name":"none"},
16      "composition model":{"name":"constant layers",
17        "layers":[{"compositions":[2], "thickness":30e3},
18        {"compositions":[3], "thickness":70e3}]}}},
```

Listing 4.4: The addition to the reference GWB input file needed to create the eastern slab weakzone.

4.8 Appendix B: ASPECT input

```
1 set World builder file = world/builder/file/location.wb
2 set Output directory = aspect/output/dir/
3 set Dimension = 3
4 set CFL number = 0.05
5 set Max nonlinear iterations = 50
6 set End time = 2e7
7 set Nonlinear solver scheme = iterated Advection and Newton Stokes
8 set Timing output frequency = 0
9 set Max nonlinear iterations in pre-refinement = 0
10 set Pressure normalization = no
11 set Nonlinear solver tolerance = 1e-6
12 set Maximum time step = 10000
13 set Maximum relative increase in time step = 25
14 set Maximum first time step = 100
15
16 subsection Solver parameters
17   subsection Newton solver parameters
18     set Max pre-Newton nonlinear iterations = 100
19     set Nonlinear Newton solver switch tolerance = 1e-4
20     set Max Newton line search iterations = 0
21     set Maximum linear Stokes solver tolerance = 1e-1
22     set Use Newton residual scaling method = false
23     set Use Newton failsafe = true
24     set Stabilization preconditioner = SPD
25     set Stabilization velocity block = SPD
26     set Use Eisenstat Walker method for Picard iterations = true
```

Chapter 4. Assessing the geodynamics of strongly arcuate subduction zones:
the eastern Caribbean subduction setting.

```

27 end
28 end
29
30 subsection Geometry model
31 set Model name = box with lithosphere boundary indicators
32 subsection Box with lithosphere boundary indicators
33 set X extent = 2500e3
34 set Y extent = 2000e3
35 set Z extent = 800e3
36 set Box origin Y coordinate = -500e3
37 set X repetitions = 25
38 set Y repetitions = 20
39 set Y repetitions lithosphere = 10
40 set Z repetitions = 6
41 set Z repetitions lithosphere= 1
42 set Lithospheric thickness = 100e3
43
44 end
45 subsection Box
46 set X extent = 3960000
47 set Y extent = 4620000
48 set Z extent = 660000
49 set X repetitions = 6
50 set Y repetitions = 7
51 end
52 subsection Ellipsoidal chunk
53 set NE corner = -58:26
54 set SW corner = -82:14
55 set Depth = 700000
56 set Semi-major axis = 6378137
57 set Eccentricity = 0
58 set East-West subdivisions = 4
59 set North-South subdivisions = 2
60 set Depth subdivisions = 1
61 end
62 end
63
64 subsection Material model
65 set Model name = visco plastic
66 subsection Visco Plastic
67 set Reference viscosity = 1.0e21
68 set Grain size = 0.01
69 set Viscosity averaging scheme = harmonic
70 set Minimum viscosity = 5e19
71 set Maximum viscosity = 5e24
72 set Reference temperature = 293 lith, lower mantle
73 set Densities = 3300 3300, 3300
74 set Thermal expansivities = 2e-5
75 # Up M., NA C.,
76 # CARBC, Wk C., Wk L., Lower Mantle
77 set Prefactors for dislocation creep = 6.51e-15, 8.57e-28, 8.57e
-28, 8.57e-28, 6.51e-15, 6.51e-16
78 set Stress exponents for dislocation creep = 3.5, 4.0,
4.0, 4.0, 3.8, 3.5
79 set Activation energies for dislocation creep = 530.e3, 167e3, 223.
e3, 167e3, 440.e3, 530.e3
80 set Activation volumes for dislocation creep = 18e-6, 36e-6, 18e
-6, 18e-6, 18e-6, 18e-6
81 set Prefactors for diffusion creep = 8.88e-15, 8.88e-15, 8.88e
-15, 8.88e-15, 8.88e-15, 8.88e-15
82 set Activation energies for diffusion creep = 335e3, 375e3, 375
e3, 375e3, 335e3, 355e3
83 set Activation volumes for diffusion creep = 5.0e-6, 6.0e-6, 6.0e
-6, 6.0e-6, 6.0e-6, 6.0e-6

```

4.8. Appendix B: ASPECT input

```
84      #                               Up M., NA C., CARBC, Wk C., Wk L., Lower
85      #                               Mantle
86      set Angles of internal friction = 15, 5, 10, 5, 10, 15
87      set Cohesions                  = 20.e6, 10e4, 10.e6, 10.e4, 10.e6, 20.e6
88      end
89      subsection Multicomponent
90      set Densities                   = 3300,3300,3000,3000
91      set Specific heats              = 1250,1250,1250,1250
92      set Reference temperature      = 273
93      set Thermal conductivities     = 2.5,2.5,2.5,2.5
94      set Thermal expansivities      = 4e-5,4e-5,4e-5,4.5
95      set Viscosities                = 1.e20,1e20,1.e24,1e22
96      end
97      subsection Compositional fields
98      set Number of fields = 5
99      end
100
101     subsection Initial composition model
102     set Model name = world builder
103     end
104     subsection Gravity model
105     set Model name = vertical
106     subsection Vertical
107     set Magnitude = 9.81
108     end
109     end
110
111
112     subsection Mesh refinement
113     set Additional refinement times =
114     set Initial adaptive refinement = 2
115     set Initial global refinement  = 1
116     set Minimum refinement level   = 0
117     set Refinement fraction        = 0.95
118     set Coarsening fraction        = 0.05
119     set Strategy                   = composition, isotherms
120     set Time steps between mesh refinement = 1
121
122     subsection Composition
123     set Compositional field scaling factors = 0,0,0,0,1
124     end
125     subsection Isotherms
126     set Exclude composition = 0
127     #minref maxref mintemp maxtemp
128     set List of isotherms = max, max, 0, 1525; \
129     min-1, max, 1525, 1535; \
130     min, min, 1535, 1650; \
131     min, max, 1650, 3000
132     end
133     end
134
135
136     subsection Initial temperature model
137     set Model name = world builder
138     end
139
140
141     subsection Boundary temperature model
142     set List of model names = initial temperature fixed surface
143     subsection Initial temperature fixed surface
144     set Surface boundary indicator = 5
145     set Surface boundary temperature = 273.15
146     set Minimal temperature = 273.15
147     set Maximal temperature = 4000
```

Chapter 4. Assessing the geodynamics of strongly arcuate subduction zones: the eastern Caribbean subduction setting.

```
148   end
149 end
150
151 subsection Discretization
152   set Use locally conservative discretization = false
153   subsection Stabilization parameters
154     set Use artificial viscosity smoothing = false
155     set alpha = 2
156     set beta = 0.078
157   end
158 end
159
160 subsection Boundary temperature model
161   set Fixed temperature boundary indicators = 0,1,2,3,4,5,6,7,8,9
162 end
163
164 subsection Boundary velocity model
165   set Prescribed velocity boundary indicators = left lithosphere xyz:function,
      right lithosphere xyz:function, front lithosphere z:function, back
      lithosphere z:function
166   set Tangential velocity boundary indicators =
167   set Zero velocity boundary indicators = bottom
168 end
169
170 subsection Boundary traction model
171   set Prescribed traction boundary indicators = front lithosphere: initial
      lithostatic pressure, back lithosphere: initial lithostatic pressure, left
      : initial lithostatic pressure, right: initial lithostatic pressure, back
      : initial lithostatic pressure, front: initial lithostatic pressure
172 end
173
174
175 subsection Free surface
176   set Free surface boundary indicators = outer
177 end
178
179 subsection Boundary traction model
180   subsection Initial lithostatic pressure
181     set Representative point = 2000e3,750e3,375e3
182   end
183 end
184
185 subsection Free surface
186   set Free surface stabilization theta = 0.75
187 end
188
189 subsection Boundary velocity model
190   subsection Function
191     set Variable names = x,y,z,t
192     set Function constants = velocity=-0.0150
193     set Function expression = if(x > 1, if(y >= -300e3, velocity, (abs(y+500
      e3)/200e3)*0.25*velocity+0.75*velocity), if(y <= 750e3 && y >= 0, 0, if(
      y >= 1000e3 || y <= -50e3, velocity, if( y >= 750e3, (abs(y-750e3)/250e3
      )*velocity, (abs(y)/50e3)*velocity)));0;0
194   end
195 end
196
197
198 subsection Postprocess
199   set List of postprocessors = visualization, velocity statistics, composition
      statistics, pressure statistics, temperature statistics
200   subsection Visualization
201     set Time between graphical output = 100e3
202     set List of output variables = viscosity, density, strain rate, depth,
      nonadiabatic pressure, spd factor, shear stress eigenvectors, maximum
```

4.8. Appendix B: ASPECT input

```
203     shear stress
204     set Interpolate output = false
205     end
206 end
207 subsection Solver parameters
208     subsection Stokes solver parameters
209     set Number of cheap Stokes solver steps = 500
210     set Linear solver tolerance = 1e-1
211     end
212 end
```

Listing 4.5: The reference input file used to run the ASPECT models. For the variables 'World builder file' and 'Output directory' dummy values are set.

5

Conclusions and outlook

In this thesis I have addressed several first-order issues that aim to contribute toward more efficient and effective modeling of 3D inherently complicated natural subduction settings. These issues define first steps opening up a large scope of new research.

The research into implementing a Newton solver for ASPECT, as reported in Chapter 2, was originally intended as a relatively straight-forward development. However, we identified a fundamental problem with solving the Stokes equations via a Newton solver as is used in many other FE or FD codes. This problem is that the linearization for setting up the Newton method may lead to a singular matrix that prevents the solution of the Newton problem, while this does not occur for the usual linearization associated with Picard iteration. We have addressed this issue in an optimal manner, in the sense of fastest non-linear convergence while retaining numerical stability. This was achieved for solving the incompressible Stokes equations, however, the singularity problem remains as yet unsolved for the compressible formulation, although we expect that a similar approach in addressing the issue is feasible. Our solution to the singularity problem is not restricted to ASPECT but also applicable to other FE and FD methods. Straightforward wall-time comparison between the Picard and Newton methods is problematic, because the computation time is problem dependent and dependent on what nonlinear tolerance is adopted for the problem. I have shown that for a complex 3D case, the Newton solver was significantly faster in terms of wall-time (see Chapter 2), which has been further enhanced by later code optimization. Currently, the Newton solver requires derivative information related to the materials used. In the current ASPECT implementation some material models allow for fast analytical evaluation of derivatives whereas others require a significant amount time for computing the derivatives through finite difference. This is especially the case for the more complex material models such as visco-plasticity. Using analytic derivatives could lead also here to a considerable speed up of the Newton iteration.

In Chapter 3 I addressed the problem of how to create geometrically complex initial conditions for numerical modelling in a user-friendly, easily extendable and easily modifiable way. To this end I developed an open source code library, the Geodynamic World Builder (GWB), which allows for the creation and (Paraview) visualization of complex initial temperature and initial composition model setups. This is realized through a simple text input-file, in the JSON language, in which users provide parameters defining tectonic features and their associated temperature and compositional models. The setup can be easily changed from Cartesian to spherical coordinates and back. The library

has been designed to be easily integrated with geodynamic codes used by the geodynamic modelling community. So far, it has already been successfully integrated in two Fortran codes, ELEFANT and SEPRAN, and one C++ code, ASPECT. As shown in both Chapters 3 and 4, the present version of the GWB is already a versatile tool and ready for use but several possibilities for extensions of this library are already envisaged. Examples are parameterization of topography/bathymetry, passive margins, variable crust and lithosphere thickness, or of mantle plumes and mantle tomography models, which all will be considered in the near future. Other output possibilities can be plate velocities, for example for use as boundary conditions. At the time that the GWB library becomes integrated in more codes by the community and sees its user base grow, the amount of available features is expected to grow organically with the number of users adding their features as were required for their own project. While users and developers alike usually implement their own initial conditions for their specific setup and code, they have now the possibility to add it to a safe and backward-compatible numerical library for their later (re-)use, or for use by others.

In Chapter 4, I have used the advantages of the computational speedup and accuracy of ASPECT offered by the Newton solver and the easy construction of 3D initial conditions by the GWB, to advance research into the geodynamics of strongly arcuate subduction in which the slab may be dragged trench-parallel through the mantle by the surface motion of the subducting plate along one segment of the curved trench. The plate-tectonic evolution of the strongly arcuate eastern Caribbean subduction since 10 Ma provided the natural setting for this investigation. By varying various initial settings, boundary conditions, and parameters defining the model rheology, I demonstrated the geodynamic feasibility of trench-parallel dragging of the northern Lesser-Antilles (NLA) slab by plate motions between 1.5-3.0 cm/yr during the past 10 Myr and 5 Myr respectively, corresponding to westward lateral slab transport of 150 km, while trench-perpendicular subduction along the eastern Lesser Antilles slab remained more or less trench stationary. The actual absolute plate motion of the North American plate along the NLA trench is about 2.0 cm/yr (Dobrovine et al. (2012)) and within the modelled range. The geodynamic setting implies considerable internal deformation and shape change of the slab which occurs on spatial scales relevant for the tectonic evolution of the plate boundary zone. In the common view, the stress field in the slab is dominantly aligned along the slab-dip direction. In the geodynamic setting of arcuate subduction where the plate convergence direction strongly changes along the trench, I generally in-

ferred strong components of maximum shear stress oblique to dip-plane which are associated with the along slab-strike deformation of the slab. This may link to the occurrence of focal mechanisms of intermediate-depth and deep earthquakes with slab-strike parallel components of displacements (e.g. Christova, 2015) which cannot be explained by the classical "along-slab-dip" view on slab stress (e.g. Houston, 2015). Intermediate depth earthquakes in the LA slab also show a strong variation of stress(-drop) directions (e.g. Meighan et al., 2013). I did not further investigate the particular relation between slab dragging and focal mechanisms but regard this an important topic for future research.

The research of this chapter puts the spotlight on new possibilities and targets for numerical subduction modelling. First, it gives new possibilities for subduction modelling of complicated 3D subduction settings. The GWB facilitates relatively easy construction of elaborate 3D initial models based the paleo-tectonic setting of incipient subduction, or on more advanced stages of subduction constrained by plate tectonic reconstructions, or on the present-day plate-tectonic setting of a subduction zone for use in instantaneous dynamics modelling. Constructing alternative starting models is equally facilitated by the GWB and allows for easy implementation of alternative initial geometries for hypothesis testing. The regional subduction models I developed in my thesis, using ASPECT with the new Newton solver, required variable computation times on 480 MPI processes of 3-6 days providing model evolution times of 10 Myr, mostly depending on the rheological complexity. Computations were performed on a modern (2018) parallel cluster. This shows that even more refined models, e.g. of natural subduction, are within practical reach for which the rheological parameters of Chapter 4 may serve as a good starting point. Some material model improvements include incorporating elastic rheology and the Peierls creep deformation mechanism, and generally allowing for compressibility. Lastly, the feasible computation times and the use of the GWB may lead to making large steps toward elaborate numerical simulation of naturally occurring subduction systems in which the interplay between slab pull, slab dragging, and the viscous slab-mantle coupling is a prime target of investigation as well as how these coupled processes affect plate-tectonics and the tectonic deformation of the crust of subduction plate boundaries.

Bibliography

- Alisic, L., Gurnis, M., Stadler, G., Burstedde, C., and Ghattas, O. (2012). Multi-scale dynamics and rheology of mantle flow with plates. *J. Geophys. Res.*, 117.
- Alzetta, G., Arndt, D., Bangerth, W., Boddu, V., Brands, B., Davydov, D., Gasmöller, R., Heister, T., Heltai, L., Kormann, K., Kronbichler, M., Maier, M., Pelteret, J.-P., Turcksin, B., and Wells, D. (2018). The deal.II library, version 9.0. *Journal of Numerical Mathematics*, 26(4):173–183.
- Amante, C. and Eakins, B. (2009). Etopo1 1 arc-minute global relief model: Procedures, data sources and analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M.
- Argus, D. F., Gordon, R. G., and DeMets, C. (2011). Geologically current motion of 56 plates relative to the no-net-rotation reference frame. *Geochemistry, Geophysics, Geosystems*, 12(11).
- Bangerth, W., Hartmann, R., and Kanschä, G. (2007). deal.II – a general purpose object oriented finite element library. *ACM Trans. Math. Softw.*, 33(4):24/1–24/27.
- Baumann, T., Kaus, B., and Popov, A. (2014). Constraining effective rheology through parallel joint geodynamic inversion. *Tectonophysics*, 631:197–211.
- Billen, M. and Arredondo, K. (2018). Decoupling of plate-asthenosphere motion caused by non-linear viscosity during slab folding in the transition zone. *Phys. Earth. Planet. Inter.*, 281:17–30.

- Billen, M. and Gurnis, M. (2003). Comparison of dynamic flow models for the Central Aleutian and Tonga-Kermadec subduction zones. *Geochem. Geophys. Geosyst.*, 4(4).
- Boschman, L. M., van Hinsbergen, D. J., Torsvik, T. H., Spakman, W., and Pindell, J. L. (2014). Kinematic reconstruction of the caribbean region since the early jurassic. *Earth-Science Reviews*, 138:102 – 136.
- Brenner, S. C. and Scott, R. L. (2002). *The Mathematical Theory of Finite Elements*. Springer, Berlin-Heidelberg-New York, second edition.
- Brune, S. and Autin, J. (2013). The rift to break-up evolution of the Gulf of Aden: Insights from 3D numerical lithospheric-scale modelling. *Tectonophysics*, 607:65–79.
- Buiter, S., Schreurs, G., Albertz, M., Gerya, T., Kaus, B., Landry, W., le Pourhiet, L., Mishin, Y., Egholm, D., Cooke, M., Maillot, B., Thieulot, C., Crook, T., May, D., Souloumiac, P., and Beaumont, C. (2016). Benchmarking numerical models of brittle thrust wedges. *Journal of Structural Geology*, 92:140–177.
- Burkett, E. and Billen, M. (2010). Three-dimensionality of slab detachment due to ridge-trench collision: Laterally simultaneous boudinage versus tear propagation. *Geochem. Geophys. Geosyst.*, 11(11).
- Burstedde, C., Wilcox, L. C., and Ghattas, O. (2011). p4est: Scalable algorithms for parallel adaptive mesh refinement on forests of octrees. *SIAM Journal on Scientific Computing*, 33(3):1103–1133.
- Calais, E., Symithe, S., de Lépinay, B. M., and Prépetit, C. (2016). Plate boundary segmentation in the northeastern caribbean from geodetic measurements and neogene geological observations. *Comptes Rendus Geoscience*, 348(1):42 – 51. From past to current tectonics.
- Capitanio, F., Faccenna, C., Zlotnik, S., and Stegman, D. (2011). Subduction dynamics and the origin of Andean orogeny and the Bolivian orocline. *Nature*, 480.
- Capitanio, F., Replumaz, A., and Riel, N. (2015). Reconciling subduction dynamics during tethys closure with large-scale asian tectonics: Insights from numerical modeling. *Geochemistry, Geophysics, Geosystems*, 16(3):962–982.

- Chertova, M., Geenen, T., van den Berg, A., and Spakman, W. (2012). Using open sidewalls for modelling self-consistent lithosphere subduction dynamics. *Solid Earth*, 3:313–326.
- Chertova, M., Spakman, W., Geenen, T., van den Berg, A., and van Hinsbergen, D. (2014a). Underpinning tectonic reconstructions of the western Mediterranean region with dynamic slab evolution from 3-D numerical modeling. *J. Geophys. Res.*, 119:10.1002/2014JB011150.
- Chertova, M. V., Spakman, W., and Steinberger, B. (2018). Mantle flow influence on subduction evolution. *Earth and Planetary Science Letters*, 489:258 – 266.
- Chertova, M. V., Spakman, W., van den Berg, A. P., and van Hinsbergen, D. J. J. (2014b). Absolute plate motions and regional subduction evolution. *Geochemistry, Geophysics, Geosystems*, 15(10):3780–3792.
- Christova, C. (2015). Spatial distribution of the contemporary stress field in the kurile wadati-benioff zone by inversion of earthquake focal mechanisms. *Journal of Geodynamics*, 83:1 – 17.
- Čížková, H., van den Berg, A., Spakman, W., and Matyska, C. (2012). The viscosity of the earth's lower mantle inferred from sinking speed of subducted lithosphere. *Phys. Earth. Planet. Inter.*
- Coltice, N. and Shephard, G. E. (2017). Tectonic predictions with mantle convection models. *Geophysical Journal International*, 213(1):16–29.
- Crameri, F., Schmeling, H., Golabek, G., Duretz, T., Orendt, R., Buitert, S., May, D., Kaus, B., Gerya, T., and Tackley, P. (2012). A comparison of numerical surface topography calculations in geodynamic modelling: an evaluation of the 'sticky air' method. *Geophys. J. Int.*, 189:38–54.
- Crameri, F. and Tackley, P. J. (2015). Parameters controlling dynamically self-consistent plate tectonics and single-sided subduction in global models of mantle convection. *Journal of Geophysical Research: Solid Earth*, 120(5):3680–3706.
- Dannberg, J., Eilon, Z., Faul, U., Gassmüller, R., Moulik, P., and Myhill, R. (2017). The importance of grain size to mantle dynamics and seismological observations. *Geochem. Geophys. Geosyst.*, 18:3034–3061.
- Dannberg, J. and Gassmüller, R. (2018). Chemical trends in ocean islands explained by plume-slab interaction. *PNAS*, 115(17):4351–4356.

- Dannberg, J. and Heister, T. (2016). Compressible magma/mantle dynamics: 3-D, adaptive simulations in ASPECT. *Geophy. J. Int.*, 207:1343–1366.
- Dobrovine, P. V., Steinberger, B., and Torsvik, T. H. (2012). Absolute plate motions in a reference frame defined by moving hot spots in the pacific, atlantic, and indian oceans. *J. Geophys. Res.*, 117(B9).
- Duretz, T., Gerya, T., and Spakman, W. (2014). Slab detachment in laterally varying subduction zones: 3-D numerical modeling. *Geophys. Res. Lett.*, 41:1951–1956.
- Eisenstat, S. C. and Walker, H. F. (1996). Choosing the forcing terms in an inexact Newton method. *SIAM Journal on Scientific Computing*, 17(1):16–32.
- Fowler, C. (2005). *The Solid Earth: An Introduction to Global Geophysics*. Cambridge University Press.
- Fraters, M. R. T. (2014). Thermo-mechanically coupled subduction modelling with ASPECT. Master’s thesis, Universiteit Utrecht, the Netherlands.
- Fraters, M. R. T., Bangerth, W., Thieulot, C., and Spakman, W. (subm., 2019a). Efficient and practical newton solvers for nonlinear stokes systems in geodynamic problems. *Geophy. J. Int.*
- Fraters, M. R. T., Thieulot, C., van den Berg, A. P., and Spakman, W. (subm., 2019b). The geodynamic world builder: a solution for complex initial conditions in numerical modelling. *Solid Earth*.
- Fritzell, E., Bull, A., and Shephard, G. (2016). Closure of the Mongol-Okhotsk Ocean: Insights from seismic tomography and numerical modelling. *Earth and Planetary Science Letters*, 445:1–12.
- Gerya, T. (2010). Dynamical instability produces transform faults at mid-ocean ridges. *Science*, 329:1047–1050.
- Giardini, D. and Woodhouse, J. (1986). Horizontal shear flow in the mantle beneath the Tonga Arc. *Nature*.
- Gleason, G. and Tullis, J. (1995). A flow law for dislocation creep of quartz aggregates determined with the molten salt cell. *Tectonophysics*, 247:1–23.
- Glerum, A., Thieulot, C., Fraters, M., Blom, C., and Spakman, W. (2018). Implementing nonlinear viscoplasticity in ASPECT: benchmarking and applications to 3D subduction modeling. *Solid Earth*, 9:267–294.

- Gresho, P. and Sani, R. (2000). *Incompressible flow and the Finite Element Method, vol II*. John Wiley and Sons, Ltd.
- Guermond, J.-L., Pasquetti, R., and Popov, B. (2010). Entropy viscosity for conservation equations. *V European Conference on Computational Fluid Dynamic*.
- Guermond, J.-L., Pasquetti, R., and Popov, B. (2011). Entropy viscosity method for nonlinear conservation laws. *J. Comp. Phys.*, page doi:10.1016/j.jcp.2010.11.043.
- Gülcher, A. J., Beaussier, S. J., and Gerya, T. V. (2019). On the formation of oceanic detachment faults and their influence on intra-oceanic subduction initiation: 3d thermomechanical modeling. *Earth and Planetary Science Letters*, 506:195 – 208.
- Hall, R. and Spakman, W. (2015). Mantle structure and tectonic history of SE Asia. *Tectonophysics*, 658:14–45.
- Harris, C. W., Miller, M. S., and Porritt, R. W. (2018). Tomographic imaging of slab segmentation and deformation in the greater antilles. *Geochemistry, Geophysics, Geosystems*, 19(8):2292–2307.
- Heister, T., Dannberg, J., Gassmöller, R., and Bangerth, W. (2017). High accuracy mantle convection simulation through modern numerical methods. II: Realistic models and problems. *Geophysical Journal International*, 210(2):833–851.
- Heroux, M. and Willenbring, J. (2012). A new overview of the Trilinos project. *Scientific Programming*, 20:83–88.
- Heuret, A. and Lallemand, S. (2005). Plate motions, slab dynamics and back-arc deformation. *Physics of the Earth and Planetary Interiors*, 149(1-2):31–51.
- Hirth, G. and Kohlstedt, D. (2003). Rheology of the upper mantle and the mantle wedge: A view from the experimentalists. *in Inside the Subduction Factory, ed. J. Eiler, Geophysical Monograph American Geophysical Union, Washington, D.C.*, 138:83–105.
- Holt, A., Becker, T., and Buffett, B. (2015). Trench migration and overriding plate stress in dynamic subduction models. *Geophy. J. Int.*, 201:172–192.

- Houston, H. (2015). 4.13 - deep earthquakes. In Schubert, G., editor, *Treatise on Geophysics (Second Edition)*, pages 329 – 354. Elsevier, Oxford, second edition edition.
- Hu, J. and Liu, L. (2016). Abnormal seismological and magmatic processes controlled by the tearing south american flat slabs. *Earth and Planetary Science Letters*, 450:40 – 51.
- Jadamec, M. and Billen, M. (2010). Reconciling surface plate motions with rapid three-dimensional mantle flow around a slab edge. *Nature*, 465:338–341.
- Jadamec, M. and Billen, M. (2012). The role of rheology and slab shape on rapid mantle flow: Three-dimensional numerical models of the Alaska slab edge. *J. Geophys. Res.*, 117(B02304).
- Jadamec, M., Billen, M., and Roeske, S. (2013). Three-dimensional numerical models of flat slab subduction and the Denali fault driving deformation in south-central Alaska. *Earth Planet. Sci. Lett.*, 376:29–42.
- Karato, S. (2012). *Deformation of Earth Materials: An Introduction to the Rheology of Solid Earth*. Cambridge University Press.
- Karato, S.-I. (2008). *Deformation of Earth Materials*. Cambridge University Press.
- Karato, S.-I. and Wu, P. (1993). Rheology of the Upper Mantle: A synthesis. *Science*, 260:771–778.
- Kaus, B., Mühlhaus, H., and May, D. (2010). A stabilization algorithm for geodynamic numerical simulations with a free surface. *Phys. Earth. Planet. Inter.*, 181:12–20.
- Kaus, B., Popov, A., Baumann, T., Pusok, A., Bauville, A., Fernandez, N., and Collignon, M. (2016). Forward and Inverse Modelling of Lithospheric Deformation on Geological Timescales. *NIC Symposium 2016*, pages 299–307.
- Kelly, C. T. (1995). *Iterative methods for Linear and Nonlinear Equations*. SIAM.
- Kiraly, A., Capitanio, F, Funicciello, E, and Faccenna, C. (2016). Subduction zone interaction: Controls on arcuate belts. *Geology*.
- Knoll, D. A. and Keyes, D. E. (2004). Jacobian-free Newton-Krylov methods: a survey of approaches and applications. *Journal of Computational Physics*, 193:357–397.

- Kreemer, C., Blewitt, G., and Klein, E. (2014). A geodetic plate motion and global strain rate model. *Geochemistry, Geophysics, Geosystems*, 15(10):3849–3889.
- Kronbichler, M., Heister, T., and Bangerth, W. (2012). High accuracy mantle convection simulation through modern numerical methods. *Geophy. J. Int.*, 191:12–29.
- Lavecchia, A., Thieulot, C., Beekman, F., Cloetingh, S., and Clark, S. (2017). Lithosphere erosion and continental breakup: Interaction of extension, plume upwelling and melting. *Earth Planet. Sci. Lett.*, 467:89–98.
- Lemiale, V., Mühlhaus, H.-B., Moresi, L., and Stafford, J. (2008). Shear banding analysis of plastic models formulated for incompressible viscous flows. *Phys. Earth. Planet. Inter.*, 171:177–186.
- Leng, W. and Gurnis, M. (2015). Subduction initiation at relic arcs. *Geophys. Res. Lett.*, 42:7014–7021.
- Liu, L. and Stegman, D. (2011). Segmentation of the Farallon slab. *Earth Planet. Sci. Lett.*, 311:1–10.
- Maffione, M., Thieulot, C., van Hinsbergen, D., Morris, A., Plümpner, O., and Spakman, W. (2015). Dynamics of intraoceanic subduction initiation: 1. Oceanic detachment fault inversion and the formation of supra-subduction zone ophiolites. *Geochem. Geophys. Geosyst.*, 16:1753–1770.
- Magni, V., Bouilhol, P., and van Hunen, J. (2014). Deep water recycling through time. *Geochem. Geophys. Geosyst.*, 15:4203–4216.
- Malatesta, C., Gerya, T., Crispini, L., Federico, L., and Capponi, G. (2013). Oblique subduction modelling indicates along-trench tectonic transport of sediments. *Nature Communications*, 4.
- Mansour, J., Kaluza, O., Giordani, J., Beucher, R., Farrington, R., Kennedy, G., Moresi, L., , Mirko, A., and Sandiford, D. (2019). `underworldcode/underworld2: v2.7.0b`.
- May, D., Brown, J., and Pourhiet, L. L. (2015). A scalable, matrix-free multigrid preconditioner for finite element discretizations of heterogeneous Stokes flow. *Computer Methods in Applied Mechanics and Engineering*, 290:496–523.

- McKenzie, D. (1970). Temperature and potential temperature beneath island arcs. *Tectonophysics*, 10(1):357–366.
- Meighan, H., Pulliam, J., ten Brink, U., and López-Venegas, A. (2013). Seismic evidence for a slab tear at the Puerto Rico trench. *Journal of Geophysical Research: Solid Earth*, 118(6):2915–2923.
- Moresi, L., Betts, P., Miller, M., and Cayley, R. (2014). Dynamics of continental accretion. *Nature*, 508:245–248.
- Müller, R. D., Royer, J.-Y., Cande, S. C., Roest, W. R., and Maschenkov, S. (1999). Chapter 2 new constraints on the late cretaceous/tertiary plate tectonic evolution of the caribbean. In Mann, P., editor, *Caribbean Basins*, volume 4 of *Sedimentary Basins of the World*, pages 33 – 59. Elsevier.
- Myhill, R. (2012). Slab buckling and its effect on the distributions and focal mechanisms of deep-focus earthquakes. *Geophysical Journal International*, 192(2):837–853.
- Myhill, R. and Warren, L. (2012). Fault plane orientations of deep earthquakes in the Izu-Bonin-Marianas subduction zone. *J. Geophys. Res.*, 117(B06307).
- Naliboff, J. and Buitter, S. (2015). Rift reactivation and migration during multi-phase extension. *Earth Planet. Sci. Lett.*, 421:58–67.
- Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer Series in Operations Research. Springer, New York.
- Pikser, J., Forsyth, D., and Hirth, G. (2012). Along-strike translation of a fossil slab. *Earth Planet. Sci. Lett.*, 331–332:315–321.
- Pindell, J. and Kennan, L. (2009). Tectonic evolution of the Gulf of Mexico, Caribbean and northern South America in the mantle reference frame: an update. *Geological Society, London, Special Publications*, 328(1):1–55.
- Plunder, A., Thieulot, C., and van Hinsbergen, D. (2018). The effect of obliquity on temperature in subduction zones: insights from 3D numerical modeling. *Solid Earth*, 9:759–776.
- Pusok, A. and Kaus, B. (2015). Development of topography in 3D continental collision models. *Geochem. Geophys. Geosyst.*, 16(1378–1400).

- Pusok, A., Kaus, B., and Popov, A. (2018). The effect of rheological approximations in 3-D numerical simulations of subduction and collision. *Tectonophysics*.
- Quinquis, M. E., Buitter, S. J., and Ellis, S. (2011). The role of boundary conditions in numerical models of subduction zone dynamics. *Tectonophysics*, 497:57–70.
- Rose, I., Buffett, B., and Heister, T. (2017). Stability and accuracy of free surface time integration in viscous flows. *Phys. Earth. Planet. Inter.*, 262:90–100.
- Rudi, J., Malossi, A. C. I., Isaac, T., Stadler, G., Gurnis, M., Staar, P. W. J., Ineichen, Y., Bekas, C., Curioni, A., and Ghattas, O. (2015). An Extreme-Scale Implicit Solver for Complex PDEs: Highly Heterogeneous Flow in Earth's Mantle. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '15, pages 5:1–5:12, New York, NY, USA. ACM.
- Ruh, J. B., Gerya, T., and Burg, J.-P. (2013). High-resolution 3d numerical modeling of thrust wedges: Influence of décollement strength on transfer zones. *Geochemistry, Geophysics, Geosystems*, 14(4):1131–1155.
- Schellart, W. and Moresi, L. (2013). A new driving mechanism for backarc extension and backarc shortening through slab sinking induced toroidal and poloidal mantle flow: Results from dynamic subduction models with an overriding plate. *J. Geophys. Res.*, 118:1–28.
- Schellart, W. P. (2017). Andean mountain building and magmatic arc migration driven by subduction-induced whole mantle flow. *Nature Communications*, 8.
- Schmeling, H., Babeyko, A., Enns, A., Faccenna, C., Funicello, F., Gerya, T., Golabek, G., Grigull, S., Kaus, B., Morra, G., Schmalholz, S., and van Hunen, J. (2008). A benchmark comparison of spontaneous subduction models - Towards a free surface. *Phys. Earth. Planet. Inter.*, 171:198–223.
- Schubert, G., Turcotte, D. L., and Olson, P. (2001). *Mantle Convection in the Earth and Planets*. Cambridge University Press.
- Silvester, D. and Wathen, A. (1994). Fast Iterative Solution of Stabilised Stokes Systems. Part II: Using general block preconditioners. *SIAM J. Numer. Anal.*, 31(5):1352–1367.

- Spakman, W., Chertova, M., van den Berg, A., and van Hinsbergen, D. (2018). Puzzling features of western Mediterranean tectonics explained by slab dragging. *Nature Geoscience*, 11:211–216.
- Spiegelman, M., May, D., and Wilson, C. (2016). On the solvability of incompressible Stokes with viscoplastic rheologies in geodynamics. *Geochem. Geophys. Geosyst.*, 17:2213–2238.
- Stegman, D., Farrington, R., Capitanio, F., and Schellart, W. (2010a). A regime diagram for subduction styles from 3-D numerical models of free subduction. *Tectonophysics*, 483:29–45.
- Stegman, D., Freeman, J., Schellart, W., Moresi, L., and May, D. (2006). Influence of trench width on subduction hinge retreat rates in 3-D models of slab rollback. *Geochem. Geophys. Geosyst.*, 7(3).
- Stegman, D., Schellart, W., and Freeman, J. (2010b). Competing influences of plate width and far-field boundary conditions on trench migration and morphology of subducted slabs in the upper mantle. *Tectonophysics*, 483(1):46 – 57. Convergent plate margin dynamics: New perspectives from structural geology, geophysics and geodynamic modelling.
- Steinberger, B., Spakman, W., Japsen, P., and Torsvik, T. (2015). The key role of global solid-earth processes in preconditioning greenland’s glaciation since the pliocene. *Terra Nova*, 27(1):1–8.
- Sternai, P., Avouac, J.-P., Jolivet, L., Faccenna, C., Gerya, T., Becker, T. W., and Menant, A. (2016). On the influence of the asthenospheric flow on the tectonics and topography at a collision-subduction transition zones: Comparison with the eastern tibetan margin. *Journal of Geodynamics*, 100:184 – 197. 200 years of geodynamic modelling.
- Sternai, P., Jolivet, L., Menant, A., and Gerya, T. (2014). Driving the upper plate surface deformation by slab rollback and mantle flow. *Earth Planet. Sci. Lett.*, 405:110–118.
- Thieulot, C. (2017). Analytical solution for viscous incompressible stokes flow in a spherical shell. *Solid Earth*, 8(6):1181–1191.
- Tosi, N., Stein, C., Noack, L., Huttig, C., Maierova, P., Samuel, H., Davies, D., Wilson, C., Kramer, S., Thieulot, C., Glerum, A., Fraters, M., Spakman, W., Rozel, A., and Tackley, P. (2015). A community benchmark for viscoplastic thermal

- convection in a 2-D square box. *Geochem. Geophys. Geosyst.*, 16(7):2175–2196.
- Turcotte, D. L. and Schubert, G. (2002). *Geodynamics*. Cambridge University Press.
- Uyeda, S. and Kanamori, H. (1979). Back-arc opening and the mode of subduction. *Journal of Geophysical Research: Solid Earth*, 84(B3):1049–1061.
- van Benthem, S., Govers, R., Spakman, W., and Wortel, R. (2013). Tectonic evolution and mantle structure of the caribbean. *Journal of Geophysical Research: Solid Earth*, 118(6):3019–3036.
- van Benthem, S., Govers, R., and Wortel, R. (2014). What drives microplate motion and deformation in the northeastern caribbean plate boundary region? *Tectonics*, 33(5):850–873.
- van de Lagemaat, S., van Hinsbergen, D., Kamp, P., and Spakman, W. (2018). Southwest Pacific Absolute Plate Kinematic Reconstruction Reveals Major Cenozoic Tonga-Kermadec Slab Dragging. *Tectonics*, 37:10.1029/2017TC004901.
- van den Berg, A., Segal, G., and Yuen, D. (2015). SEPRAN: A Versatile Finite-Element Package for a Wide Variety of Problems in Geosciences. *Journal of Earth Science*, 26(1):089–095.
- van den Berg, A., Yuen, D., Umemoto, K., Jacobs, M., and Wentzcovitch, R. (2019). Mass-dependent dynamics of terrestrial exoplanets using ab initio mineral properties. *Icarus*, 317:412–426.
- van der Meer, D. G., van Hinsbergen, D. J., and Spakman, W. (2018). Atlas of the underworld: Slab remnants in the mantle, their sinking history, and a new outlook on lower mantle viscosity. *Tectonophysics*, 723:309 – 448.
- van Hunen, J. and Allen, M. B. (2011). Continental collision and slab break-off: A comparison of 3-d numerical models with observations. *Earth and Planetary Science Letters*, 302(1):27 – 37.
- van Keken, P., Currie, C., King, S., Behn, M., Cagnioncle, A., Hee, J., Katz, R., Lin, S.-C., Parmentier, E., Spiegelman, M., and Wang, K. (2008). A community benchmark for subduction zone modelling. *Phys. Earth. Planet. Inter.*, 171:187–197.

- Yamato, P., Husson, L., Braun, J., Loiselet, C., and Thieulot, C. (2009). Influence of surrounding plates on 3D subduction dynamics. *Geophys. Res. Lett.*, 36(L07303).
- Zhao, Y., de Vries, J., van den Berg, A., Jacobs, M., and van Westrenen, W. (2019). The participation of ilmenite-bearing cumulates in lunar mantle overturn. *Earth Planet. Sci. Lett.*, 511:1–11.
- Zhong, S., Zuber, M., Moresi, L., and Gurnis, M. (2000). The role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection. *J. Geophys. Res.*, 105(B5):11,063–11,082.
- Zhou, X., Li, Z.-H., Gerya, T., Stern, R., Xu, Z., and Zhang, J. (2018). Subduction initiation dynamics along a transform fault control trench curvature and ophiolite ages. *Geology*, 46(7):607–610.

My thesis aims to resolve two main outstanding problems that occur when trying to numerically simulate the 3D complexity of lithosphere subduction evolution through geological time: high computational cost to solve the associated non-linear systems accurately and the efficient constructing of complicated 3D initial conditions mimicking a subduction setting in the geological past or at the present-day.

The first issue of high computational cost is mostly due to the required use of nonlinear rheologies which are needed to realistically include temperature-, pressure- and stress dependent material properties of rock deformation. The resulting nonlinear Stokes equations require an iterative solving strategy, which can require a lot of iterations with the simple to implement Picard iteration. The Picard iteration is robust, but slow to converge. The convergence rate can be greatly improved by using a Newton iteration, which may fail to converge if the iteration solution far from the real solution, but will converge very fast when it is near the real solution. For very nonlinear systems of equations, a combination of both methods is optimal. We discovered that the linear system produced by the Newton iteration may in some cases not be solvable. We provide a general remedy for this issue by forcing the system to be solvable at a minimum amount of convergence loss. We also show that this solution works well for large 3D tectonic settings.

The second issue concerns setting up initial conditions for the geometric complexity of realistic 3D subduction settings involving creating 3D fields of initial temperature and of materials (e.g. crust, mantle, tectonic plates, faults and weakness zones, 3D slab geometry). An initial condition may concern the earliest stage of a subduction or other geodynamic system as well as an advanced stage, depending on our knowledge of the tectonic evolution. A simple approach to this, such as writing simple functions directly into the code, works well enough for 2D cases and very simple 3D cases. With increasing complexity of the 3D geodynamic setting, it becomes however practically very difficult to set up and modify the initial conditions in this way. In this thesis, I developed a more generic way for setting up initial numerical models leading to a stand-alone code library called the Geodynamic World Builder (GWB), which implements this. We show that the code library is already successfully coupled to three different geodynamic codes. The new generic approach together with the code library renders relatively easy initial-model construction as well as modification of initial settings. In addition, it creates a platform to make initial setups reproducible for use in different geodynamic codes.

In the last main chapter, I use these new developments to investigate a

complicated and realistic 3D subduction evolution that resembles the eastern Caribbean subduction setting since 10 Ma. The GWB allowed for flexible construction of the initial settings of tectonic plates, weakness zones, and 3D slab geometry at 10 Ma, while the Newton solver allowed for obtaining modelling results relatively fast, typically between 3 and 6 days depending on the rheological complexity. The modelling results show that slab dragging, i.e. lateral transport of the slab through the mantle by the trench-parallel motion of the North American plate, has a strong effect on the slab geometry and the slab stress field. Slab stress is displayed with the direction of maximum shear stress in the slab, which may align with potential fault slip directions of earthquakes. This shows that when slab dragging occurs major stress orientations are generally not trench-perpendicular, as is usually perceived for slab stress.

By this research described in the three main chapters of my thesis, I realized important new steps towards numerical simulation of natural subduction systems.

Het doel van mijn proefschrift is om twee belangrijke openstaande problemen op te lossen die voorkomen bij numeriek simuleren van complexe 3D interactie tussen de aardse tektonische platen en mantel processen: de computerkracht die het kost om de geassocieerde niet-lineaire systemen accuraat op te lossen en het efficiënt en reproduceerbaar maken van initiële condities voor het nabootsen van subductie processen in het geologische verleden of het heden.

Het eerste probleem, het probleem van de hoge rekenkosten komt voornamelijk door het noodzakelijke gebruik van niet-lineaire rheologieën. Deze rheologieën zijn nodig voor het realistisch modelleren van temperatuur, druk en stress afhankelijkheden in gesteente. De resulterende niet-lineaire Stokes vergelijkingen vereisen een iteratieve oplossingsstrategie. Deze strategie vereist in sommige gevallen echter heel veel iteraties met de standaard Picard methode, wat een iteratie is die simpel is om te implementeren maar in sommige gevallen heel langzaam naar de juiste oplossing convergeert. De convergentie snelheid kan flink versneld worden door gebruik te maken van een Newton iteratie. Deze iteratie divergeert wanneer de iteratie ver van de echte oplossing is, maar convergeert heel snel wanneer het dicht bij de werkelijke oplossing is. Voor systemen die erg niet-lineair zijn is een combinatie van beide methoden optimaal. We ontdekten echter dat het lineaire systeem dat voor de Newton iteratie opgelost dient te worden in sommige gevallen niet oplosbaar is. In hoofdstuk twee geven we een generieke oplossing voor dit probleem door het systeem te forceren om oplosbaar te blijven met een minimum hoeveelheid convergentie verlies. We laten ook zien dat dit goed werkt voor grote 3D tektonische problemen.

Het tweede probleem heeft te maken met het opzetten van initiële condities voor geometrisch complexe realistische 3D subductie zones. Dit vereist het opzetten van 3D velden van initiële temperatuur en van materialen (e.g. korst, mantel, tektonische platen, breuken en zwakte zones, 3D subducerende platen, etc.). Een initiële conditie kan de beginfase van een geodynamisch systeem behelzen, zoals een subductie systeem, maar ook latere fasen, afhankelijk van onze kennis van de tektonische evolutie. Een eenvoudige aanpak van dit probleem, zoals het schrijven van simpele functies direct in de code, werkt prima voor zowel 2D als simpele 3D gevallen. Het wordt echter met toenemende complexiteit van de 3D geodynamische setting steeds moeilijker om de initiële condities op deze manier op te zetten en aan te passen. Voor dit proefschrift heb ik een meer generieke methode voor het opzetten van initiële condities voor numerieke modellen ontwikkeld, wat heeft geleid tot een stuk software

genaamd 'the Geodynamic World Builder' or GWB, die de methode implementeert. We laten zien dat de software al succesvol gekoppeld is aan drie verschillende geodynamische modellen. Deze nieuwe generieke aanpak, samen met de software, maakt het relatief makkelijk om initiële condities voor modellen te maken en ze ook weer aan te passen. Deze aanpak zorgt ook voor een platform die het mogelijk maakt om initiële modellen op te zetten in het ene platform en reproduceerbaar te maken voor verschillende andere geodynamische modelleerplatformen.

In hoofdstuk 4 gebruik ik deze nieuwe verbeteringen om een gecompliceerde en realistische 3D subductie evolutie te onderzoeken die goed overeenkomt met het oostelijk Caribisch gebied vanaf 10 miljoen jaar geleden. De GWB maakt de flexibele constructie van initiële begin situaties van tektonische platen, zwakte zones en 3D slab geometrie mogelijk. De geïmplementeerde Newton iteratie zorgt er vervolgens voor dat de tijd die het kost om dit soort modellen voldoende accuraat voor 10 miljoen jaar door te rekenen relatief laag is. Om 10 miljoen jaar door te rekenen in een model kost meestal tussen de 3 en 6 dagen op 240 hyperthreading processors (CPU's), afhankelijk van de rheologische complexiteit. De modellen laten 'slab dragging' zien, wat een proces is waarbij de slab lateraal door de mantel getransporteerd wordt. De spanning wordt bekeken in de richting van maximale schuifspanning, die uitgelijnd kan zijn met mogelijke richtingen van breukslip van aardbevingen. Dit laat zien dat wanneer de slab door de mantel gesleept wordt, de grote schuifspanningen over het algemeen niet loodrecht op de trog staan, wat normaal gesproken gedacht wordt.

In dit proefschrift heb ik door middel van het onderzoek van de drie kernhoofdstukken belangrijke nieuwe stappen gezet naar het numeriek simuleren van natuurlijke subductie zones.

This thesis would not have been possible without the years of help and support from colleagues, collaborators, friends and of course my family. I would very much like to explicitly thank all those here.

First of all, I want to thank my promotor Wim Spakman who not only gave me the opportunity to continue doing scientific research after my master thesis but also gave me the freedom and trust to follow research directions I thought were important for the project and allowed me to set up collaborations with other universities. Always being available and willing to discuss problems, give advice and thinking up new solutions to problems has been very important to me.

Secondly, I would like to thank Cedric Thieulot (co-promotor) for involving me in the development of ASPECT for my master's thesis, and for all the years of help and support both during my master thesis and PhD. You were always there to help me when I got stuck on a problem and you gave me the confidence to try to start the collaboration to implement a Newton solver in ASPECT. Through the countless conversations, you gave me immense insight into all kinds of problems related to geodynamics and geodynamic modelling and the structure of the scientific geodynamic modelling community.

I would also like to thank Wolfgang Bangerth (promotor) for, together with Timo Heister, creating the very open and welcoming community around the open source code ASPECT, through which I was able to greatly improve and develop my skills and get to know many new amazing people. Furthermore, I would also like to thank you very much for the time, financial support and especially the patience you had with me during my several visits to your lab to implement a Newton solver in ASPECT. Your advice on the Geodynamic World Builder project was greatly appreciated.

Douwe van Hinsbergen (promotor), I would like to thank you for all the advice and help you have given me through the years. Your focus on the importance of giving good presentations has inspired me since I joined the group as a master's student. Your constant energy and flow of new ideas helped me to think outside of the box and broaden my thinking of what is possible and the thinking up new interesting questions which are still unanswered.

Timo Heister, I would like to thank you as well for creating the ASPECT community together with Wolfgang Bangerth which has been invaluable for my development. I really appreciated your help with the Newton solver and the World Builder project and have learned much. This brings me to also thank the ASPECT community as a whole, but especially the four maintainers of the project (Wolfgang Bangerth, Juliane Dannberg, Rene Gassmüller and Timo Heister) for

the welcoming community they created and continue to maintain and nourish, and also for trusting me to be a principal developer of the project. Special thanks to CIG for inviting me every year to the ASPECT hackathon which has enabled me to make the progress which resulted in this thesis.

I would also like to thank Magali Billen, for being very accommodating and helpful in the transition phase between my PhD and post-doc, Arie van den Berg for providing help with the mathematics behind the Newton solver and taking the time to implement and test the Geodynamic World Builder in SEPRAN, and Lydian Boschman for providing the data on tectonic reconstructions used in the modelling of the Caribbean region. The very professional and personal technical support by Theo van Zessen en Lukas van de Wiel is also very appreciated. Being able to just walk over to your offices and together quickly figure out problems has saved me a lot of time and stress.

I would also like to thank all my roommates for the interesting discussions and the sharing of all the frustrations and successes. Especially the weekly (or sometimes daily) brexit (frustration) updates by Pete McPhee where greatly appreciated.

My paranympths have also been invaluable over the years. Anne Glerum, when I met you I was starting my master's thesis project. We together went on an exciting journey to discover (and eventually improve) this cool brand new modelling code. Through the years you have helped and advised me, and together we had our own journal clubs and of course music playlists on the Beast. It has always been fun to share the latest tricks, failures and discoveries with you. Oscar Hover, although we had met before, our journey really started when we started as full-time board members of SIB-Utrecht (DUNSA-Utrecht). Besides thanking you for all your help, support and friendship through the years, I want to thank you for all the discussions since the beginning of our SIB-Utrecht board year and helping with setting up with what in the end would become UC-MUN. This has given me the opportunity to reflect on a broad palette of issues and to learn a set of skills otherwise difficult to obtain.

Having a supporting network outside the academic world, challenging and supporting me on a whole different level has been very important for me in finishing my PhD. This included acquiring presentation and organizational skills, learning to help and guide others, staying involved in current political discussion, forcing me every time to try and explain what I was doing for my PhD in an even more explainable way, and having very interesting non-academic projects to work on and simply relax and party. Since I strongly believe this had immense value in my academic work, I would like to thank them hereby.

Acknowledgements

To start, I would like to thank Marjolein Bomhof, Jesse Bassant, Oscar Hover and Roeland Smit for being part of the SIB-Utrecht board with me. All your inspiring ideas, great conversations, getting us all through that difficult year and the continued fun dinners, weekends trips to remote Dutch places and all your support beyond the board year is greatly appreciated.

Within SIB-Utrecht it is a common saying that the most important thing you do as a Board is making sure you have good successors. I am very proud and thankful for our successors on the Board of SIB-Utrecht. Thankful that they were willing to invest a year in the society, and proud of what they achieved in that year. This allowed me to learn and grow even more. I am also very thankful for the society in general for all the opportunities to learn skills which turned out to be useful in academics and all the great and diverse people I got to meet. Being part of the committee which organizes the weekly lecture, inviting and hosting ministers, parliamentarians, journalist and academics, was a great learning experience, and I am thankful for all the interesting discussions on the organization of lectures and current politics we had. In this regard, I would like to especially thank Miriam Saridi and Damaz de Jong for their support, perspective and collaboration on some of these great non-academic projects.

Last, but certainly not least, I would like to thank my family for their support through the years. Especially my parents and my brothers who have helped and supported me over the years and helped me overcome my learning difficulties. Without their support and encouragement, this thesis would never have been written. Thank you for always being there for me.

Funding

This work is funded by the Netherlands Organization for Scientific Research (NWO), as part of the Caribbean Research program, grant 858.14.070. This work was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 223272., the Netherlands Research Centre for Integrated Solid Earth Science (ISES), the Computational Infrastructure in Geodynamics initiative (CIG), through the National Science Foundation under Award No. EAR-0949446 and The University of California - Davis, and by the National Science Foundation under awards OCI-1148116 and OAC-1835673 as part of the Software Infrastructure for Sustained Innovation (SI2) program (now the Cyberinfrastructure for Sustained Scientific Innovation, CSSI).



Utrecht University



COMPUTATIONAL
INFRASTRUCTURE
for GEODYNAMICS



Colorado
State
University



Curriculum Vitae

3 october 1989	born in Wageningen (NL)
2002 - 2008	Pre-university education (WVO) Het Nieuwe Lyceum, Bilthoven, Netherlands
2008 - 2011	Bachelor in Earth Sciences Utrecht University, Utrecht, the Netherlands Bachelor thesis: Drainage history of the Margaritifer region on Mars
2010 - 2011	International diploma of geology of Imperial College London Imperial College London, London, UK Exchange year
2011 - 2014	Master in Earth Structure and Dynamics Utrecht University, Utrecht, the Netherlands Master Thesis: Coupling between surface and mantle with ASPECT
2014 - 2018	PhD candidate Utrecht University, Utrecht, the Netherlands
2019	Staff Research Associate UC Davis, California, USA

I started using ASPECT in 2014 for my work on Thermo-mechanically coupled subduction modelling as part of my MSc thesis (Fraters, 2014). As a Master and PhD student I was fortunate to attend 5 times, between 2014 and 2018, the CIG annual ASPECT Hackathon. At the time of writing I have contributed more than 250 times to the code (which ranks me at #1 contributor after the official 4 core maintainers). In 2017 I was granted the role of official core developer of ASPECT. This brought the responsibility of reviewing pull request (user contributions) of others developers and the right to merge them into the main code. Since I became an official developer of ASPECT, I have reviewed and merged more then 30 pull requests into ASPECT.

Implementing new features in ASPECT for my research also led me to contribute to deal.II itself: One related to a bug in the split string function and two related to adding a json input file parser. Finally, I am a co-author on two geodynamics articles published during my PhD:

1. *A community benchmark for viscoplastic thermal convection in a 2D square box*, N. Tosi, C. Stein, L. Noack C. Hüttig 2 , P. Maierova, S.C.

Kramer, C. Thieulot, A. Glerum, M. Fraters, W. Spakman, A. Rozel and P. J. Tackley, *Geochemistry, Geophysics, Geosystems*, 16, 2015.

2. *Implementing nonlinear viscoplasticity in ASPECT: benchmarking and applications to 3D subduction modeling*, A. Glerum, C. Thieulot, M. Fraters, C. Blom and W. Spakman, *Solid Earth*, 9, pp267-294, 2018.

Besides bug or documentation fixes, and the code contributions described in the following chapters, I was also a contributor to the initial topography plugin system which allows the user to prescribe an initial topography on the surface of the computational domain either from digital elevation models or from user-input polygonal parameters and the ellipsoidal geometry model which allows the construction of ellipsoidal (e.g. WGS84) meshes in ASPECT.

During my PhD I have also (co-)supervised two master thesis projects. The first project was by Constantijn Blom, with the title: 'State of the art numerical subduction modelling with ASPECT; thermo-mechanically coupled viscoplastic compressible rheology, free surface, phase changes, latent' (<https://dspace.library.uu.nl/handle/1874/348133>). The second project was by Luuk Schuurmans titled 'Numerical modelling of overriding plate deformation and slab rollback in the Western Mediterranean' (<https://dspace.library.uu.nl/handle/1874/366408>).

Finally, I have given the following scientific presentations during my PhD:

09/2014	Geomod: <i>Thermo-mechanical subduction modeling with ASPECT.</i>	Poster
01/2015	SINK Symposium: <i>From Kinematics to Dynamics in the Caribbean region.</i>	talk
03/2015	CIG Webinar: <i>From Kinematics to Dynamics in the Caribbean Region.</i>	Talk
04/2015	EGU Spring meeting: <i>Thermo-mechanically coupled subduction.</i>	talk
05/2015	CIG Research highlight: <i>Thermo-mechanically coupled subduction modeling using ASPECT.</i>	Article
08/2015	XIV International Workshop on Numerical Modelling of Mantle and Lithosphere Dynamics: <i>Two Applications of ASPECT to Geodynamics.</i>	Poster
02/2016	Computational mathematic group presentation Texas A&M: <i>From Kinematics to Dynamics in the Caribbean Region.</i>	Talk
03/2016	NAC: <i>Toward coupling deep driving processes to surface evolution in the Caribbean region.</i>	Poster
04/2017	EGU Spring Meeting: <i>Coupling Lithosphere-Mantle Dynamics to the Cenozoic Tectonic Evolution in the Caribbean Region.</i>	Poster

Curriculum Vitae

- | | | |
|---------|---|--------|
| 08/2017 | NetherMod (XV International Workshop on Numerical Modeling of Mantle and Lithosphere Dynamics): <i>Coupling Lithosphere-Mantle Dynamics to the Cenozoic Tectonic Evolution in the Caribbean Region.</i> | Poster |
| 02/2018 | Fort Collins Applied Math seminar: <i>Efficient and Practical Newton Solvers for Nonlinear Stokes Systems in Geodynamic Problems.</i> | Talk |
| 03/2018 | NAC: <i>Coupling Lithosphere-Mantle Dynamics to the Cenozoic Tectonic Evolution in the Caribbean Region.</i> | Poster |
| 04/2018 | EGU Spring Meeting: <i>Efficient and Practical Newton Solvers for Nonlinear Stokes Systems in Geodynamic Problems.</i> | Poster |
| 05/2018 | CIG Webinar: <i>Newton solver in ASPECT.</i> | Talk |

