Modelling lithospheric deformation using a compressible visco-elasto-viscoplastic rheology and the effective viscosity approach

Thibault Duretz, René Borst, Philippe Yamato

Modelling lithospheric deformation using a compressible visco-elasto-viscoplastic rheology and the effective viscosity approach

Thibault Duretz[1,2], **René de Borst**[3], **Philippe Yamato**[1,4]

[1]Univ Rennes, CNRS, Géosciences Rennes UMR 6118, F-35000 Rennes, France
[2]Institut des Sciences de la Terre, University of Lausanne, Lausanne, Switzerland
[3]University of Sheffield, Department of Civil and Structural Engineering, Sheffield S1 3JD, UK
[4]Institut Universitaire de France (IUF)

Key Points:

- The effective viscosity approach is extended to include compressible elasto-plasticity, plastic dilatancy, power-law viscoplasticity and softening.
- This extended effective viscosity approach gives results as accurate as algorithms in engineering which use return mapping and consistent linearisation.
- Applications to crustal-scale shear banding and long-term lithospheric deformations are provided.

Corresponding author: T. Duretz, thibault.duretz@univ-rennes1.fr

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2021GC009675.

Abstract

Deformations of the colder regions of the lithosphere mainly occur in the frictional regime. In geodynamic models, frictional plastic deformations are often highly localised (shear bands) and are used as proxies for faults. However, capturing the generation and evolution of shear bands in geodynamic models is troublesome. Indeed, mesh dependency and lack of convergence affect, to some extent, the results of geodynamic models. Here we extend the most common plasticity implementation used in geodynamic codes (effective viscosity approach) to include the combined effects of elasto-plastic compressibility, plastic dilatancy, strain softening and viscoplasticity. The latter acts as a regularisation that cures most of the known issues of geodynamic models related to frictional plasticity. Using regularised models based on the M2Di MATLAB routines, we show that volumetric elasto-plastic deformations can significantly impact crustal-scale shear banding. We also show that the artificial overstress caused by viscoplasticity can be mitigated by employing power-law models. Furthermore, we demonstrate that plasticity algorithms common in geodynamics (based on the effective viscosity approach) can be as accurate as those obtained with algorithms typically used in engineering (return mapping with a consistent tangent operator). Finally, we show examples of long-term tectonic deformations using the state-of-the art geodynamic code MDoodz. They indicate that viscoplastic regularisation can be used efficiently to obtain reliable simulations in geodynamics.

1 Introduction

In most geodynamic settings a large volume of the crust exhibits frictional-plastic deformations. Together with the equilibrium condition this rheological model can explain the occurrence of high-angle faults at the onset of extensional faulting and low-angle faults in compression. Moreover, this rheological model has been well calibrated by laboratory experiments of rock deformations [e.g. *Byerlee*, 1978; *Rutter and Glover*, 2012] and has further been validated in the context of deep continental drilling [*Zoback et al.*, 1993].
In order to properly capture geologically relevant stress states and structures, geodynamic models must therefore include frictional plasticity. However, accounting for such a rheological model is not trivial and remains challenging. In particular, the modelling of localised shear bands, which serve as proxies for faults, causes issues. The frictional plastic rheology becomes unstable when softening or non-associated plastic flow are included

and this can lead to the formation of localised shear bands [*Rudnicki and Rice*, 1975]. Yet, most current implementations do not incorporate an internal length scale, which is necessary to constrain the shear band to a finite width. The absence of such a length scale causes patterns of the modelled faults to be fractal [*Poliakov et al.*, 1993] and their properties (dimensions, number, stresses and strains) to depend on the numerical resolution [e.g. *Duretz et al.*, 2019]. Another consequence is that the models are numerically unstable in the sense that divergence of the equilibrium-finding iterative process is often observed [*Spiegelman et al.*, 2016; *Duretz et al.*, 2018].

To address these issues numerous regularisation schemes have been proposed. All of them introduce an internal length scale. Non-local plasticity involves a typical dimension of the area over which the plastic strain is averaged [*Bažant and Lin*, 1988]. Gradient plasticity includes the spatial gradients of the plastic strain in the yield function [*de Borst and Mühlhaus*, 1992]. Cosserat models include micro-rotations to reflect the micro-structure of the material. The bending modulus which sets the stiffness between the ensuing micro-curvatures and the couple stresses then introduces a length scale as the quotient of this bending modulus and Young's modulus has the dimension of length [*Mühlhaus and Vardoulakis*, 1987; *Stefanou et al.*, 2017; *Sabet and de Borst*, 2019]. While all of these approaches have been applied successfully, they typically require more computational power than continuum models which are not enriched, since they either require additional degrees of freedom (rotational degrees of freedom in Cosserat media) or complicated, time-consuming averaging procedures. Herein we focus on viscoplastic regularisation, which has recently been applied to problems in geodynamics [*Duretz et al.*, 2019; *Jacquey and Cacace*, 2020; *Duretz et al.*, 2020]. This regularisation approach is based on the inclusion of rate dependence of the yield function [*Wang et al.*, 1997; *de Borst and Duretz*, 2020]. Viscoplasticity thus introduces explicitly a time scale rather than a length scale [e.g. *Wang*, 2019]. The methodology is purely local and hence does not require the introduction of additional degrees of freedom at the global level. Likely, viscoplasticity is the simplest possible regularisation technique and its implementation in existing codes is fairly trivial. While it is not the most rigorous regularisation technique, especially for quasi-static (slow) process, the benefits of viscoplasticity are that a divergence of the equilibrium-finding iterative procedure is usually avoided and that mesh dependence is vastly reduced. Nevertheless, the method introduces an artificial overstress, which may affect the solution – an issue that we will thoroughly address in this paper.

In this study we provide a detailed description of the implementation of viscoplasticity for geodynamic codes that are based on the velocity-pressure formulation and on the effective viscosity approach (EVA) for the plastic rheology [e.g. *Willett*, 1992; *Moresi et al.*, 2003; *Gerya and Yuen*, 2007; *Lemiale et al.*, 2008; *May et al.*, 2014; *Kaus et al.*, 2016; *Spiegelman et al.*, 2016; *Glerum et al.*, 2018; *Naliboff et al.*, 2020]. We consider a compressible visco-elasto-viscoplastic (V-E-VP) formulation, power-law viscoplasticity and various softening laws. Both local rheological computations (local iterations, return mapping) and global computations (Newton-Raphson iterations) are explained in detail. Numerical implementation is based on the Finite Difference Method. A description of an equivalent formulation in the context of the Finite Element Method and displacement-pressure formulation can be found in *Commend et al.* [2004]. All the results shown can be reproduced using the open source MATLAB routines based on M2Di [*Räss et al.*, 2017] (https://bitbucket.org/lraess/m2di/src/master/M2Di2_VEVP_Compressible) and the geodynamic modelling code MDoodz (https://github.com/tduretz/MDOODZ6.0).

2 Model formulation

We consider steady-state deformations of a compressible V-E-VP medium, so that the balance of momentum takes the form

$$\frac{\partial \sigma_{ij}}{\partial x_j} + \rho g_i = 0, \tag{1}$$

The body force acting on the medium is due to gravity acceleration g. x contains the spatial coordinates, ρ corresponds to the density and σ is the total stress tensor. The latter relates to the deviatoric stress τ and to the pressure p via $\sigma_{ij} = p\delta_{ij} + \tau_{ij}$, with δ_{ij} the Kronecker delta. The total strain rate $\dot{\epsilon}$ relates to the deviatoric and the volumetric strain rates, $\dot{\varepsilon}$ and $v_{k,k}$, respectively, as follows: $\dot{\epsilon}_{ij} = \dot{\varepsilon}_{ij} + \frac{1}{3} v_{k,k} \delta_{ij}$, with v the velocity vector. The Einstein summation convention applies and the notation $_{,k}$ implies differentiation with respect to x_k.

The rheological model is based on the additive decomposition of the deviatoric strain rate tensor:

$$\dot{\varepsilon}_{ij} = \dot{\varepsilon}^{\mathrm{v}}_{ij} + \dot{\varepsilon}^{\mathrm{e}}_{ij} + \dot{\varepsilon}^{\mathrm{vp}}_{ij}, \tag{2}$$

where the superscripts v, e and vp stand for elastic, viscous and viscoplastic, respectively. We consider isotropic power-law creep, linear isotropic elasticity and viscoplasticity such

that the additive decomposition can be expressed as:

$$\dot{\varepsilon}_{ij} = \frac{\tau_{ij}}{2\eta^{\mathrm{v}}} + \frac{\dot{\tau}_{ij}}{2G} + \dot{\lambda}\frac{\partial Q}{\partial \tau_{ij}}, \tag{3}$$

where η^{v} is the effective creep viscosity, G is the shear modulus, $\dot{\lambda}$ is the rate of the plastic multiplier and Q is the plastic flow potential. The effective power-law creep viscosity is formulated as:

$$\eta^{\mathrm{v}} = \eta_0^{\mathrm{v}}\ \dot{\varepsilon}_{\mathrm{II}}^{\mathrm{v}\ \frac{1}{n^{\mathrm{v}}}-1}, \tag{4}$$

where η_0^{v} and n^{v} are material parameters which can be calibrated using laboratory experiments. $\dot{\varepsilon}_{\mathrm{II}}^{\mathrm{v}}$ is the second invariant of the viscous part of the deviatoric strain rate tensor defined as $\dot{\varepsilon}_{\mathrm{II}}^{\mathrm{v}} = \sqrt{\frac{1}{2}(\dot{\varepsilon}_{xx}^{\mathrm{v}\ 2} + \dot{\varepsilon}_{yy}^{\mathrm{v}\ 2} + \dot{\varepsilon}_{zz}^{\mathrm{v}\ 2}) + \dot{\varepsilon}_{xy}^{\mathrm{v}\ 2}}$. Here we consider the case of plane strain deformation, hence the out-of-plane deviatoric strain rate and deviatoric stress components do not vanish. Henceforth, we will consider a viscoplastic Drucker-Prager model with the yield function

$$F = \tau_{\mathrm{II}} - p\sin\phi - c\cos\phi - \dot{\lambda}\eta^{\mathrm{vp}}, \tag{5}$$

where τ_{II} is the second invariant of the deviatoric stress tensor defined as $\tau_{\mathrm{II}} = \sqrt{\frac{1}{2}(\tau_{xx}{}^2 + \tau_{yy}{}^2 + \tau_{zz}{}^2) + \tau_{xy}{}^2}$, c is the cohesion, ϕ is the friction angle and η^{vp} is the viscoplastic viscosity. The parameters of the Drucker-Prager plasticity model are typically calibrated using experimental data [e.g. *Byerlee*, 1978; *Rutter and Glover*, 2012]. However classical local and rate-independent non-associated plasticity models cause mesh-dependence of numerical solutions. We employ a viscoplastic model to reduce the effects of mesh dependence during non-associated plastic flow. For generality, we express the viscoplastic viscosity using a power-law relation, so that:

$$\eta^{\mathrm{vp}} = \eta_0^{\mathrm{vp}}\ \dot{\lambda}^{\frac{1}{n^{\mathrm{vp}}}-1}, \tag{6}$$

where η_0^{vp} and n^{vp} are material parameters. In this contribution, we do not aim to relate the parameters of the viscoplastic model to experimental rock deformation data. Viscoplasticity is here only used for its time regularisation effect.

Compared to inviscid plasticity models viscoplastic models introduce a rate-dependent overstress, $\bar{\sigma}$ [e.g. *Heeres et al.*, 2002; *Niazi et al.*, 2013; *de Borst and Duretz*, 2020]. The use of a power-law exponent in excess of 1 can significantly reduce overstress variations induced by changes in the magnitude of viscoplastic strain rate, see Figure (1A). Hardening/softening laws are defined for the variables that control the evolution of plastic-

ity:

$$\xi = \xi_0 - \frac{\Delta_\xi}{2} \operatorname{erfc}\left(-\frac{\varepsilon^{\mathrm{vp}} - \mu_\xi}{\sigma_\xi}\right), \tag{7}$$

where ξ represents either the cohesion, friction or dilatancy angle. ξ_0 indicates the initial value of ξ, while Δ_ξ, μ_ξ and σ_ξ correspond to the amplitude, the mean and the standard deviation of the prescribed variation of ξ, respectively. σ_ξ controls the rate of softening. Its effect on the variation of ξ with viscoplastic strain is depicted in Figure (1B). Such a non-linear hardening/softening law differs from the piecewise-linear laws which are commonly used in geodynamics. The advantage is that it does not include singularities and is continuously differentiable, which is very suitable in implicit rheological modelling using Newton-Raphson linearisation.

The accumulated viscoplastic strain tensor is expressed as

$$\varepsilon^{\mathrm{vp}} = \int \sqrt{\frac{2}{3}\frac{\partial Q}{\partial \tau_{ij}}^T \frac{\partial Q}{\partial \tau_{ij}}} \mathrm{d}\varepsilon^{\mathrm{vp}} \equiv \int h \mathrm{d}\varepsilon^{\mathrm{vp}}. \tag{8}$$

The plastic flow potential is defined as:

$$Q = \tau_{\mathrm{II}} - p \sin\psi, \tag{9}$$

where ψ is the dilatancy angle. This plasticity model is non-associated ($\frac{\partial F}{\partial p} \neq \frac{\partial Q}{\partial p}$) unless ψ equals ϕ. Plastic flow only occurs if $F \geq 0$. Then, the rate of the plastic multiplier is positive.

The volumetric rheological model is based on an additive decomposition of the divergence of velocity:

$$v_{k,k} = v^{\mathrm{e}}_{k,k} + v^{\mathrm{vp}}_{k,k}. \tag{10}$$

Thus, we assume that volumetric deformation is caused by either elasticity or viscoplasticity, and that volumetric viscous creep is excluded. This leads to:

$$v_{k,k} = -\frac{\dot{p}}{K} - \dot{\lambda}\frac{\partial Q}{\partial p}, \tag{11}$$

where K stands for the bulk modulus. When $\psi = 0$, the term $\frac{\partial Q}{\partial p} = -\sin\psi$ vanishes and the model reaches viscoplastic incompressibility. The fully incompressible limit $v_{k,k} = 0$ is obtained by assuming elastic incompressibility and setting the dilatancy angle to zero.

3 Numerical implementation

The momentum balance has been discretised using the staggered grid finite-difference method. The components of the velocity vector (v_i) and the pressure (p) are considered

as the primitive variables (i.e. a velocity-pressure formulation). The determination of velocity and pressure fields that satisfy momentum balance equation is a global procedure that requires the solution of a system of equations. Moreover, due to non-linear rheological models, this procedure is iterative and requires successive global non-linear iterations. The deviatoric rheological implementation is based on the effective viscosity approach (EVA), which is used in most geodynamic codes. We use a predictor-corrector procedure whereby the trial visco-elastic stress is computed at each global iteration and corrected in case of viscoplastic flow. Both the predictor and corrector steps involves non-linearities. The considered rheological model is local as, e.g., the yield function does not dependent on the gradients of the plastic strain. Hence all rheological computations are local procedures in the sense that there are applied to each cell/vertex or integration points independently.

3.1 Local rheological procedures: non-linear visco-elastic predictor

The stress rate is integrated using a backward Euler scheme, $\tau_{ij} = \tau^0_{ij} + \dot{\tau}_{ij}\Delta t$, with τ^0 the deviatoric stress tensor at the previous time step and Δt is the time step. We now define $\eta^{\mathrm{e}} = G\Delta t$. Then, Eq. (3) can be integrated to yield:

$$\tau_{ij} = 2\eta^{\mathrm{ve}} \left(\dot{\varepsilon}'_{ij} - \dot{\lambda} \frac{\partial Q}{\partial \tau_{ij}} \right) \tag{12}$$

where $\eta^{\mathrm{ve}} = \left(\frac{1}{\eta^{\mathrm{v}}} + \frac{1}{\eta^{\mathrm{e}}} \right)^{-1}$ and $\dot{\varepsilon}'_{ij} = \dot{\varepsilon}_{ij} + \frac{\tau^0_{ij}}{2\eta^{\mathrm{e}}}$ is an effective deviatoric strain rate tensor that accounts for the time discretisation of the stress rate [e.g. *Moresi et al.*, 2003; *Kaus et al.*, 2016; *Bauville et al.*, 2020]. Similarly, we define the effective divergence $v'_{k,k} = v_{k,k} - \frac{p^0}{K\Delta t}$, where p^0 is the pressure from the previous time step.

For a non-linear viscous creep model there is no closed-form expression for η^{ve} and a local non-linear iteration is needed [*Popov and Sobolev*, 2008]. An exact additive partitioning of the elastic and viscous deviatoric strain rates is ensured by the non-linear identity:

$$f(\eta^{\mathrm{ve}}) = \dot{\varepsilon}'_{\mathrm{II}} - \frac{\tau_{\mathrm{II}}}{2\eta^{\mathrm{e}}} - \dot{\varepsilon}^{\mathrm{v}}_{\mathrm{II}} = 0, \tag{13}$$

where $\dot{\varepsilon}'_{\mathrm{II}}$ is the second invariant of $\dot{\varepsilon}'_{ij}$, $\tau_{\mathrm{II}} = 2\eta^{\mathrm{ve}}\dot{\varepsilon}'_{\mathrm{II}}$ is the second invariant of the deviatoric stress tensor and $\dot{\varepsilon}^{\mathrm{v}}_{\mathrm{II}} = C^{\mathrm{v}}\tau^{n_{\mathrm{v}}}_{\mathrm{II}}$ is the viscous strain rate with $C^{\mathrm{v}} = (2\eta^{\mathrm{v}}_0)^{-n_{\mathrm{v}}}$.

A local Newton-Raphson scheme can be used to solve $f(\eta^{\mathrm{ve}}) = 0$. The effective visco-elastic viscosity is then determined iteratively by incrementing successive correc-

tions

$$\delta\eta^{\mathrm{ve}} = -\left(\frac{\partial f}{\partial \eta^{\mathrm{ve}}}\right)^{-1} f(\eta^{\mathrm{ve}}), \tag{14}$$

where

$$\frac{\partial f}{\partial \eta^{\mathrm{ve}}} = -2\frac{\dot{\varepsilon}'_{\mathrm{II}}}{2\eta^{\mathrm{e}}} - C^{\mathrm{v}} n^{\mathrm{v}} \frac{\tau_{\mathrm{II}}}{\eta^{\mathrm{ve}}}. \tag{15}$$

This procedure typically converges to machine precision in less than five iterations. The algorithm is strain-rate driven and $\dot{\varepsilon}'_{\mathrm{II}}$ is constant during these local iterations.

The viscous creep model does not include volumetric deformation. Hence, viscous creep is assumed to be purely deviatoric and does not feed back on the pressure field. The inclusion of a composite creep model based on different flow law expressions (exponential creep, grain-size evolution) is straightforward [e.g. *Popov and Sobolev*, 2008; *Schmalholz and Duretz*, 2017; *Bessat et al.*, 2020]. The resulting deviatoric stress serves as a deviatoric trial stress for the subsequent plasticity computations ($\tau_{\mathrm{II}}^{\mathrm{trial}} = \tau_{\mathrm{II}}$).

3.2 Local rheological procedures: non-linear viscoplastic corrector

Once the deviatoric trial stress has been determined, the condition for yielding,

$$F^{\mathrm{trial}} = \tau_{\mathrm{II}}^{\mathrm{trial}} - p^{\mathrm{trial}} \sin\phi - c\cos\phi \tag{16}$$

can be evaluated. As discussed, $\tau_{\mathrm{II}}^{\mathrm{trial}}$ is determined in the predictor stage, while p^{trial} is extracted from the global solution vector. The trial yield function assumes no viscoplastic flow, hence $\dot{\lambda} = 0$ at this stage. For viscoplastic yielding ($F^{\mathrm{trial}} \geq 0$), the rate of plastic multiplier ($\dot{\lambda}$) needs to be evaluated, which will be used to determine the deviatoric viscoplastic strain rate tensor, $\dot{\varepsilon}_{ij}^{\mathrm{vp}} = \dot{\lambda}\frac{\partial Q}{\partial \tau_{ij}} = \dot{\lambda}\left[\frac{\tau_{xx}}{2\tau_{\mathrm{II}}} \frac{\tau_{yy}}{2\tau_{\mathrm{II}}} \frac{\tau_{zz}}{2\tau_{\mathrm{II}}} \frac{\tau_{xy}}{\tau_{\mathrm{II}}}\right]^T$ and the volumetric viscoplastic strain rate, $v_{k,k}^{\mathrm{vp}} = -\dot{\lambda}\frac{\partial Q}{\partial p} = \dot{\lambda}\sin\psi$. Viscoplastic flow then implies a reduction of the deviatoric stress such that:

$$\tau_{\mathrm{II}}^{\mathrm{corr}} = \tau_{\mathrm{II}}^{\mathrm{trial}} - \eta^{\mathrm{ve}}\dot{\lambda}. \tag{17}$$

The individual corrected deviatoric stress components may be evaluated as:

$$\tau_{ij}^{\mathrm{corr}} = 2\eta^{\mathrm{ve}}\left(\dot{\varepsilon}'_{ij} - \dot{\varepsilon}_{ij}^{\mathrm{vp}}\right) = \tau_{ij}^{\mathrm{trial}} - 2\eta^{\mathrm{ve}}\dot{\lambda}\frac{\partial Q}{\partial \tau_{ij}}. \tag{18}$$

If elasto-plastic volume changes are considered, viscoplastic flow feeds back into the pressure. The pressure update then includes a correction proportional to the amount of viscoplastic volume change:

$$p^{\mathrm{corr}} = p^{\mathrm{trial}} + \dot{\lambda} K \Delta t \sin\psi \tag{19}$$

In the incompressible limit, the pressure can be interpreted as a Lagrange multiplier that enforces incompressibility. Hence, a pressure update may not be explicitly written and the rheological model does not require any local pressure corrections. The accumulated viscoplastic strain is updated with the following increment:

$$\Delta\varepsilon_{\mathrm{II}}^{\mathrm{vp}} = h\dot{\lambda}\Delta t \tag{20}$$

The corrected yield function is thus expressed as

$$F = \tau_{\mathrm{II}}^{\mathrm{trial}} - \eta^{\mathrm{ve}}\dot{\lambda} - p^{\mathrm{corr}}\sin\phi - c\cos\phi - \eta^{\mathrm{vp}}\dot{\lambda} \tag{21}$$

For ideal, linear viscoplasticity (so without hardening or softening), setting $F = 0$ yields the following expression of plastic multiplier rate [*de Borst and Feenstra*, 1990]:

$$\dot{\lambda} = \frac{F^{\mathrm{trial}}}{\eta^{\mathrm{ve}} + \eta^{\mathrm{vp}} + K\Delta t\sin\psi\sin\phi}. \tag{22}$$

However, as soon as either η^{vp}, c, ϕ or ψ involve non-linear expressions, the determination of $\dot{\lambda}$ necessitates non-linear iterations. This procedure is again achieved via local Newton-Raphson iterations, where corrections to the rate of the plastic multiplier are expressed as:

$$\delta\dot{\lambda} = -\left(\frac{\partial F}{\partial\dot{\lambda}}\right)^{-1} F(\dot{\lambda}), \tag{23}$$

with

$$\frac{\partial F}{\partial\dot{\lambda}} = -K\Delta t\sin\psi\sin\phi - \eta^{\mathrm{ve}} - \frac{\eta^{\mathrm{vp}}}{n^{\mathrm{vp}}} - H, \tag{24}$$

where $H = \frac{\partial c}{\partial\dot{\lambda}}\cos\phi - \frac{\partial\phi}{\partial\dot{\lambda}}\left(c\sin\phi - p^{\mathrm{corr}}\cos\phi\right) + K\Delta t\dot{\lambda}\cos\psi\sin\phi\frac{\partial\psi}{\partial\dot{\lambda}}$. The partial derivatives of c, ϕ or ψ are of the form:

$$\frac{\partial\xi}{\partial\dot{\lambda}} = -\frac{h\Delta t\Delta_\xi}{\sqrt{\pi}\sigma_\xi}\exp\left[-\frac{\left(\mu_\xi - \varepsilon_{\mathrm{II}}^{\mathrm{pl}}\right)^2}{\sigma_\xi^2}\right] \tag{25}$$

where ξ is either c, ϕ or ψ.

Upon convergence, the stress update is be formulated as:

$$\tau_{ij} = 2\eta^{\mathrm{ve}}\left(\dot{\varepsilon}'_{ij} - \dot{\lambda}\frac{\partial Q}{\partial\tau_{ij}}\right) \equiv 2\eta^{\mathrm{vep}}\dot{\varepsilon}'_{ij} \tag{26}$$

where

$$\eta^{\mathrm{vep}} = \frac{p^{\mathrm{corr}}\sin\phi + c\cos\phi + \eta^{\mathrm{vp}}\dot{\lambda}}{2\dot{\varepsilon}'_{\mathrm{II}}}. \tag{27}$$

The effective visco-elasto-viscoplastic viscosity can then be used for the resolution of the global non-linear problem. Implementation details regarding local rheological computations (predictor and corrector phases) can be found in the Supporting Material (Code S1).

3.3 Global Newton-Raphson iterations

The activation of non-linear rheological elements (e.g. power-law creep, frictional plasticity) introduces a non-linearity at the global level. In order to reach global equilibrium, successive global Newton-Raphson iterations are applied. The corrections to the velocity and the pressure are given by:

$$\begin{bmatrix} \delta v \\ \delta p \end{bmatrix} = \boldsymbol{J}^{-1} \begin{bmatrix} f_v \\ f_p \end{bmatrix}. \tag{28}$$

where $\boldsymbol{J}$ represents the Jacobian matrix and f_v, f_p are the residuals, defined as:

$$\begin{aligned} f_{v_i} &= \frac{\partial \tau_{ij}}{\partial x_j} - \frac{\partial p}{\partial x_i} + \rho g_i, \\ f_p &= -\frac{\mathrm{d}p}{\mathrm{d}t} - K\left(v_{k,k} - v_{k,k}^{\mathrm{vp}}\right). \end{aligned} \tag{29}$$

It should be noted that the evaluation of the above residual equations requires knowledge of corrected deviatoric stress components and corrected pressure in case of viscoplastic deformation (see Sec 3.2). If p^{trial} is considered as the global pressure variable, the time discretized residuals are expressed as:

$$\begin{aligned} \tilde{f_{v_i}} &= \frac{\partial \tau_{ij}}{\partial x_j} - \frac{\partial p^{\mathrm{trial}}}{\partial x_i} - \frac{\partial K \Delta t v_{k,k}^{\mathrm{vp}}}{\partial x_i} + \rho g_i \\ \tilde{f}_p &= -\frac{p^{\mathrm{trial}} - p^0}{\Delta t} - K v_{k,k}, \end{aligned} \tag{30}$$

where p^0 is the pressure from the preceding time increment and Δt is the time step. Alternatively one may define p^{corr} as the global pressure variable, in this case the time-discretized residuals are expressed as:

$$\begin{aligned} \tilde{f_{v_i}} &= \frac{\partial \tau_{ij}}{\partial x_j} - \frac{\partial p^{\mathrm{corr}}}{\partial x_i} + \rho g_i, \\ \tilde{f}_p &= -\frac{p^{\mathrm{corr}} - p^0}{\Delta t} - K\left(v_{k,k} - v_{k,k}^{\mathrm{vp}}\right). \end{aligned} \tag{31}$$

The assembly of the Jacobian matrix requires the evaluation of the tangential operators, $\frac{\partial \tau_{ij}}{\partial \dot{\varepsilon}_{kl}}$ and $\frac{\partial p}{\partial v_{k,k}}$, at each cell or integration point. The latter can be obtained either via a consistent linearisation of the finite-step visco-elasto-viscoplastic relation (as customary in computational engineering), or by explicitly evaluating the derivatives of the effective viscosity η^{vep}, which is commonly done in computational geodynamics. Herein, we have tested both approaches and the corresponding analytical expressions are given in the Appendices A and B, for the Newton linearisation of the effective viscosity approach and the finite-step consistent tangent linearisation, respectively. It is emphasised that, when the linearisation and differentiation are done correctly, both approaches result in

the same Jacobian, and hence, the convergence behaviour is exactly the same (see Fig. 2 and Supporting Material Figure S8).

At each global iteration solving for the correction vector requires the application of the inverse of the Jacobian matrix to the current residual vector. To this end, we use a sparse direct factorisation based on UMFPACK [*Bates*, 2007] in the M2Di examples. For the MDOODZ applications, we rely on a direct-iterative scheme involving Powell and Hestenes iterations, Generalised Conjugate Residual iterations and pre-conditioning using Cholesky factorisation of the symmetrised Jacobian [e.g. *Räss et al.*, 2017]. These techniques require the explicit assembly of the Jacobian matrix and it is thus necessary to compute the partial derivatives of the momentum and continuity equations with respect to the velocities.

3.4 The choice of parameters for viscoplastic regularisation

The viscoplastic rheological model relies on the inclusion of a rate-dependent viscous component in the yield function. The latter causes an extra stress, typically coined the overstress $\bar{\sigma}$, compared to rate-independent models. The magnitude of the overstress depends on the strain rate can be expressed as:

$$\bar{\sigma} = \eta^{\mathrm{vp}}\dot{\lambda} = \eta_0^{\mathrm{vp}}\ \dot{\lambda}^{\frac{1}{n^{\mathrm{vp}}}}\ . \tag{32}$$

For a linear viscoplastic model, the overstress depends linearly on the rate of plastic multiplier. Large variations of strain rate result in proportional variations of the overstress. This effect can be mitigated by using power-law models, which reduces the magnitude of the overstress. For example, Fig. (1A) shows variations of the overstress for different values of the stress exponent, n^{vp}. A linear model predicts a variation of overstress of an order of magnitude for a variation of the strain rate of an order of magnitude. By contrast, a power law model using $n^{\mathrm{vp}} = 2.0$ predicts only half of this overstress. In the following model we compute the reference viscosity factor ($\mathrm{Pa}\cdot\mathrm{s}^{n^{\mathrm{vp}}}$) by defining a reference overstress $\hat{\sigma}^{\mathrm{ref}}$ for a given reference value of strain rate ($\dot{\lambda} = \dot{\varepsilon}^{\mathrm{ref}}$). For example, we set $\bar{\sigma}^{\mathrm{ref}} = 1$ MPa for $\dot{\varepsilon}^{\mathrm{ref}} = 1\times10^{-15}\ \mathrm{s}^{-1}$ as in *Duretz et al.* [2020]. The reference viscosity factor can thus be written as a function of the reference overstress, the reference strain rate and the stress exponent:

$$\eta_0^{\mathrm{vp}} = \bar{\sigma}^{\mathrm{ref}}\ \dot{\varepsilon}^{\mathrm{ref}\,-\frac{1}{n^{\mathrm{vp}}}}\ . \tag{33}$$

In the linear case ($n^{\text{vp}} = 1.0$), such as in *Duretz et al.* [2020], the viscosity coefficient would be equal to 10^{21} Pa·s for the above stated reference overstress and strain rate. In practice, we set the reference strain rate equal to that of the bulk strain rate applied to the model boundaries. Variations of the reference overstress (or viscoplastic viscosity) will influence the width of shear bands [*Duretz et al.*, 2019] and thus will have an impact the evolution of geodynamic models. We provide some examples in the following sections as well as in the Supporting Material (Figure S3).

4 Results

4.1 Comparison with previously published results

We first check whether the above model can successfully capture visco-elasto-(visco)plastic shear banding. To this end, we compare the results of 3 shear banding simulations to previously published results. The latter simulations were obtained using a different model formulation (displacement-based) and linearisation technique (consistent tangent linearisation), but with the same spatial and temporal discretisation (staggered grid finite differences, backward Euler). The first simulation uses an elasto-plastic (E-P) rheology and corresponds to *Test 1* of *Duretz et al.* [2018] (Fig. 2A). The second test accounts for deviatoric viscous creep (V-E-P rheology) and corresponds to *Test 4* of *Duretz et al.* [2018] (Fig 2B). The third test accounts for an elasto-viscoplastic rheology (E-VP) and was presented in *Duretz et al.* [2019] as *Model 1* (Fig 2C). All tests include bulk elasticity and plastic dilation. For each test, we report the evolution of the minimum, the maximum and the mean value of the second deviatoric stress invariant in the domain. In all cases excellent agreement between results obtained with the different formulations is obtained. We conclude that models based on the effective viscosity approach with a correct differentiation and a velocity-pressure formulation capture visco-elasto-(visco)plastic shear banding as accurately as models based on the displacement-based formulations and using a consistent tangent linearisation. Moreover we have also derived a finite-step consistent tangent linearisation suitable for the velocity-pressure formulation (Appendix A). We could show that this approach delivers similar non-linear convergence than the Newton linearisation of the effective viscosity approach (see Supporting Material, Figures S7 and S8).

4.2 Crustal shear banding

We now apply the rheological model to model shear banding at the crustal scale (Fig. 3 to 6). The design of the experiments is similar to that in *Duretz et al.* [2020]. The model domain has a width of 100 km and a height of 30 km, and accounts for gravity acceleration. The crust is represented by a V-E-VP rheology and accounts for the Westerly granite flow parameters, constant shear and bulk modulus, and a viscoplastic pressure-dependent flow rule. The temperature varies linearly between 20°C and 466°C from top to bottom. The only difference with *Duretz et al.* [2020] is that we consider the top boundary to be a free surface. The simulations have been carried out in compression with a constant rate of 10^{-15} s^{-1}. We have assumed small strains and thus neither advection, nor rotation are taken into account. The models were run with a resolution of 404×124 cells up to a final time of 1.2811 My (101 time steps of 4×10^{11} s). Non-linear iterations were performed until the norm of momentum residuals dropped below 5×10^{-12}. The results can be reproduced using the corresponding M2Di MATLAB routines.

4.2.1 Bulk elastic deformation and plastic dilation

The assumption of incompressible deformations is often made in geodynamic modelling. We now investigate the role of elastic compressibility as well as that of plastic dilation on the patterns of crustal scale shear banding. Fig. 3A shows shear banding patterns obtained with an incompressible V-E-VP formulation [*Duretz et al.*, 2020]. Including elastic bulk deformation (bulk modulus $K = 5\times10^{10}$ Pa) significantly changes the shear banding pattern. The strain rates tend to concentrate within the main shear band and secondary shear bands tend to disappear. The effect of shear banding on the pressure field is still noticeable, and strong pressure gradients across shear bands are preserved (Fig. 3B). The inclusion of a constant plastic dilatancy angle ($\psi = 10°$) smears out most of the shear bands observed in the incompressible counterpart. The main shear bands are broader, hence both the intensity of deformation and pressure variations are attenuated (Fig. 3C).

The effect of varying the bulk modulus while keeping the dilatancy angle constant is shown in Fig. 4. Panels A to C show a variation of K within the range 10^{10} – 10^{11} Pa. The progressive increase of K promotes strain localisation and shows that plastic strain localisation is favoured as elastic incompressibility is approached. On the other

hand, models with ψ ranging between 2.5 and 7.5 are shown in panels D to F. The results further confirm that plastic strain localisation is more likely to occur for low values of the dilatancy angle. The magnitude of the deviatoric stress (max. 500 MPa) and the depth of the brittle-ductile transition (-17 km) are not much affected by variations of the bulk modulus and the dilatancy angle. We conclude that shear banding is promoted by either elastic or plastic incompressibility. Strain localisation is the most intense in the limiting case for both elastic and plastic incompressibility (Fig. 3A).

In general we observe that simulations in which the effects of bulk elasticity and plastic dilation are included exhibit a better global non-linear convergence than incompressible simulations. This is also linked to the fact that both elastic and plastic incompressibility promote shear banding, which renders the iteration procedure more challenging (see Supporting Information Figure S11).

4.2.2 Power-law viscoplasticity

We have also explored the impact of different values of the power-law viscoplasticity exponent on the patterns of shear banding. All models are designed such that the reference overstress is 1 MPa. In the reference model ($K = 5\times10^{10}$ Pa, $\psi = 10°$), the shear bands are wide and the overstress rises above 10 MPa (Fig. 3B, Fig. 5A). This is because the strain rate in the shear band is one order larger in magnitude than the background value and the viscoplastic model is linear ($n^{\mathrm{vp}} = 1.0$). Setting the power-law exponent to 1.5 locally reduces the overstress to maximum values of about 10 MPa and amplifies strain localisation (Fig. 5B). Increasing the exponent to 2.0 further decreases the width of the shear bands and reduces the overstress to maximum values of about 5 MPa (Fig. 5C). By increasing the stress exponent, models tend towards the rate-independent plastic limit. Moderate values of the stress exponent (¡3.0) are hence recommended to benefit from the advantages of viscoplasticity (global non-linear convergence, regularisation) while keeping moderate values of overstress.

4.2.3 Strain softening

Material strain softening is often used to trigger strain localisation in the frictional domain [*Huismans and Beaumont*, 2003; *Naliboff et al.*, 2017, e.g.]. We have tested the effect of softening on the cohesion, the friction and the dilatancy angle for the V-E-VP

class of models. Implicit and explicit implementations were tested (Fig. 6). In the implicit approach plastic variables are updated during the non-linear global iterations. In the explicit approach, plastic variables are updated once per time step, i.e. after the global non-linear iterations. Models with cohesion strain softening exhibit the largest sensitivity with regard to the type of implementation, Fig. 6A,B. The model with implicit strain softening features numerous secondary shear bands, Fig. 6B. In both cases the minimum cohesion (0 MPa) is reached within the shear bands. For friction softening, we have only considered a small decrease of the friction angle (5°). This ensures a stable implicit time integration at each time step without requiring any form of adaptative time stepping. Both implementations deliver very similar shear banding patterns with most reduction of the friction angle occurring in the shear bands, Fig. 6C,D. For dilation softening, we also allowed for a 5° reduction of the dilatancy angle. Whatever the implementation, model results are virtually similar (Fig. 6E,F) and very much resemble the reference model, Fig. 3B. A large part of the plastic region is not affected by shear localisation. Reduction of the dilatancy angle is also observed where shear bands do not develop. In all cases we observe that 'thick' shear bands (i.e. several elements/cells wide) develop. This indicates that the regularising properties are preserved when strain softening is considered.

5 Application to lithosphere dynamics simulations

In order to demonstrate the practical use of the V-E-VP model, we have run long-term lithospheric deformation simulations (Fig. 7 to 10). The V-E-VP model was implemented in the thermo-mechanical code MDoodz that can handle tectonics deformations [*Duretz et al.*, 2016a; *Kiss et al.*, 2019; *Poh et al.*, 2020; *Candioti et al.*, 2020] and accounts for composite rheological models [*Yamato et al.*, 2019; *Bessat et al.*, 2020] and a true free surface [*Duretz et al.*, 2016b]. The model domain has a 300 km width and 100 km height. The crust consists of three layers (15 km thick upper crust, 10 km thick middle crust and 10 km thick lower crust). To seed the deformation, a weak elliptical inclusion is located within the lower crust. Its orientation will trigger either symmetric or asymmetric deformation patterns (10 km long axis, 2 km short axis, with a 0° dip for the symmetric case and a 30° dip to the left for the asymmetric case). The underlying mantle is represented by a single material phase. For the mechanical problem, we consider constant normal velocities at the left (± 0.5 cm/y), right (± 0.5 cm/y) and bottom (± 0.31 cm/y) faces of the model, and a free surface at the top. The transient thermal problem

was initialised with an equilibrated temperature field (shear and adiabatic heating are neglected in the initialisation step) and accounts for zero fluxes at the left and right sides, and constant temperatures at the top (0° C) and at the bottom (1450 ° C). We have used a linear viscoplastic model ($n^{\mathrm{vp}} = 1.0$). Strain softening was applied for crustal materials only and was based on an explicit implementation (see the previous section). We have used an explicit marker advection scheme (4^{th} order Runger-Kutta in space) with a variable time step (Courant number: 0.25). Governing equations and material properties are given in Appendix C and Table C.1.

5.1 Evolution of reference models

5.1.1 The symmetric case

In order to test the robustness of our algorithm, we first consider a case where both initial and boundary conditions are symmetric. The viscoplastic viscosity was set to 2×10^{20} Pa·s ($n^{\mathrm{vp}} = 1.0$), which corresponds to a reference overstress of 2×10^5 Pa. The spatial resolution was set to 187 m (1600×592 cells). To enforce symmetry, the dip of the weak elliptical inclusion is set to 0. After 1 My of extension, stress is built up and frictional deformation occurs in the brittle regions of the whole crust (Fig. 7A). After 4.5 My, a symmetric neck has developed and the crust has thinned by a factor 3 in the center of the model. With exhumation, the mantle cools down which leads to embrittlement (Fig. 7B,C). At 7.6 My, the crust is hyperthinned and is about to break-up and locally reaches 1/8 of its original thickness (Fig. 7C). In compression, the yield stress is higher and more time is needed for fully buiding-up stresses. After 1 My, only the upper crust is affected by frictional plastic deformation (Fig. 7D). At 4.5 My, the lithosphere has started to buckle and frictional deformation occurs throughout the entire crust (Fig. 7E). After 7.6 My, the crust was locally thickened by almost a factor 2. It is striking that, either in compression or extension, all modelled structures are strictly symmetric despite the use of frictional strain softening.

5.1.2 The asymmetric case

Now we introduce asymmetry by initially tilting the weak elliptical inclusion. Again, models were run under either extension or compression for a total duration of 7.6 My. The viscoplasticty parameters and model resolution were similar to that of the previ-

ous example (187 m). After 1.9 My of extension, a graben has developed at the centre of the domain and is laterally limited by two conjugate crustal-scale shear bands. The 500°C isotherm is located at an average depth of 30 km. Asymmetry is inherited from the initial tilt of the elliptical inclusion. As a result, the right shoulder rises higher than the left one (Fig. 8A). After 4.4 My of stretching, the crust has necked down to a thickness of about 10 km and the 500°C isotherm now coincides with the base of the thinned crust. The isocontour of strain reveals accumulated strain in the necked crust and at the base of the lithosphere (Fig. 8B). After 7.6 My, the crust reaches the stage of break-up and two margins can be identified. An asymmetry can be observed as the right margin is almost twice as wide as the left margin. The strain isocontour indicates that both plates are weakly deformed above 90 km depth (Fig. 8C). After 1.9 My of compression, a crustal scale pop-up structure develops in the centre of the domain. As in extension, asymmetry is inherited from the orientation of the perturbation. Hence, the thrust that has the closest orientation to that of the weak elliptical inclusion becomes the main thrust. The right part of the domain thus becomes the lower plate (Fig. 8D). After 4.4 My, the Moho of the lower plate reaches 60 km depth. The left plate starts to bend downwards and the plateau is slightly tilted towards the hinterland. A secondary thrust develops in the foreland. The 500°C isocontour is characterised by an upward deflection below the plateau (Fig. 8E). At 7.6 My, new thrusts have developed in the foreland and the tilt of the plateau reaches 10°. The 500°C isocontour has started to diffuse laterally and the lower plate Moho reaches a depth of 80 km. The accumulated strain contours reveal that foreland deformation was mostly accommodated by the two deeply rooted thrusts (Fig. 8F).

5.2 Behaviour upon mesh refinement

We have next studied the behaviour of the models upon mesh refinement. In extension, the differences are minor when considering simulations for a low spatial resolution (200 × 72, Fig. 9A), a medium resolution (400 × 148, Fig. 9B), a high resolution (800×296, Fig. 9C), and the reference case (1600×592, Fig. 8C). Invariably, two plates that are weakly deformed above 90 km depth were individualised in response to stretching. The margins exhibit similar morphologies and the stress distributions are in broad agreement. The maximum stress values are located in the frictional plastic regions of the lower crust of both plates (250-300 MPa).

In compression, the sensitivity of models to the spatial resolution is bigger. In particular, the low resolution model (Fig. 9D) does not capture some important characteristics of the reference simulation (Fig. 8F). Instead, the upper plate is bent upwards, the plateau is not developed and the upper plate deformation is limited. These differences progressively diminish with increasing resolution. For example, the upper plate depicts a slight downward bending for the medium resolution (Fig. 9E). For the high resolution, the amplitude of bending reaches that of the reference model (Fig. 9F). The shape of the 500°C isotherm also converges with increasing resolution. A stress of 500 MPa (and above) is reached in the frictional portions on the middle crust, lower crust and the mantle lithosphere. The morphology of the frontal part of the wedge is the most critical region with regard to mesh convergence. Strain isocontours indicate that plateau growth was mainly accommodated by two deeply rooted thrusts in both high and reference resolution models. However, there are still important differences in terms of the number of secondary thrusts, which locally affect the topography.

5.3 Effect of the reference overstress

In V-E-VP models the rate-dependence of the plastic model is controlled by the viscoplastic viscosity. The subsequent overstress is thus expected to influence the results of numerical simulations. For example, setting a too small value of reference overstress (or too small η^{vp}) is equivalent to running a model in the rate independent limit (V-E-P). In this limit, the benefits of viscoplastic regularisation vanish and models will thus likely fail at satisfying a global force equilibrium. For a too large value of overstress, the computed stress will be well above the value predicted with the rate-independent Drucker-Prager. This will impede shear banding and strain localisation in the frictional domain. We have examined the impact of this parameter on the lithospheric extension and on the compression models for the medium resolution (800×296 cells). For the reasons mentioned above, we have varied η^{vp} within a narrow range (10^{20} – 10^{21} Pa·s). In extension, the magnitude of the viscoplastic viscosity has little influence on the morphology of the margins and the timing of break-up (Fig. 10,A,B,C). Small values of η^{vp} promote strain localisation in the frictional plastic domain, enhances asymmetry as well as topographic gradients. However, the overall stress distribution and the stress levels are not affected strongly. The value of η^{vp} also has an influence on lithosphere compression models, in particular on the distribution of strain around the plateau. For example, we observe that

deeply rooted foreland-dipping thrusts are better captured for $\eta^{vp} \leq 4\times10^{20}$ Pa·s (Fig. 10D,E,F). Interestingly, we notice that variations of η^{vp} affect the bending of the upper plate in a similar way as the numerical resolution (see section above). However, the overall stress distribution and the stress level, and also the topography, are influenced rather weakly by the value of η^{vp}.

6 Discussion

Geodynamic models generally rely on continuum mechanics and can, so far, at best achieve resolutions of 100 m scale [*Naliboff et al.*, 2017; *Petri et al.*, 2019]. Such resolutions are far larger than the actual width of fault zones that develop in the brittle part of the lithosphere [*Shipton et al.*, 2006]. The latter could also be modelled using discontinuous representations [e.g. *de la Puente et al.*, 2009]. However this approach does not seem suitable for large deformation tectonic models that should capture the self-consistent emergence and activation of frictional plastic shear bands. Moreover, such models will not be able to resolve fault zones dimensions, in particular their width, when a sufficiently high resolution will have come within reach. It is therefore essential to continue developing geodynamic models based on a continuum approach that can capture frictional plastic strain localisation while providing stable and accurate numerical solutions.

To this end we have used viscoplastic regularisation, which offers a simple and direct way to obtain convergence during the non-linear iterative process and vastly reduces mesh dependence. However, it also clear that viscoplastic regularisation is a time regularisation and not a spatial regularisation approach [*Wang*, 2019], in contrast with nonlocal plastic models [*Bažant and Lin*, 1988], gradient-based models [*de Borst and Mühlhaus*, 1992] or models based on Cosserat medium [*Mühlhaus and Vardoulakis*, 1987]. Viscoplasticity is thus not expected to resolve all mesh convergence problems. Future models should also consider spatial regularisation approach, which despite their algorithmic and computational cost become more and more affordable in the context of fully iterative solving strategies (e.g. pseudo-transient chemes in *Räss et al.* [2019]). In particular, it will be interesting to consider models that include both temporal and spatial regularisation in the context of geodynamic modelling [*Wang*, 2019]. In general care has to be taken when selecting the value of the viscoplastic viscosity which controls the overstress, since the evolution of lithospheric models can be affected by this parameter (see Supporting Information Figure S3). For the parameter range considered, these variations fortunately

appeared less important than variations caused by mesh refinement. In fact, even with regularisation and for converged flow and pressure fields, secondary shear bands still develop as the resolution increases. This is likely a limitation of viscoplastic regularisation.

With viscoplastic regularisation, the main parameter that allows for obtaining non-linear convergence while preserving shear localisation is the reference overstress. We recommend to select values in the order of, or below, 1 MPa. Local variations of the overstress may become important in cases where strong localisation occurs. The above presented power-law model appears as a suitable way to limit this drawback. We have also noticed that, in some cases, checkerboard-style shear banding patterns can arise (see Supporting Information Figure S6). This was observed for very low values of the overstress ($\bar{\sigma} < 10^5$ Pa), in the limit where viscoplastic regularisation becomes inefficient and where shear band can not be captured by the model resolution. This effect occurred in compressible models and is unrelated to the stability of the velocity-pressure discretisation. Further work is needed to investigate what controls this behaviour and to further improve the use of V-E-VP geodynamic models.

The effects of elastic bulk deformation can be included in geodynamic models [e.g *Poliakov et al.*, 1993; *Gerbault et al.*, 1998; *Popov and Sobolev*, 2008; *Choi and Petersen*, 2015]. Herein we demonstrate that variations in the bulk modulus (K) or in Poisson ratio (ν) within a realistic range can have a strong impact on the patterns and the intensity of shear banding. More intense strain localisation was obtained for high values of the elastic bulk modulus, in particular near the limit of elastic incompressibility. This can be explained by the fact that an increase of bulk modulus decreases effective elasto-plastic hardening, which can promote strain localisation. For a larger bulk modulus, the strength contrast between shear bands and their less deformed host is increased, which further contribution to localisation. A comparison of models involving or not the effect of elastic compressibility was presented in *Choi and Petersen* [2015]. In comparison to geoFLAC models ($\nu = 0.25$), those based on the code 2DPIC (elastically incompressible, $\nu = 0.5$) displayed larger accumulated strain inside shear bands at high resolution (their Fig. 2). This would also indicate that elastic incompressibility further contribute to strain localisation. This is however to be taken with care since the models presented in this study are based on fundamentally different algorithms. Models with a positive dilatancy angle exhibit a more diffuse localisation pattern than plastic incompressible models in agreement with the results of *Choi and Petersen* [2015]. The use of dilation

strain softening as suggested by *Choi and Petersen* [2015], for example, seems a reasonable choice to capture the initial volume changes due to frictional plastic deformation, while still promoting the localisation of strain. It is also in agreement with laboratory measurements [*Zhao and Cai*, 2010, e.g.].

We have also demonstrated that for frictional plasticity a velocity-based Effective Viscosity Approach (EVA) can deliver the same results as algorithms more common in the engineering literature, which are based on a displacement formulation and which linearise a return-mapping scheme to compute the stresses in order to derive a consistently linearised tangent operator (see also *Lemiale et al.* [2008]). This is of course subject to the use of the same spatio-temporal discretisation. Moreover, it is also conditional on a proper derivation of the visco-elastic-viscoplastic viscosity η^{vep} and a proper differentiation of this quantity. This includes that (i) elastic deformations are taken into account, (ii) plastic incompressibility is not assumed a priori, and (iii) a proper loading-unloading criterion is utilised. It is noted that, at variance with these conditions, the original formulation [*Willett*, 1992] assumes the absence of elastic deformations, plastic incompressibility, and continued plastic loading. For the original formulation, identical results are therefore obtained only for a small subset of constitutive models.

The cause for the generation of non-symmetric structures in geodynamic models is typically attributed to plastic strain softening [e.g. *Huismans and Beaumont*, 2003]. In this study, strictly symmetric deformation structures arise from models with symmetric initial and boundary conditions, despite the use plastic strain softening (Fig. 7). In fact, it was necessary to introduce asymmetry in either the initial or boundary conditions to model the formation of asymmetric rifts and mountain belts (Fig. 8). Here, asymmetry is thus induced by either inheritance or far-field kinematics. This is provided that all discretisation elements (mesh, markers, stencils) are symmetric and that frictional plasticity is solved to reasonable accuracy (here, machine precision) at each time increment. Given these conditions, we argue that monitoring the symmetry of structures can demonstrate the robustness of geodynamic models that encapsulate non-linear rheologies upon large strain. A comparison with the models presented in *Huismans and Beaumont* [2007] is provided in the Supporting Information. It shows that both symmetric and asymmetric deformation can be modelled using a either a linear or power-law V-E-VP model (Supporting Material Figure S1 to S3) and further confirms that the occurrence of asymmetric deformation is controlled by the amplification of initial perturba-

tions (Supporting Material Figure S1). Our result hence confirm those of *Huismans and Beaumont* [2007] but also clarifies the role of strain softening which acts as a catalyser rather than a source of asymmetry.

Numerical geodynamic models often exhibit lack of global non-linear convergence [*Spiegelman et al.*, 2016] and the consequences of this is debatable. Here we have used viscoplastic regularisation to obtain global convergence in geodynamic models. We have observed that when the internal dynamics of the model is not entirely controlled by frictional plastic strain localisation, the evolution of models is not very sensitive to the convergence of global equilibrium (see Supporting information Figure S9). This is a positive outcome which is particularly true when the plastic layer (i.e. the crust) is considered as a passive stress limiter or when the internal dynamics is controlled by a mesh-convergent strain localisation phenomenon (e.g., shear heating in [*Schmalholz et al.*, 2014]). However, the propagation of frictional plastic shear bands is affected by the convergence of global equilibrium (see Supporting information Figure S9). This can become problematic in cases where frictional plastic shear banding is the essential ingredient of a model. It is therefore important to keep on developing aspects related to frictional plasticity and to continue improving the reliability and robustness of future geodynamic models.

7 Conclusions

We have designed and tested new aspects of a frictional viscoplastic rheology for geodynamic modelling, with emphasis on viscoplastic regularisation. Elastic compressibility, power-law viscoplasticity, plastic dilatancy, non-associated plastic flow and strain softening can be all combined, and incorporated in a numerical approach that is common in geodynamic modelling, i.e. a velocity-pressure formulation with an Effective Viscosity Approach. Moreover, since we accounted for elastic deformation and compressibility, the approach can deliver results which are as accurate as those in computational engineering. We have shown that elastic volumetric deformations, usually neglected in geodynamic modelling, can have a noticeable impact on patterns and intensity of shear banding. Power-law viscoplasticity can be used to limit the overstress inherent in viscoplastic modelling. Finally, we have shown successful applications of the visco-elastic-viscoplastic (V-E-VP) model in the context of state-of-the geodynamic simulations. Hence, viscoplastic regularisation may be used for practical purposes when modelling long-term tectonic deformations.

8 Figures

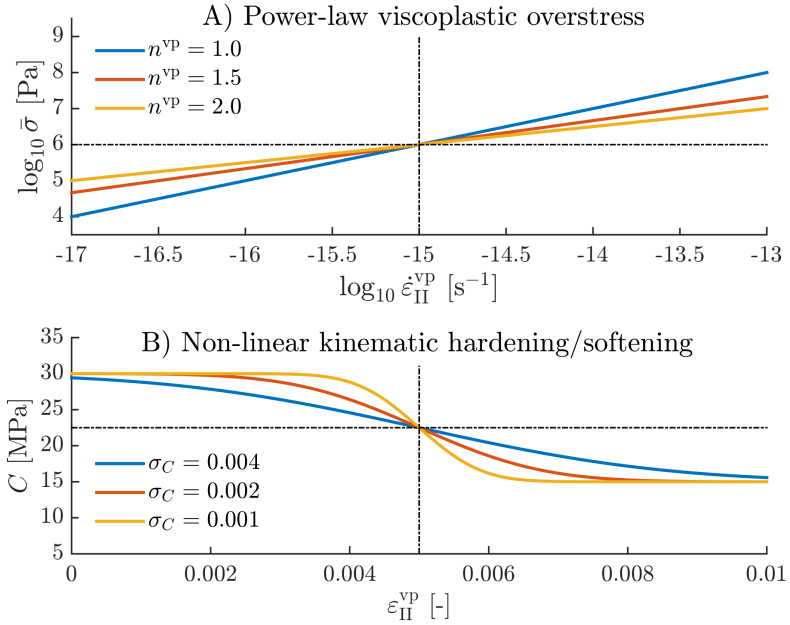

Figure 1. A) Effect of power-law viscoplasticity exponent n^{vp} on the magnitude of viscoplastic overstress $\hat{\sigma}$ for the variable viscoplastic strain rate $\dot{\varepsilon}_{II}^{vp}$. B) Example of a non-linear hardening/softening law. Effect of variable standard deviation σ_c on the evolution of cohesion c with the viscoplastic strain ε_{II}^{vp}.

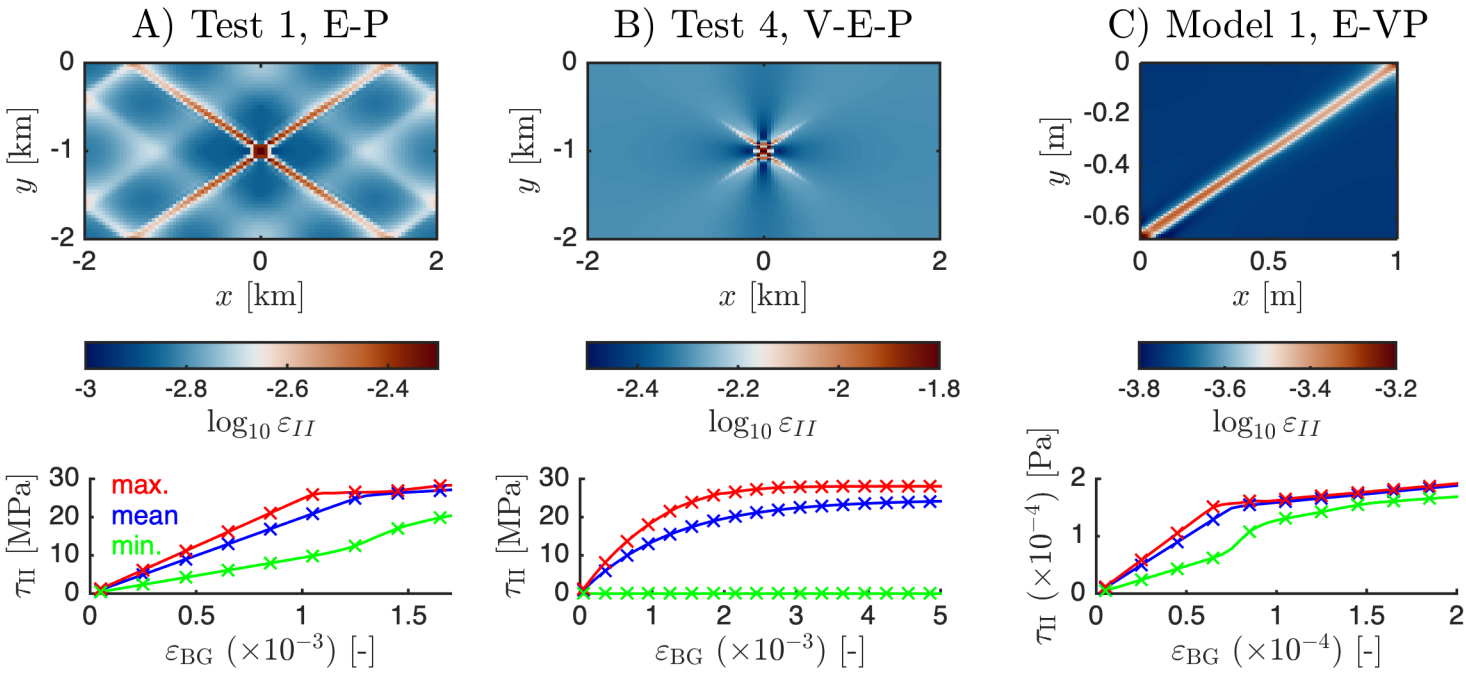

Figure 2. Simulation of shear banding using compressible elastic and plastic rheology. A) Test 1 from *Duretz et al.* [2018]: shear banding in the elasto-plastic regime. B) Test 4 from *Duretz et al.* [2018]: shear banding in the visco-elasto-plastic regime. C) Reference model from *Duretz et al.* [2019]: shear banding in the elasto-viscoplastic regime. Lower panel plots show comparison between measurements of the second deviatoric stress invariant (minimum, mean, maximum). Solid lines are from published studies [*Duretz et al.*, 2018, 2019] and are based on incremental displacement-based consistent tangent formulation. The crosses correspond to this study and are based on velocity-pressure effective viscosity formulation. The colour maps are *vik* from *Crameri* [2018].

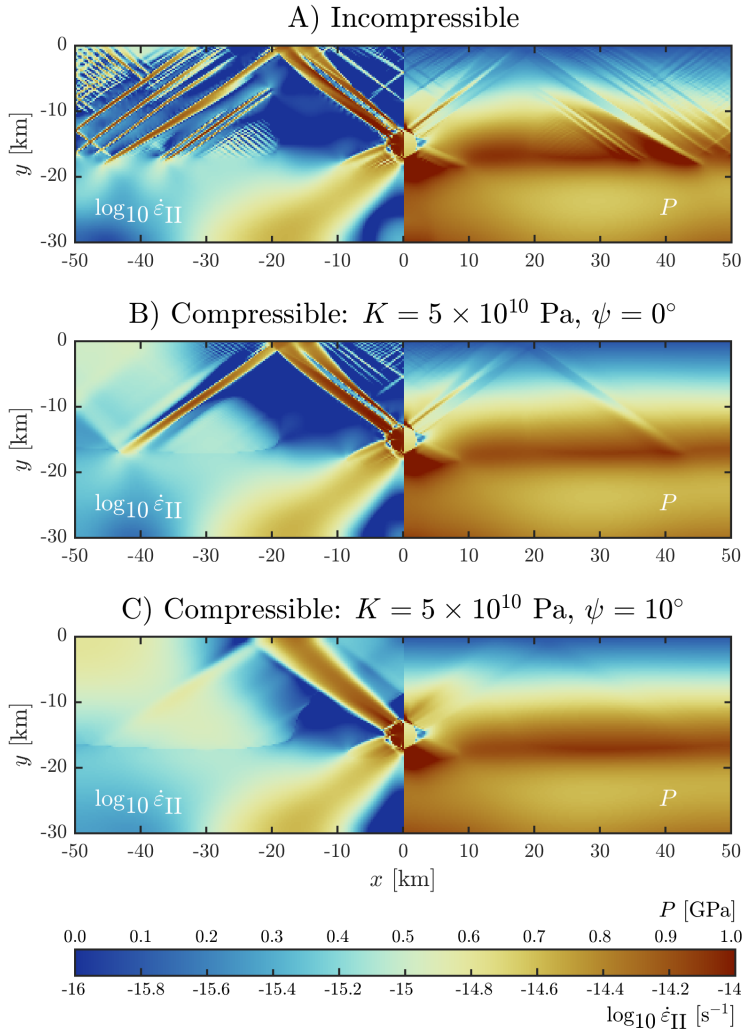

Figure 3. Role of compressibility on crustal shear banding. A) Strictly incompressible model. B) Model incorporating elastic volumetric deformations. C) Model incorporating both elastic volumetric deformation and plastic dilatancy. For each model, the second invariant of deviatoric strain rate tensor (logarithmic scale, left) and the pressure (right) are depicted. The colour map is *roma* (inverted) from *Crameri* [2018]. Initial and boundary conditions are similar to those of *Duretz et al.* [2020] excepted that we consider the top boundary to be a free surface (see text for details)

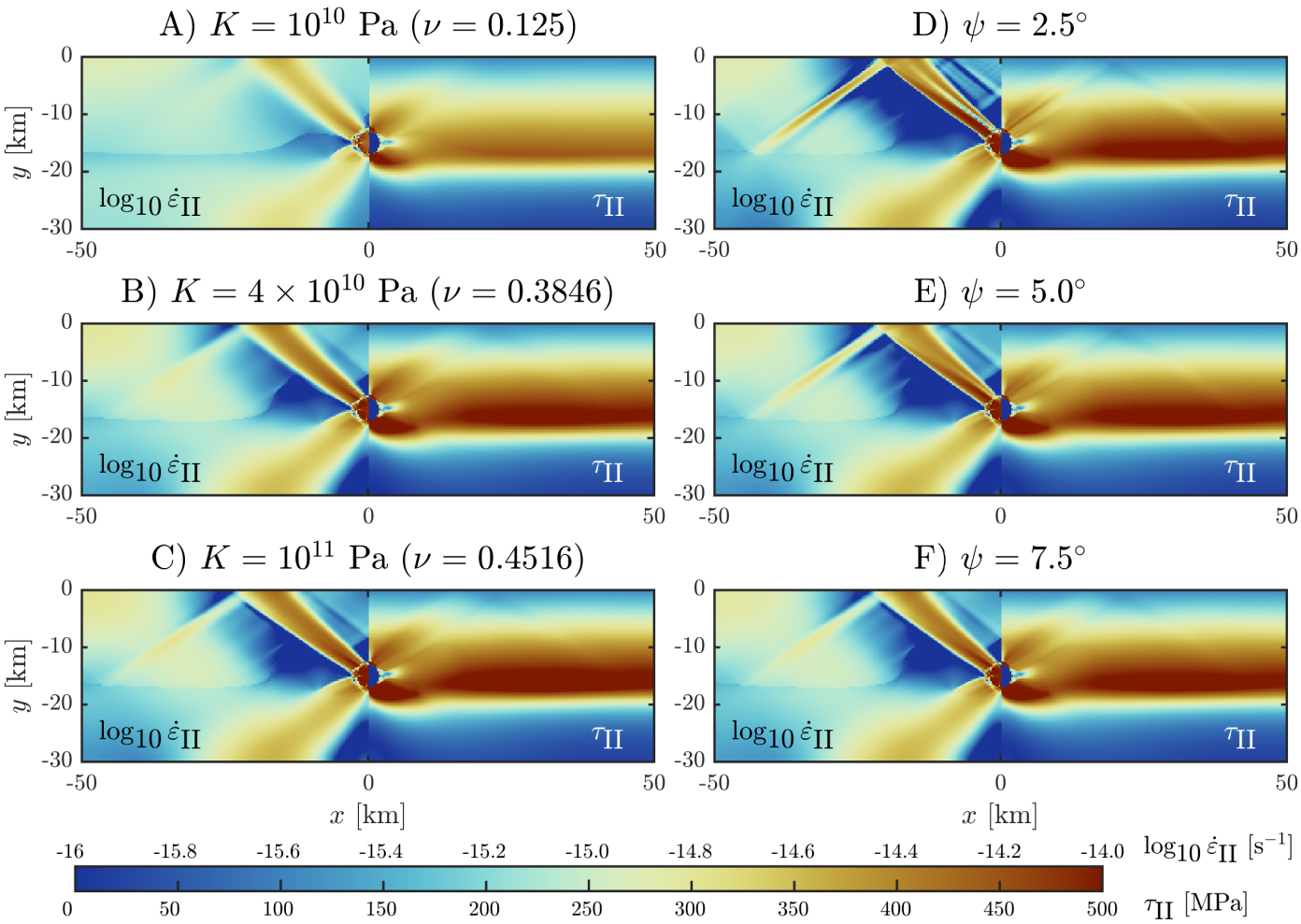

Figure 4. Effect of varying elastic compressibility and plastic dilation. A), B) and C) shows models with different elastic compressibilities. E), F) and G) correspond to runs with different dilatancy angles (constant). The colour map is *roma* (inverted) from *Crameri* [2018].

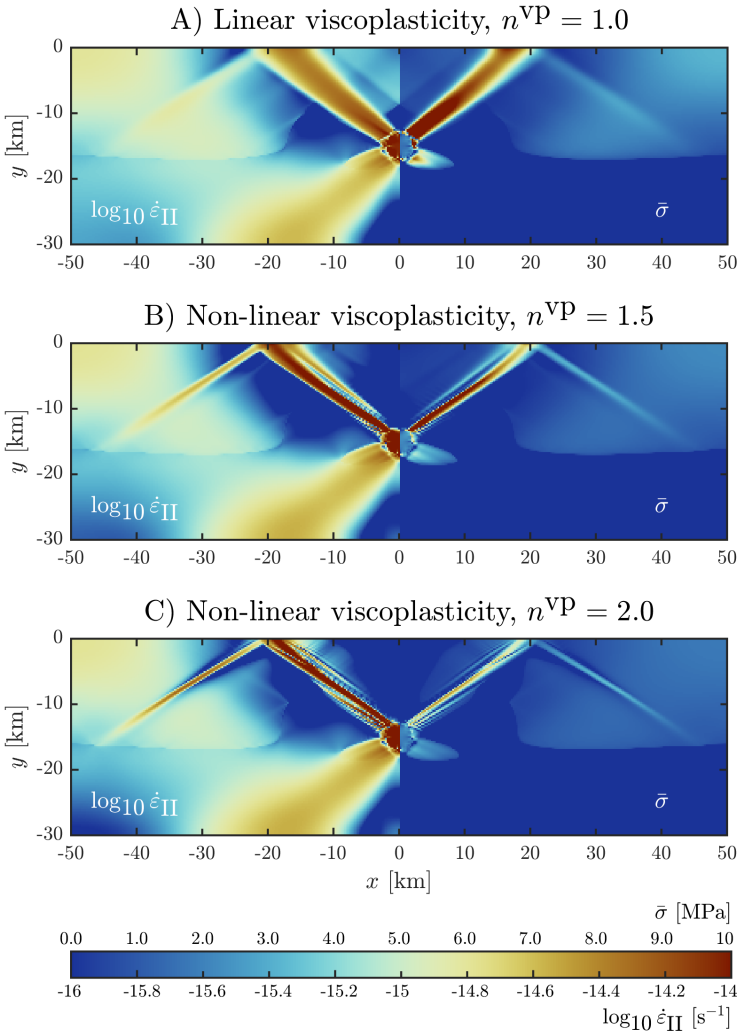

Figure 5. Effect of power-law viscoplasticity. A) Compressible model ($K = 5 \times 10^{10}$ Pa, $\psi = 10°$) with linear viscoplasticity. B) Similar model as in A) but with power-law viscoplastic exponent set to 1.5. C) Similar model as in A) but with power-law viscoplastic exponent set to 2.0. For each model, both the second invariant of deviatoric strain rate tensor (logarithmic scale, left) and the viscoplastic overstress (right) are depicted. The colour map is *roma* (inverted) from *Crameri* [2018].

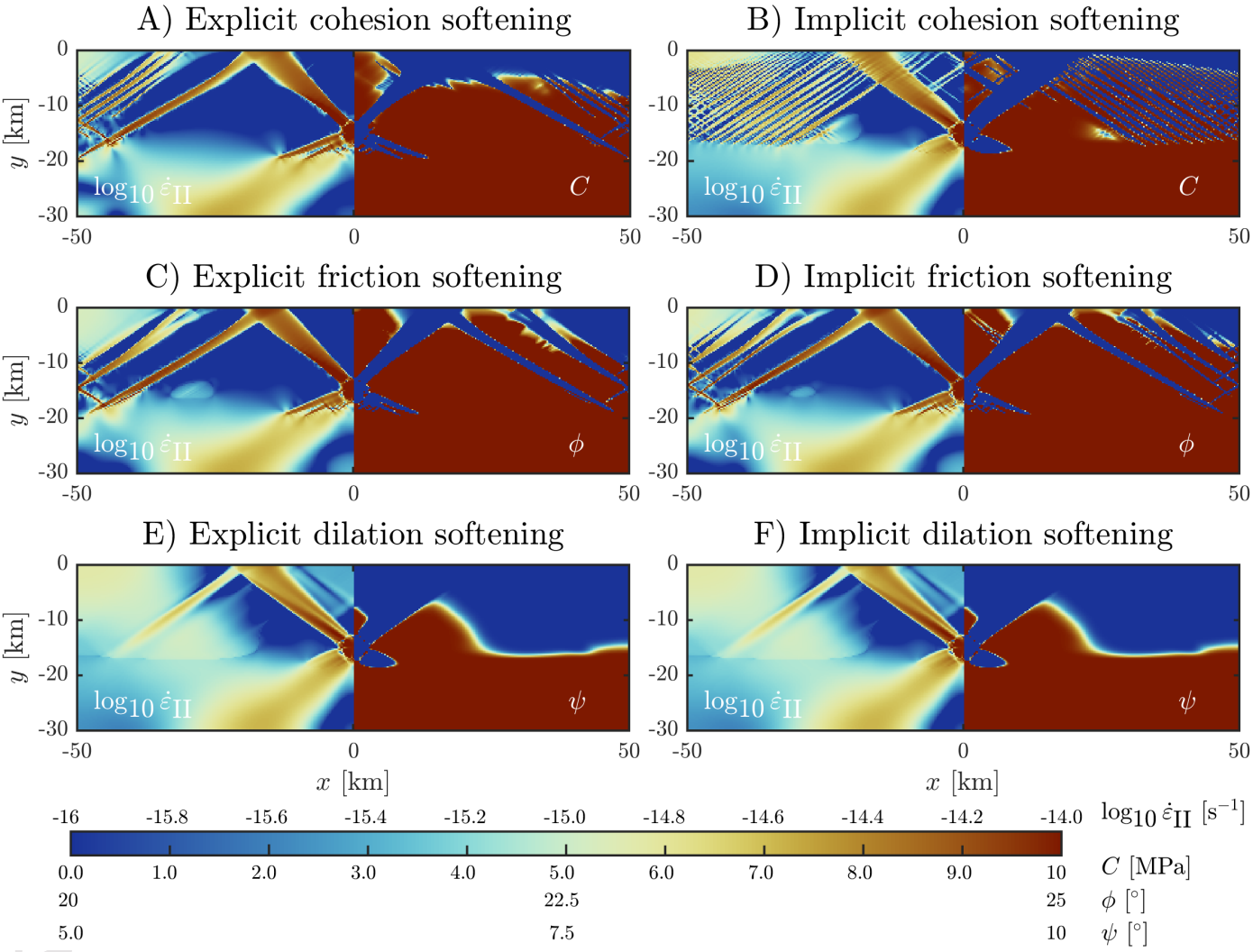

Figure 6. Effect of strain softening on crustal shear banding. The left column shows results obtained with an explicit strain softening implementation. The right column correspond to an implicit implementation of strain softening. A) and B) correspond to cohesion softening from 50 MPa to 25 MPa. C) and D) correspond to a friction softening from 25 to 20°. E) and F) correspond to a dilation softening from 10 to 5°. The colour map is *roma* (inverted) from *Crameri* [2018].

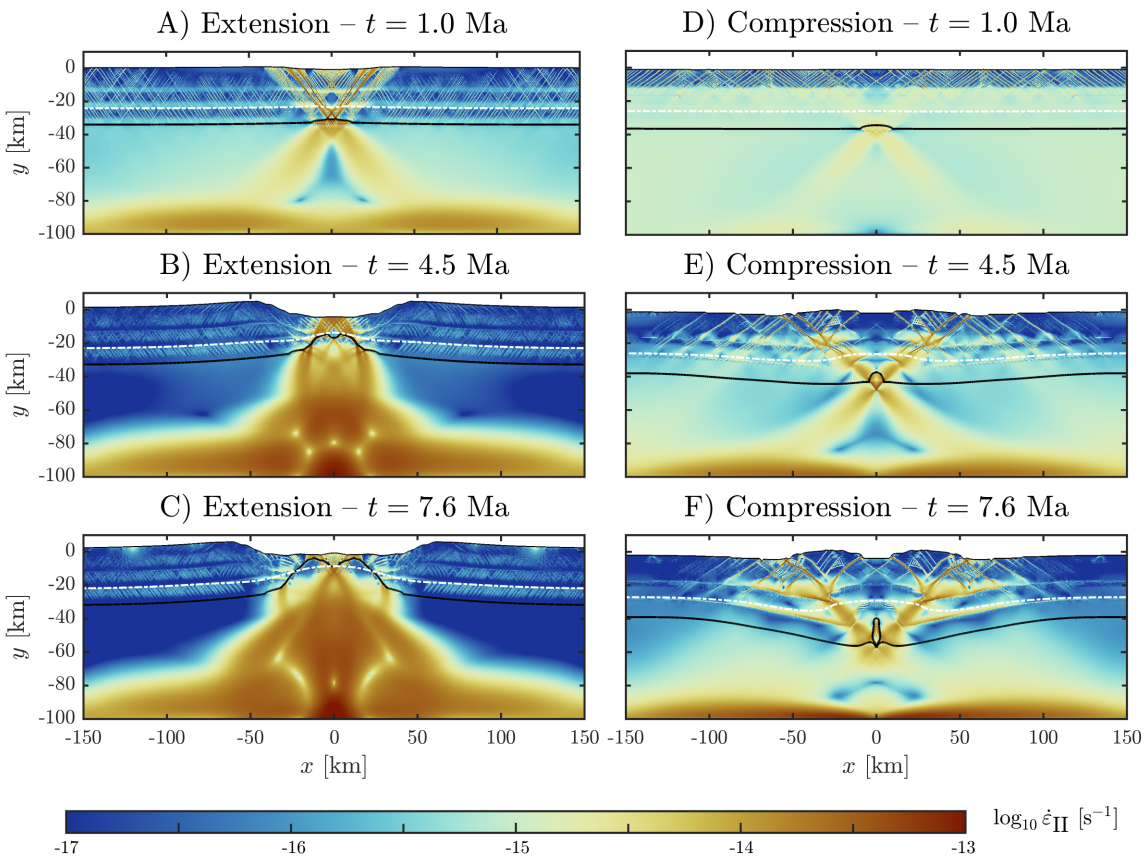

Figure 7. Time evolution of the lithospheric models in extension and compression with symmetric initial and boundary conditions. The colour map corresponds to the magnitude of the second deviatoric strain-rate invariant (logarithmic scale). Panels A, B, and C depict a model undergoing extension, panels D, E, F correspond to compression. The black lines indicate the location of the Moho, the dash-dotted line is the 500° isotherm.

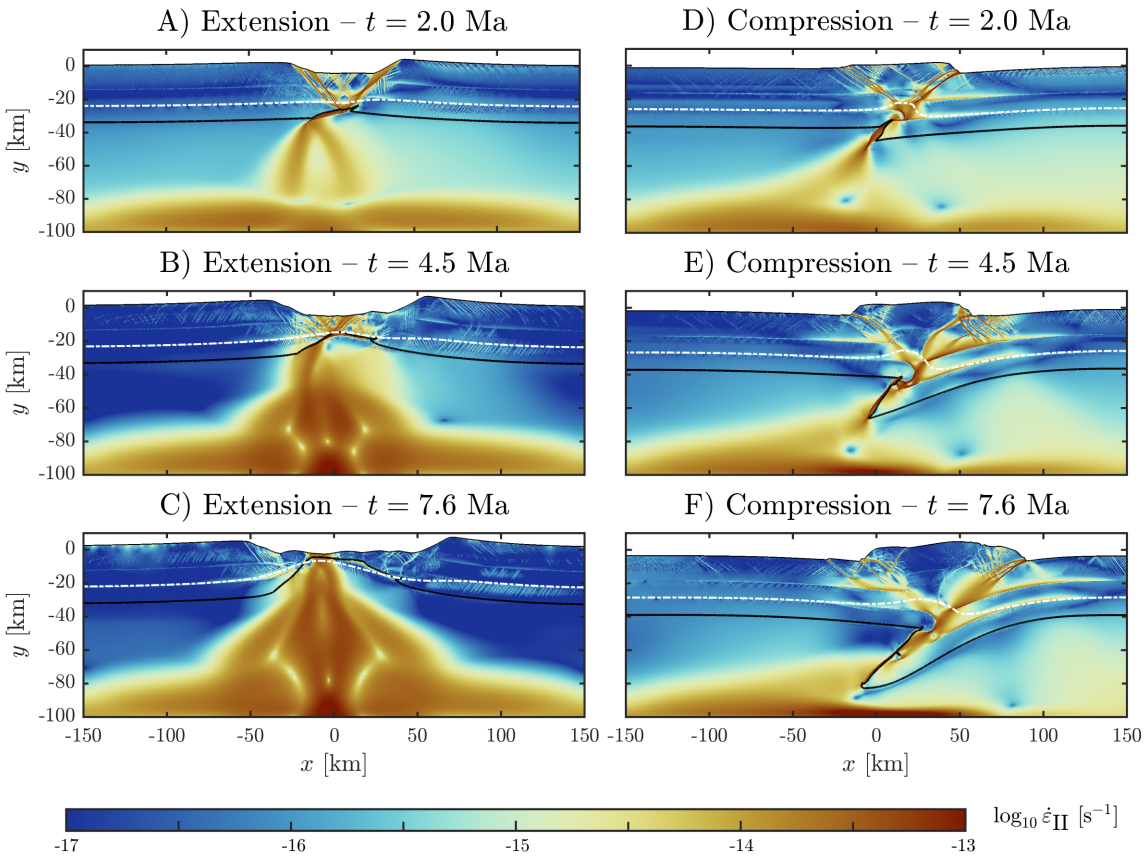

Figure 8. Time evolution of the reference lithospheric models in extension and compression with a non-symmetric initial condition. The colour map corresponds to the magnitude of the second deviatoric strain-rate invariant (logarithmic scale). Panels A, B, and C depict a model undergoing extension, panels D, E, F correspond to compression. The black lines indicate the location of the Moho, and the dash-dotted line is the 500° isotherm.

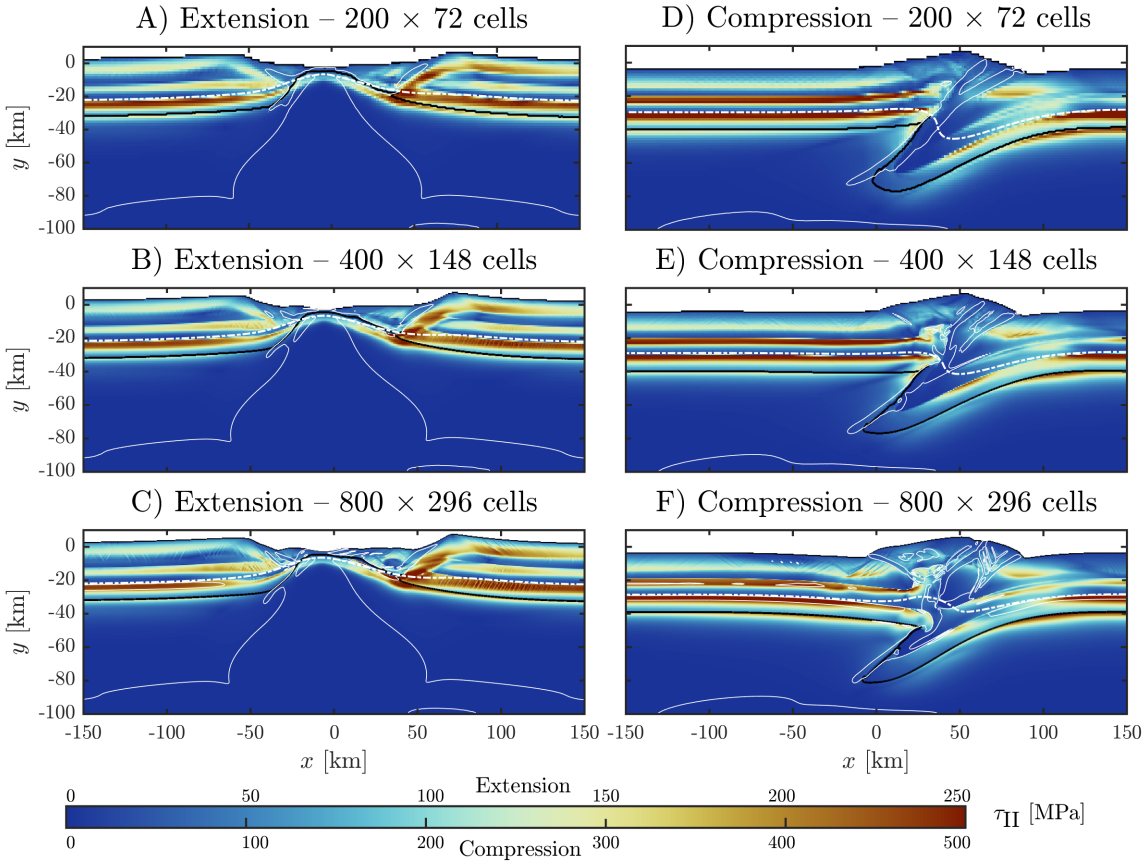

Figure 9. Sensitivity of lithospheric models to numerical resolution. Extension to the left, compression to the right. Three different resolutions are depicted (200 × 72, 400 × 148 and 800 × 296 cells). The model time is about 7.6 My for all six simulations. The colormap corresponds the second deviatoric stress invariant (τ_{II}). The black lines indicate the location of the Moho, the dash-dotted line is the 500° isotherm and the thin solid white is the 1.8 accumulated strain contour.

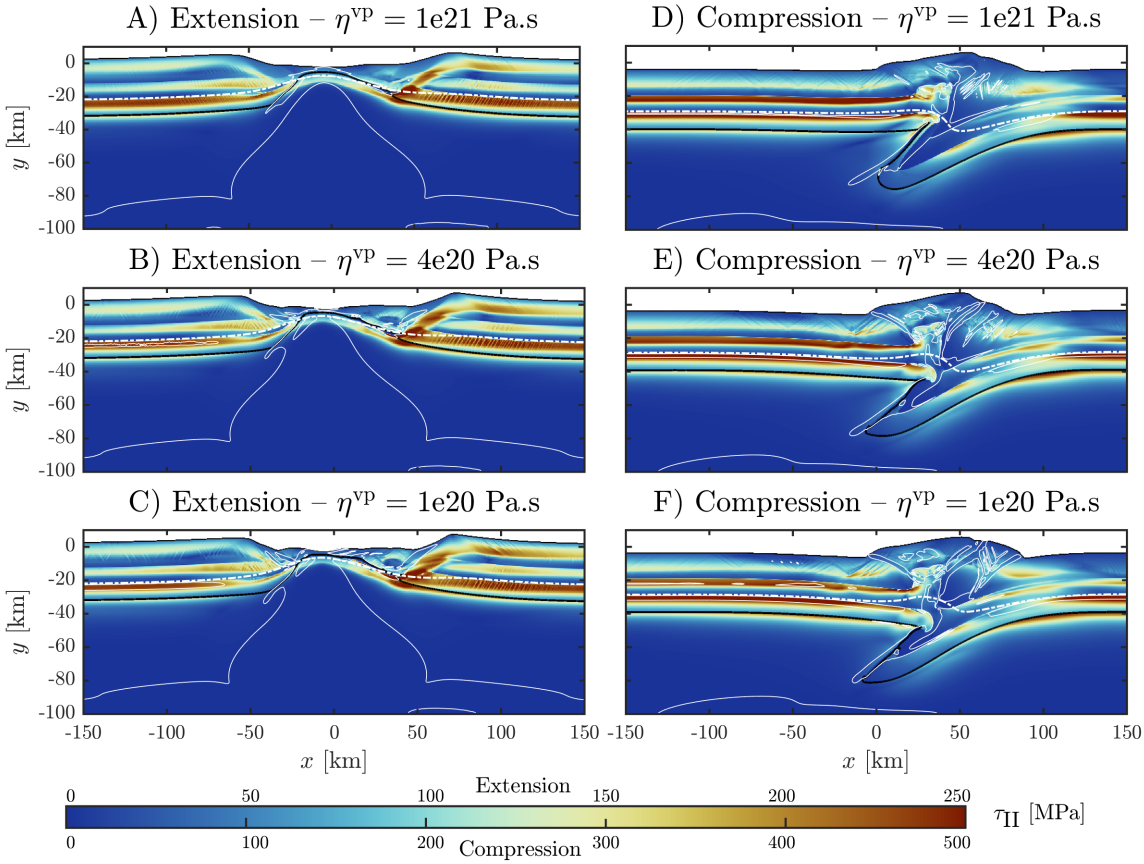

Figure 10. Influence of the viscoplastic viscosity on the evolution of lithospheric models in both extension (left panels) and compression (right panels). Three values of η^{vp} were investigated (10^{21}, 4×10^{20} and 10^{20} Pa· s). Models were run up to about 7.6 My. The colormap corresponds the second deviatoric stress invariant (τ_{II}). The black lines indicate the location of the Moho, the dash-dotted line is the 500° isotherm and the thin solid white is the 1.8 accumulated strain contour.

Table 1. Model parameters used for Fig. 2. The friction and dilatancy angles were set to 30 and 10 degrees, respectively.

Parameter	*Test 1*	*Test 4*	*Model 1*
$L_x[m]$	4000	4000	1
$L_y[m]$	2000	2000	0.685
$n_x \times n_y$	100×50	100×50	100×69
K [Pa]	2×10^{10}	2×10^{10}	2.0
G_{mat} [Pa]	10^{10}	10^{10}	1.0
G_{inc} [Pa]	2.5×10^{9}	10^{10}	0.25
η_{mat} [Pa.s]	-	2.5×10^{21}	-
η_{inc} [Pa.s]	-	10^{17}	-
C [Pa]	3×10^{7}	3×10^{7}	1.75×10^{-4}
η^{vp} [Pa.s]	-	-	2.5×10^{2}
$\dot{\varepsilon}_{\mathrm{BG}}$ [s^{-1}]	10^{-15}	10^{-15}	5×10^{-10}
Δt [s]	10^{10}	10^{10}	10^{4}

A: The tangent operator: Newton linearisation of the effective viscosity approach (EVA Newton)

The tangent operator needed for the global Newton-Raphson iterations is expressed as: $\mathbf{D} \equiv \frac{\delta \boldsymbol{\tau}}{\delta \dot{\boldsymbol{\varepsilon}}}$, where $\boldsymbol{\tau} = \left[\tau_{xx}\ \tau_{yy}\ \tau_{zz}\ \tau_{xy}\ p^{\mathrm{corr}}\right]^T$, $\dot{\boldsymbol{\varepsilon}} = \left[\epsilon_{xx}\ \epsilon_{yy}\ \epsilon_{zz}\ \epsilon_{xy}\ v'_{k,k}\right]^T$ and may be explicitly written as:

$$\mathbf{D}^{\mathrm{eva}} = \begin{bmatrix} \frac{\partial \tau_{xx}}{\partial \dot{\varepsilon}'_{xx}} & \frac{\partial \tau_{xx}}{\partial \dot{\varepsilon}'_{yy}} & \frac{\partial \tau_{xx}}{\partial \dot{\varepsilon}'_{zz}} & \frac{\partial \tau_{xx}}{\partial \dot{\varepsilon}'_{xy}} & \frac{\partial \tau_{xx}}{\partial v'_{k,k}} \\ \frac{\partial \tau_{yy}}{\partial \dot{\varepsilon}'_{xx}} & \frac{\partial \tau_{yy}}{\partial \dot{\varepsilon}'_{yy}} & \frac{\partial \tau_{yy}}{\partial \dot{\varepsilon}'_{zz}} & \frac{\partial \tau_{yy}}{\partial \dot{\varepsilon}'_{xy}} & \frac{\partial \tau_{yy}}{\partial v'_{k,k}} \\ \frac{\partial \tau_{zz}}{\partial \dot{\varepsilon}'_{xx}} & \frac{\partial \tau_{zz}}{\partial \dot{\varepsilon}'_{yy}} & \frac{\partial \tau_{zz}}{\partial \dot{\varepsilon}'_{zz}} & \frac{\partial \tau_{zz}}{\partial \dot{\varepsilon}'_{xy}} & \frac{\partial \tau_{zz}}{\partial v'_{k,k}} \\ \frac{\partial \tau_{xy}}{\partial \dot{\varepsilon}'_{xx}} & \frac{\partial \tau_{xy}}{\partial \dot{\varepsilon}'_{yy}} & \frac{\partial \tau_{xy}}{\partial \dot{\varepsilon}'_{zz}} & \frac{\partial \tau_{xy}}{\partial \dot{\varepsilon}'_{xy}} & \frac{\partial \tau_{xy}}{\partial v'_{k,k}} \\ \frac{\partial p^{\mathrm{corr}}}{\partial \dot{\varepsilon}'_{xx}} & \frac{\partial p^{\mathrm{corr}}}{\partial \dot{\varepsilon}'_{yy}} & \frac{\partial p^{\mathrm{corr}}}{\partial \dot{\varepsilon}'_{zz}} & \frac{\partial p^{\mathrm{corr}}}{\partial \dot{\varepsilon}'_{xy}} & \frac{\partial p^{\mathrm{corr}}}{\partial v'_{k,k}} \end{bmatrix} \tag{A.1}$$

To construct the tangent operator matrix needed, we first reformulate the rheology as:

$$\begin{aligned} \tau_{ij} &= 2\eta \left(\dot{\varepsilon}'_{ij}, p^{\mathrm{trial}}\right) \dot{\varepsilon}'_{ij} \\ p^{\mathrm{corr}} &= p^{\mathrm{trial}} + K \Delta t v^{\mathrm{vp}}_{k,k} \left(\dot{\varepsilon}'_{ij}, p^{\mathrm{trial}}\right) \end{aligned} \tag{A.2}$$

where $\eta = \eta^{\mathrm{ve}}$ below the plastic yield and $\eta = \eta^{\mathrm{vep}}$ at the yield. Then, the components of $\mathbf{D}^{\mathrm{eva}}$ can be computed as follows:

$$\begin{aligned}
\frac{\partial \tau_{ij}}{\partial \dot{\varepsilon}'_{kl}} &= 2\eta \left(\delta_{ik}\delta_{jl} + \dot{\varepsilon}'_{ij} \frac{\partial \eta}{\partial \dot{\varepsilon}'_{kl}} \right) \\
\frac{\partial \tau_{ij}}{\partial v_{k,k}} &= 2\dot{\varepsilon}'_{ij} \frac{\partial \eta}{\partial v_{k,k}} \\
\frac{\partial p^{\mathrm{corr}}}{\partial \dot{\varepsilon}'_{kl}} &= K\Delta t \frac{\partial v^{\mathrm{vp}}_{i,i}}{\partial \dot{\varepsilon}'_{kl}} \\
\frac{\partial p^{\mathrm{corr}}}{\partial v_{k,k}} &= -K\Delta t \left(1 - \frac{\partial v^{\mathrm{vp}}_{i,i}}{\partial v_{k,k}} \right).
\end{aligned} \tag{A.3}$$

In case of visco-elastic creep or incompressible plastic flow, the partial derivatives of p^{corr} vanish. Since we consider p^{trial} as our global variable for the pressure, it is necessary to express partial derivatives with regard to the trial pressure. The latter can be further computed using the chain rule as, $\frac{\partial \tau_{ij}}{\partial p^{\mathrm{trial}}} = \frac{\partial \tau_{ij}}{\partial v_{k,k}} \frac{\partial v_{k,k}}{\partial p^{\mathrm{trial}}}$ and $\frac{\partial p^{\mathrm{corr}}}{\partial p^{\mathrm{trial}}} = \frac{\partial p^{\mathrm{corr}}}{\partial v_{k,k}} \frac{\partial v_{k,k}}{\partial p^{\mathrm{trial}}}$, where $\frac{\partial v_{k,k}}{\partial p^{\mathrm{trial}}} = -\frac{1}{K\Delta t}$.

For completeness we also provide the partial derivatives of η^{ve} and η^{vep}. Upon completion of the local iterations that determine the visco-elastic trial stress, the effective viscosity is expressed as:

$$\eta^{\mathrm{ve}} = \left(\frac{1}{\eta^{\mathrm{v}}} + \frac{1}{\eta^{\mathrm{e}}} \right)^{-1} \text{with} \quad \eta^{\mathrm{v}} = C^{\mathrm{v}} \, {\tau_{II}}^{n^{\mathrm{v}}-1}. \tag{A.4}$$

The partial derivatives of η^{ve} take the following form:

$$\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{ij}} = \frac{\partial \eta^{\mathrm{ve}}}{\partial \tau_{ij}} \frac{2\eta^{\mathrm{ve}}}{1 - 2\left(\frac{\partial \eta^{\mathrm{ve}}}{\partial \tau_{xx}} \dot{\varepsilon}'_{xx} + \frac{\partial \eta^{\mathrm{ve}}}{\partial \tau_{yy}} \dot{\varepsilon}'_{yy} + \frac{\partial \eta^{\mathrm{ve}}}{\partial \tau_{zz}} \dot{\varepsilon}'_{zz} + \frac{\partial \eta^{\mathrm{ve}}}{\partial \tau_{xy}} \dot{\varepsilon}'_{xy} \right)} \tag{A.5}$$

Differentiation with regard to τ_{ij} yields:

$$\frac{\partial \eta^{\mathrm{ve}}}{\partial \tau_{ij}} = -b(n^{\mathrm{v}} - 1) \left(\eta^{\mathrm{ve}}\right)^2 C^{\mathrm{v}} \tau_{\mathrm{II}}^{n^{\mathrm{v}}-3} \tau_{ij}, \tag{A.6}$$

with $b = 1$ for normal components and $b = 2$ for the shear components. Note that the derivative of η^{ve} with regard to p^{trial} is 0 since we consider that visco-elastic creep in pressure-independent in this study.

In case of plastic flow, the partial derivatives of η^{vep} and $v^{\mathrm{vp}}_{i,i}$ need to be determined. The latter may be evaluated upon completion of the local iteration that determines the values of $\dot{\lambda}$ may be formulated as:

$$\begin{aligned}
\frac{\partial \eta^{\mathrm{vep}}}{\partial \dot{\varepsilon}'_{ij}} &= -\frac{\dot{\varepsilon}'_{ij} \tau_{\mathrm{II}}}{4\left(\dot{\varepsilon}'_{\mathrm{II}}\right)^3} + \frac{1}{2\dot{\varepsilon}'_{\mathrm{II}}} \left(a \frac{\partial \dot{\lambda}}{\partial \dot{\varepsilon}'_{ij}} + \dot{\lambda} \frac{\partial \eta^{\mathrm{vp}}}{\partial \dot{\varepsilon}'_{ij}} \right) \\
\frac{\partial \eta^{\mathrm{vep}}}{\partial p^{\mathrm{trial}}} &= \frac{1}{2\dot{\varepsilon}'_{\mathrm{II}}} \left(\sin\phi + a \frac{\partial \dot{\lambda}}{\partial p^{\mathrm{trial}}} + \dot{\lambda} \frac{\partial \eta^{\mathrm{vp}}}{\partial p^{\mathrm{trial}}} \right)
\end{aligned} \tag{A.7}$$

where $a = H + \eta^{\mathrm{vp}} + K\Delta t \sin\psi \sin\phi$. The partial derivatives of η^{vp} are non-zero only when $n^{\mathrm{vp}} > 1$. They are then expressed as:

$$\begin{aligned}\frac{\partial \eta^{\mathrm{vp}}}{\partial \dot{\varepsilon}'_{ij}} &= \left(\frac{1}{n^{\mathrm{vp}}} - 1\right) \frac{\eta^{\mathrm{vp}}}{\dot{\lambda}} \frac{\partial \dot{\lambda}}{\partial \dot{\varepsilon}'_{ij}} \\ \frac{\partial \eta^{\mathrm{vp}}}{\partial p^{\mathrm{trial}}} &= \left(\frac{1}{n^{\mathrm{vp}}} - 1\right) \frac{\eta^{\mathrm{vp}}}{\dot{\lambda}} \frac{\partial \dot{\lambda}}{\partial p^{\mathrm{trial}}}.\end{aligned} \tag{A.8}$$

and the derivatives of the plastic multiplier rate read:

$$\begin{aligned}\frac{\partial \dot{\lambda}}{\partial \dot{\varepsilon}'_{ij}} &= -\frac{\partial F}{\partial \dot{\lambda}}^{-1} \left(\frac{\partial F}{\partial \dot{\varepsilon}'_{ij}} - \dot{\lambda} \frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{ij}}\right) \\ \frac{\partial \dot{\lambda}}{\partial p^{\mathrm{trial}}} &= -\frac{\partial F}{\partial \dot{\lambda}}^{-1} \frac{\partial F}{\partial p^{\mathrm{trial}}}.\end{aligned} \tag{A.9}$$

Finally, the partial derivatives of the yield function can be expressed as:

$$\begin{aligned}\frac{\partial F}{\partial \dot{\varepsilon}'_{ij}} &= b\eta^{\mathrm{ve}} \frac{\dot{\varepsilon}'_{ij}}{\dot{\varepsilon}'_{\mathrm{II}}} + 2\dot{\varepsilon}'_{\mathrm{II}} \frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{ij}} \\ \frac{\partial F}{\partial p^{\mathrm{trial}}} &= -\sin\phi.\end{aligned} \tag{A.10}$$

The derivatives of plastic strain rate are expressed as:

$$\begin{aligned}\frac{\partial v^{\mathrm{vp}}_{i,i}}{\partial \dot{\varepsilon}'_{ij}} &= -\left(\frac{\partial Q}{\partial P} \frac{\partial \dot{\lambda}}{\partial \dot{\varepsilon}'_{ij}} + \dot{\lambda} \frac{\partial^2 Q}{\partial \dot{\varepsilon}'_{ij} \partial P}\right) \\ \frac{\partial v^{\mathrm{vp}}_{i,i}}{\partial p^{\mathrm{trial}}} &= -\left(\frac{\partial Q}{\partial P} \frac{\partial \dot{\lambda}}{\partial p^{\mathrm{trial}}} + \dot{\lambda} \frac{\partial^2 Q}{\partial\, {p^{\mathrm{trial}}}^2}\right)\end{aligned} \tag{A.11}$$

where the second order partial derivatives of Q are:

$$\begin{aligned}\frac{\partial^2 Q}{\partial \dot{\varepsilon}'_{ij} \partial p^{\mathrm{trial}}} &= \cos(\psi) \frac{\partial \psi}{\partial \dot{\lambda}} \frac{\partial \dot{\lambda}}{\partial \dot{\varepsilon}'_{ij}} \\ \frac{\partial^2 Q}{\partial^2 p^{\mathrm{trial}}} &= \cos(\psi) \frac{\partial \psi}{\partial \dot{\lambda}} \frac{\partial \dot{\lambda}}{\partial p^{\mathrm{trial}}}.\end{aligned} \tag{A.12}$$

In the case of Picard iterations, partial derivatives of effective viscosity and corrected pressure with regard to $\dot{\boldsymbol{\varepsilon}}$ are neglected, the tangent operator then simply reads:

$$\mathbf{D}^{\mathrm{pic}} = \begin{bmatrix} 2\eta & 0 & 0 & 0 & 0 \\ 0 & 2\eta & 0 & 0 & 0 \\ 0 & 0 & 2\eta & 0 & 0 \\ 0 & 0 & 0 & 2\eta & 0 \\ 0 & 0 & 0 & 0 & -K\Delta t \end{bmatrix}, \tag{A.13}$$

where $\eta = \eta^{\mathrm{ve}}$ below the plastic yield and $\eta = \eta^{\mathrm{vep}}$ at the yield.

B: The tangent operator: consistent tangent finite-step approach (Cons. Tangent)

The consistent tangent operator can also be derived by differentiating a finite-step visco-elasto-viscoplastic relation, which is a customary approach in computational en-

gineering. However, this is generally done for a displacement-based approach. Geodynamic codes are usually based on a velocity-pressure formulation, and it is therefore useful to express an appropriate form of the consistent tangent operator.

To this end we first formulate the rheology as:

$$\begin{aligned} \tau_{ij} &= 2\eta^{\mathrm{ve}}\left(\dot{\varepsilon}'_{ij}\right)\left[\dot{\varepsilon}'_{ij} - \dot{\varepsilon}'^{\ \mathrm{vp}}_{ij}\left(\dot{\varepsilon}'_{ij}, p^{\mathrm{trial}}\right)\right] \\ p^{\mathrm{corr}} &= p^{\mathrm{trial}} + K\Delta t v^{\mathrm{vp}}_{k,k}\left(\dot{\varepsilon}'_{ij}, p^{\mathrm{trial}}\right) = -K\Delta t\left[v'_{k,k} - v^{\mathrm{vp}}_{k,k}\left(\dot{\varepsilon}'_{ij}, p^{\mathrm{trial}}\right)\right], \end{aligned} \tag{B.1}$$

where $v'_{k,k} = v_{k,k} - \frac{p^0}{K\Delta t}$. Different from the preceding section, the viscoplastic strain rates and divergence appear explicitly in the deviatoric stress update. There is no need to define an effective V-E-VP viscosity as the trial visco-elastic viscosity will be used. The constitutive relation can be recast further as:

$$\boldsymbol{\tau} = \mathbf{D}^{\mathrm{ve}}\left(\dot{\boldsymbol{\varepsilon}} - \dot{\boldsymbol{\varepsilon}}^{\mathrm{p}}\right) = \mathbf{D}^{\mathrm{ve}}\dot{\boldsymbol{\varepsilon}} - \dot{\lambda}\mathbf{D}^{\mathrm{ve}}\mathbf{m} \tag{B.2}$$

where $\mathbf{m} = \left[\frac{\partial Q}{\partial \tau_{xx}} \frac{\partial Q}{\partial \tau_{yy}} \frac{\partial Q}{\partial \tau_{zz}} \frac{\partial Q}{\partial \tau_{xy}} - \frac{\partial Q}{\partial p^{\mathrm{trial}}}\right]^T$, and thus contains derivatives with regard to the stress deviators as well as the pressure. Due to power-law viscous rheology, $\mathbf{D}^{\mathrm{ve}}$ depends on the deviatoric strain rate, thus on $\dot{\boldsymbol{\varepsilon}}$, where $\boldsymbol{\tau} = \left[\tau_{xx}\ \tau_{yy}\ \tau_{zz}\ \tau_{xy}\ p^{\mathrm{corr}}\right]^T$, $\dot{\boldsymbol{\varepsilon}} = \left[\epsilon_{xx}\ \epsilon_{yy}\ \epsilon_{zz}\ \epsilon_{xy}\ v'_{k,k}\right]^T$. The operator $\mathbf{D}^{\mathrm{ve}}$ is expressed as:

$$\mathbf{D}^{\mathrm{ve}} = \begin{bmatrix} 2\eta^{\mathrm{ve}} & 0 & 0 & 0 & 0 \\ 0 & 2\eta^{\mathrm{ve}} & 0 & 0 & 0 \\ 0 & 0 & 2\eta^{\mathrm{ve}} & 0 & 0 \\ 0 & 0 & 0 & 2\eta^{\mathrm{ve}} & 0 \\ 0 & 0 & 0 & 0 & -K\Delta t \end{bmatrix}, \tag{B.3}$$

For generality, we will consider the case of power-law visco-elastic flow, so that η^{ve} is a non-linear function of the strain rate.

We consider an infinitesimal perturbation of stress-strain rate relationship such that:

$$\delta\boldsymbol{\tau} = \mathbf{D}^{\mathrm{ve}}\delta\dot{\boldsymbol{\varepsilon}} + \delta\mathbf{D}^{\mathrm{ve}}\dot{\boldsymbol{\varepsilon}} - \delta\dot{\lambda}\mathbf{D}^{\mathrm{ve}}\mathbf{m} - \delta\mathbf{D}^{\mathrm{ve}}\dot{\lambda}\mathbf{m} - \dot{\lambda}\mathbf{D}^{\mathrm{ve}}\frac{\partial \mathbf{m}}{\partial \boldsymbol{\tau}}\delta\boldsymbol{\tau}. \tag{B.4}$$

The above expression can be recast as:

$$\delta\boldsymbol{\tau} = \mathbf{E}^{-1}\mathbf{D}^{\mathrm{ve}}\delta\dot{\boldsymbol{\varepsilon}} + \mathbf{E}^{-1}\delta\mathbf{D}^{\mathrm{ve}}\dot{\boldsymbol{\varepsilon}}^{\mathrm{ve}} - \mathbf{E}^{-1}\mathbf{D}^{\mathrm{ve}}\mathbf{m}\delta\dot{\lambda} \tag{B.5}$$

where $\mathbf{E} = \left(\mathbf{I} + \mathbf{D}^{\mathrm{ve}}\dot{\lambda}\frac{\partial \mathbf{m}}{\partial \boldsymbol{\tau}}\right)$ and $\dot{\boldsymbol{\varepsilon}}^{\mathrm{ve}} = \dot{\boldsymbol{\varepsilon}} - \dot{\boldsymbol{\varepsilon}}^{\mathrm{p}}$. The consistency condition implies that $\delta F = 0$ during plastic loading, hence:

$$\left(\frac{\partial F}{\partial \boldsymbol{\tau}}\right)^T \delta\boldsymbol{\tau} + \frac{\partial F}{\partial \dot{\lambda}}\delta\dot{\lambda} + \frac{\partial F}{\partial \lambda}\delta\lambda = 0. \tag{B.6}$$

The infinitesimal perturbation of plastic multiplier rate, $\delta\dot{\lambda}$, can be obtained by introducing infinitesimal perturbation of stress-strain rate relationship into the consistency condition:

$$\delta\dot{\lambda} = \frac{\left(\frac{\partial F}{\partial \boldsymbol{\tau}}\right)^T \mathbf{E}^{-1}\mathbf{D}^{\mathrm{ve}}\delta\dot{\boldsymbol{\varepsilon}} + \left(\frac{\partial F}{\partial \boldsymbol{\tau}}\right)^T \mathbf{E}^{-1}\delta\mathbf{D}^{\mathrm{ve}}\dot{\boldsymbol{\varepsilon}}^{\mathrm{ve}}}{\left(\frac{\partial F}{\partial \boldsymbol{\tau}}\right)^T \mathbf{E}^{-1}\mathbf{D}^{\mathrm{ve}}\mathbf{m} - \frac{\partial F}{\partial \dot{\lambda}} - \Delta t \frac{\partial F}{\partial \lambda}}, \tag{B.7}$$

where we have assumed that $\delta\lambda = \delta\dot{\lambda}\Delta t$.

Finally, the consistent tangent operator may be expressed by substituting $\delta\dot{\lambda}$ into expression B.5 and by letting $\mathbf{D}^{\mathrm{ctl}} \equiv \frac{\partial \boldsymbol{\tau}}{\partial \dot{\boldsymbol{\varepsilon}}}$:

$$\mathbf{D}^{\mathrm{ctl}} = \mathbf{E}^{-1}\mathbf{B} + \mathbf{E}^{-1}\mathbf{D}^{\mathrm{ve}} \left(\mathbf{I} - \frac{\mathbf{m}\left(\frac{\partial F}{\partial \boldsymbol{\tau}}\right)^T \mathbf{E}^{-1}\left(\mathbf{B} + \mathbf{D}^{\mathrm{ve}}\right)}{\left(\frac{\partial F}{\partial \boldsymbol{\tau}}\right)^T \mathbf{E}^{-1}\mathbf{D}^{\mathrm{ve}}\mathbf{m} - \frac{\partial F}{\partial \dot{\lambda}} - \Delta t \frac{\partial F}{\partial \lambda}} \right) \tag{B.8}$$

where $\mathbf{B} = \frac{\partial \mathbf{D}^{\mathrm{ve}}\dot{\boldsymbol{\varepsilon}}^{\mathrm{ve}}}{\partial \tau}$, which can be written explicitly written as:

$$\mathbf{B} = \begin{bmatrix} 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{xx}}\dot{\varepsilon}^{\mathrm{ve}}_{xx} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{yy}}\dot{\varepsilon}^{\mathrm{ve}}_{xx} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{zz}}\dot{\varepsilon}^{\mathrm{ve}}_{xx} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{xy}}\dot{\varepsilon}^{\mathrm{ve}}_{xx} & 0 \\ 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{xx}}\dot{\varepsilon}^{\mathrm{ve}}_{yy} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{yy}}\dot{\varepsilon}^{\mathrm{ve}}_{yy} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{zz}}\dot{\varepsilon}^{\mathrm{ve}}_{yy} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{xy}}\dot{\varepsilon}^{\mathrm{ve}}_{yy} & 0 \\ 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{z}}\dot{\varepsilon}^{\mathrm{ve}}_{zz} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{yy}}\dot{\varepsilon}^{\mathrm{ve}}_{zz} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{zz}}\dot{\varepsilon}^{\mathrm{ve}}_{zz} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{xy}}\dot{\varepsilon}^{\mathrm{ve}}_{zz} & 0 \\ 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{xx}}\dot{\varepsilon}^{\mathrm{ve}}_{xy} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}_{yy}}\dot{\varepsilon}^{\mathrm{ve}}_{xy} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{zz}}\dot{\varepsilon}^{\mathrm{ve}}_{xy} & 2\frac{\partial \eta^{\mathrm{ve}}}{\partial \dot{\varepsilon}'_{xy}}\dot{\varepsilon}^{\mathrm{ve}}_{xy} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{B.9}$$

The partial derivatives of the trial viscosity with regard to effective deviatoric strain rates are given in the previous appendix section. In practice, one may further reduce the size of systems to 3×3 by introducing constraint on out-of-plane deformation. The inverse of the matrix $\mathbf{E}$ can then be computed analytically.

The continuum tangent operator (Cont. Tangent) can formulated by neglecting the second order derivatives of the plastic flow potential ($\frac{\partial \mathbf{m}}{\partial \boldsymbol{\tau}} = \mathbf{0}$) and thus setting $\mathbf{E} = \mathbf{I}$.

C: Governing equations for lithospheric models (MDoodz6.0)

The models presented in Sec 5.1 were intended to show the evolution of a V-E-VP in a state-of-the-art geodynamic simulation, thus involving large deformations and com-

posite rheological modelling. The following set of thermo-mechanical equations were solved:

$$\begin{aligned} \frac{\partial \tau_{ij}}{\partial x_j} - \frac{\partial p}{\partial x_i} &= -\rho g_i, \\ \frac{1}{K}\frac{\mathrm{d}p}{\mathrm{d}t} &= -\left(v_{i,i} - v_{i,i}^{\mathrm{vp}} - v_{i,i}^{\mathrm{th}}\right) \\ \rho c_{\mathrm{P}}\frac{\mathrm{d}T}{\mathrm{d}t} &= -\frac{\partial q_i}{\partial x_i} + H_{\mathrm{R}} + H_{\mathrm{S}} + H_{\mathrm{A}}, \end{aligned} \tag{C.1}$$

where T is the temperature and ρ, c_{P}, k and α are the density, the heat capacity at constant pressure, the the thermal conductivity, and the thermal expansivity, respectively. The term $q_i = -k\frac{\partial T}{\partial x_i}$ is the heat flux vector and $v_{i,i}^{\mathrm{th}} = \alpha\frac{\mathrm{d}T}{\mathrm{d}t}$ is the thermal divergence rate. H_{R} is the radioctive heating, $H_{\mathrm{S}} = \tau_{ij}\left(\dot{\varepsilon}_{ij} - \dot{\varepsilon}_{ij}^{\mathrm{e}}\right)$ is shear heating and $H_{\mathrm{A}} = \alpha T\frac{\mathrm{d}p}{\mathrm{d}t}$ is adiabatic heating.

The deviatoric stress rate includes term arising from advection, rotation and stretching:

$$\breve{\boldsymbol{\tau}} = \dot{\boldsymbol{\tau}} - \boldsymbol{v}\nabla\boldsymbol{\tau} - \nabla\boldsymbol{v}^T\cdot\boldsymbol{\tau} - \boldsymbol{\tau}\cdot\nabla\boldsymbol{v} \tag{C.2}$$

In the current implementation, this operation is split such that the update is semi-implicit. At each time step, the deviatoric stress tensor is advected and rotated using the velocity and stress field obtained after convergence of the non-linear solver. The rheological model is composite and is based on an additive decomposition the different creep strain rates. Such rheological model is used to represent the creep mechanisms in the lithospheric mantle (dislocation, diffusion and Peierls). The different mechanisms are combined within local rheological iterations and allow for determination of trial visco-elastic stress (See *Popov and Sobolev* [2008] and Sec. 13). Diffusion and dislocation creep are expressed as power-laws, such that:

$$\eta^{\mathrm{Dif,Dis}} = C^{\mathrm{Dif,Dis}}\ \dot{\varepsilon}_{\mathrm{II}}{}^{\frac{1}{n^{\mathrm{Dif,Dis}}}-1}\ d^{\frac{m^{\mathrm{Dif,Dis}}}{n^{\mathrm{Dif,Dis}}}} \tag{C.3}$$

where d is the grain size, m is the grain size exponent, n is the stress exponent. The factor C is expressed as:

$$C^{\mathrm{Dif,Dis}} = \left[2F^{\mathrm{Dif,Dis}}\left(A^{\mathrm{Dif,Dis}}\right)^{-\frac{1}{n}}\exp\left(\frac{Q^{\mathrm{Dif,Dis}} + PV^{\mathrm{Dif,Dis}}}{n^{\mathrm{Dif,Dis}}RT}\right)\right]^{-n^{\mathrm{Dif,Dis}}} \tag{C.4}$$

where Q is activation energy, V is activation volume, F is the correction factor for conversion of experimental data to invariant formulation (see *Schmalholz and Fletcher* [2011]) and R is the gas constant. Peierls creep is implemented using the effective power-law formulation of *Kameyama et al.* [1999], the subsequent Peierls effective viscosity is thus expressed as:

$$\eta^{\mathrm{Peierls}} = C^{\mathrm{Peierls}}\ \dot{\varepsilon}_{\mathrm{II}}{}^{\frac{1}{n^{\mathrm{Peierls}}}-1} \tag{C.5}$$

Table C.1. Material parameters used in the lithospheric simulations. Additional parameters for Peierls creep are $\gamma = 0.1$ and $\sigma^{\text{Peierls}} = 8.5 \times 10^9$ Pa. Thermal expansivity, heat capacity, bulk modulus, shear modulus are assumed constant for each phases ($\alpha = 3.2 \times 10^{-5}$ K^{-1}, $c_{\text{P}} = 1050$ J.kg^{-1}.K^{-1}, $K = 2 \times 10^{10}$ Pa, $G = 3 \times 10^{10}$ Pa). Softening of the friction angle (upper and middle crust) occurs within an accumulated plastic strain of 0.5. Flow law parameter are taken from *Hansen and Carter* [1983] for the upper crust, Rybacki *Mackwell et al.* [1998] for the lower crust, *Hirth and Kohlstedt* [2004] and *Goetze and Evans* [1979] for the lithospheric mantle.

	A [Pa^{-n}.s^{-1}]	Q [J.mol^{-1}]	V [m^3.mol^{-1}]	n	m	C [Pa]	ϕ [°]	ψ [°]	ρ_0 [kg.m^{-3}
Upper crust	3.1623×10^{-26}	186.5×10^3	0.0	3.3	0.0	2×10^7	$30 \rightarrow 10$	5	2700
Middle crust	3.9811×10^{-16}	356×10^3	0.0	3.0	0.0	2×10^7	$30 \rightarrow 20$	5	2750
Lower crust	5.0477×10^{-28}	485×10^3	0.0	4.7	0.0	2×10^7	30	5	2800
Inclusion	1.1×10^{-16}	530×10^3	11×10^{-6}	3.5	0.0	10^6	0	0	3260
Mantle Diffusion	1.5×10^{-15}	375×10^3	4×10^{-6}	1.0	3.0	10^7	30	5	3260
Mantle Dislocation	1.1×10^{-16}	530×10^3	11×10^{-6}	3.5	0.0	10^7	30	5	3260
Mantle Peierls	5.7×10^{11}	540×10^3	–	–	-	10^7	30	5	3260

where n^{Peierls} is formulated as

$$n^{\text{Peierls}} = \frac{Q^{\text{Peierls}}}{RT} q\gamma \left(1-\gamma\right)^{q-1} \tag{C.6}$$

and

$$C^{\text{Peierls}} = A^{\text{Peierls}} \left(\gamma\sigma^{\text{Peierls}}\right)^{-n^{\text{Peierls}}} \exp\left(-\frac{Q^{\text{Peierls}}}{RT}(1-\gamma)^2\right) \tag{C.7}$$

where Q^{Peierls}, γ, σ^{Peierls} are material parameters (see table below).

The density is assumed to be pressure and temperature dependent and follows the equation of state:

$$\rho = \rho_0 \exp\left(-\alpha T + \frac{P}{K}\right) \tag{C.8}$$

where ρ_0 is the reference density.

References

Bates, D. M. (2007), Direct Methods for Sparse Linear Systems by Timothy A. Davis, *International Statistical Review*, *75*(2), 260–261.

Bauville, A., M. Furuichi, and M. (2020), Control of fault weakening on the structural styles of underthrusting-dominated non-cohesive accretionary wedges, *Journal of Geophysical Research: Solid Earth*, *125*(3), e2019JB019,220, doi: https://doi.org/10.1029/2019JB019220, e2019JB019220 10.1029/2019JB019220.

Bažant, Z. P., and F.-B. Lin (1988), Non-local yield limit degradation, *International Journal for Numerical Methods in Engineering*, *26*(8), 1805–1823.

Bessat, A., T. Duretz, G. Hetényi, S. Pilet, and S. M. Schmalholz (2020), Stress and deformation mechanisms at a subduction zone: Insights from 2-D thermomechanical numerical modelling, *Geophysical Journal International*, *221*(3), 1605–1625.

Byerlee, J. (1978), Friction of rocks, *Pure and Applied Geophysics*, *116*, 615–626.

Candioti, L. G., S. M. Schmalholz, and T. Duretz (2020), Impact of upper mantle convection on lithosphere hyper-extension and subsequent convergence-induced subduction, *Solid Earth Discussions*, *2020*, 1–41, doi:10.5194/se-2020-88.

Choi, E., and K. D. Petersen (2015), Making Coulomb angle-oriented shear bands in numerical tectonic models, *Tectonophysics*, *657*, 94–101.

Commend, S., A. Truty, and T. Zimmermann (2004), Stabilized finite elements applied to elastoplasticity: I. mixed displacementpressure formulation, *Computer Methods in Applied Mechanics and Engineering*, *193*(33), 3559–3586, doi: https://doi.org/10.1016/j.cma.2004.01.007.

Crameri, F. (2018), Geodynamic diagnostics, scientific visualisation and staglab 3.0, *Geoscientific Model Development*, *11*(6), 2541–2562.

de Borst, R., and T. Duretz (2020), On viscoplastic regularisation in strain-softening rocks and soils, *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*, 890–903.

de Borst, R., and P. H. Feenstra (1990), Studies in anisotropic plasticity with reference to the Hill criterion, *International Journal for Numerical Methods in Engineering*, *29*, 315–336.

de Borst, R., and H. B. Mühlhaus (1992), Gradient-dependent plasticity: Formulation and algorithmic aspects, *International Journal for Numerical Methods in*

Engineering, *35*, 521–539.

de la Puente, J., J.-P. Ampuero, and M. Käser (2009), Dynamic rupture modeling on unstructured meshes using a discontinuous Galerkin method, *Journal of Geophysical Research: Solid Earth*, *114*(B10302), 1–17.

Duretz, T., D. May, and P. Yamato (2016a), A free surface capturing discretization for the staggered grid finite difference scheme, *Geophysical Journal International*, *204*(3), 1518–1530.

Duretz, T., B. Petri, G. Mohn, S. M. Schmalholz, F. L. Schenker, and O. Müntener (2016b), The importance of structural softening for the evolution and architecture of passive margins, *Scientific Reports*, *6*, 38,704.

Duretz, T., A. Souche, R. de Borst, and L. Le Pourhiet (2018), The benefits of using a consistent tangent operator for viscoelastoplastic computations in geodynamics, *Geochemistry, Geophysics, Geosystems*, *19*(2018GC007877), 4904–4024.

Duretz, T., R. de Borst, and L. Le Pourhiet (2019), On finite thickness of shear bands in frictional viscoplasticity, and implications for lithosphere dynamics, *Geochemistry, Geophysics, Geosystems*, *20*(2019GC008531), 5598–5616.

Duretz, T., R. de Borst, P. Yamato, and L. Le Pourhiet (2020), Toward robust and predictive geodynamic modeling: The way forward in frictional plasticity, *Geophysical Research Letters*, *47*(5), e2019GL086,027.

Gerbault, M., A. N. B. Poliakov, and M. Daignieres (1998), Prediction of faulting from the theories of elasticity and plasticity: what are the limits?, *Journal of Structural Geology*, *20*(2), 301 – 320.

Gerya, T. V., and D. A. Yuen (2007), Robust characteristics method for modelling multiphase visco-elasto-plastic thermo-mechanical problems, *Physics of the Earth and Planetary Interiors*, *163*, 83–105.

Glerum, A., C. Thieulot, M. Fraters, C. Blom, and W. Spakman (2018), Nonlinear viscoplasticity in ASPECT: benchmarking and applications to subduction, *Solid Earth*, *9*(2), 267–294, doi:10.5194/se-9-267-2018.

Goetze, C., and B. Evans (1979), Stress and temperature in the bending lithosphere as constrained by experimental rock mechanics, *Geophysical Journal International*, *59*(3), 463–478.

Hansen, F., and N. Carter (1983), Semibrittle creep of dry and wet westerly granite at 1000 mpa.

Heeres, O. M., A. S. J. Suiker, and R. de Borst (2002), A comparison between the Perzyna viscoplastic model and the consistency viscoplastic model, *European Journal of Mechanics: A/Solids*, *21*, 1 – 12.

Hirth, G., and D. Kohlstedt (2004), *Rheology of the Upper Mantle and the Mantle Wedge: A View from the Experimentalists*, pp. 83–105, American Geophysical Union (AGU).

Huismans, R. S., and C. Beaumont (2003), Symmetric and asymmetric lithospheric extension: Relative effects of frictional-plastic and viscous strain softening, *Journal of Geophysical Research: Solid Earth*, *108*(B10), doi:10.1029/2002JB002026.

Huismans, R. S., and C. Beaumont (2007), Roles of lithospheric strain softening and heterogeneity in determining the geometry of rifts and continental margins, *Geological Society, London, Special Publications*, *282*(1), 111–138, doi: 10.1144/SP282.6.

Jacquey, A. B., and M. Cacace (2020), Multiphysics modeling of a brittle-ductile lithosphere: 1. explicit visco-elasto-plastic formulation and its numerical implementation, *Journal of Geophysical Research: Solid Earth*, *125*(1), e2019JB018,474.

Kameyama, M., D. A. Yuen, and S.-I. Karato (1999), Thermal-mechanical effects of low-temperature plasticity (the peierls mechanism) on the deformation of a viscoelastic shear zone, *Earth and Planetary Science Letters*, *168*(1), 159 – 172.

Kaus, B. J. P., A. A. Popov, T. S. Baumann, A. E. Püsök, A. Bauville, N. Fernandez, and M. Collignon (2016), Forward and inverse modelling of lithospheric deformation on geological timescales, in *NIC Symposium 2016*, vol. 48, edited by K. Binder, M. Müller, A. Kremer, and A. Schnurpfeil, pp. 299–307, Forschungszentrum Jülich, Jülich.

Kiss, D., L. G. Candioti, T. Duretz, and S. M. Schmalholz (2019), Thermal softening induced subduction initiation at a passive margin, *Geophysical Journal International*, *220*(3), 2068–2073.

Lemiale, V., H. B. Mühlhaus, L. Moresi, and J. Stafford (2008), Shear banding analysis of plastic models formulated for incompressible viscous flows, *Physics of the Earth and Planetary Interiors*, *171*, 177–186.

Mackwell, S. J., M. E. Zimmerman, and D. L. Kohlstedt (1998), High-temperature deformation of dry diabase with application to tectonics on venus, *Journal of Geophysical Research: Solid Earth*, *103*(B1), 975–984.

May, D. A., J. Brown, and L. L. Pourhiet (2014), ptatin3d: High-performance methods for long-term lithospheric dynamics, in *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 274–284, doi:10.1109/SC.2014.28.

Moresi, L., F. Dufour, and H. Muhlhaus (2003), A Lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials, *Journal of Computational Physics*, *184*, 476–497.

Mühlhaus, H. B., and I. Vardoulakis (1987), The thickness of shear bands in granular materials, *Géotechnique*, *37*, 271–283.

Naliboff, J. B., S. J. H. Buiter, G. Péron-Pinvidic, P. T. Osmundsen, and J. Tetreault (2017), Complex fault interaction controls continental rifting, *Nature Communications*, *8*, 1179.

Naliboff, J. B., A. Glerum, S. Brune, G. Pron-Pinvidic, and T. Wrona (2020), Development of 3-d rift heterogeneity through fault network evolution, *Geophysical Research Letters*, *47*(13), e2019GL086,611, doi:https://doi.org/10.1029/2019GL086611, e2019GL086611 2019GL086611.

Niazi, M. S., H. H. Wisselink, and T. Meinders (2013), Viscoplastic regularization of local damage models: Revisited, *Comput. Mech.*, *51*(2), 203–216.

Petri, B., T. Duretz, G. Mohn, S. M. Schmalholz, G. D. Karner, and O. Müntener (2019), Thinning mechanisms of heterogeneous continental lithosphere, *Earth and Planetary Science Letters*, *512*, 147 – 162.

Poh, J., P. Yamato, T. Duretz, D. Gapais, and P. Ledru (2020), Precambrian deformation belts in compressive tectonic regimes: A numerical perspective, *Tectonophysics*, *777*, 228,350.

Poliakov, A., Y. Podladchikov, and C. Talbot (1993), Initiation of salt diapirs with frictional overburdens: numerical experiments, *Tectonophysics*, *228*(3), 199 – 210.

Popov, A. A., and S. V. Sobolev (2008), SLIM3D: A tool for three-dimensional thermomechanical modeling of lithospheric deformation with elasto-visco-plastic rheology, *Physics of the Earth and Planetary Interiors*, *171*, 55–75.

Räss, L., T. Duretz, Y. Y. Podladchikov, and S. M. Schmalholz (2017), M2Di: Concise and efficient MATLAB 2-D Stokes solvers using the Finite Difference Method, *Geochemistry, Geophysics, Geosystems*, *18*(2), 755–768.

Räss, L., T. Duretz, and Y. Y. Podladchikov (2019), Resolving hydromechanical coupling in two and three dimensions: spontaneous channelling of porous fluids owing to decompaction weakening, *Geophysical Journal International*, *218*(3), 1591–1616, doi:10.1093/gji/ggz239.

Rudnicki, J. W., and J. R. Rice (1975), Conditions for the localization of deformation in pressure sensitive dilatant materials, *Journal of the Mechanics and Physics of Solids*, *23*, 371–394.

Rutter, E. H., and C. T. Glover (2012), The deformation of porous sandstones: are Byerlee friction and the critical state line equivalent?, *Journal of Structural Geology*, *44*, 129 – 140.

Sabet, S. A., and R. de Borst (2019), Structural softening, mesh dependence, and regularisation in non-associated plastic flow, *International Journal for Numerical and Analytical Methods in Geomechanics*, *43*, 2170–2183.

Schmalholz, S., T. Duretz, F. Schenker, and Y. Podladchikov (2014), Kinematics and dynamics of tectonic nappes: 2-d numerical modelling and implications for high and ultra-high pressure tectonism in the western alps, *Tectonophysics*, *631*(C), 160–175, doi:10.1016/j.tecto.2014.05.018.

Schmalholz, S. M., and T. Duretz (2017), Impact of grain size evolution on necking in calcite layers deforming by combined diffusion and dislocation creep, *Journal of Structural Geology*, *103*, 37 – 56.

Schmalholz, S. M., and R. C. Fletcher (2011), The exponential flow law applied to necking and folding of a ductile layer, *Geophysical Journal International*, *184*(1), 83–89.

Shipton, Z. K., A. M. Soden, J. D. Kirkpatrick, A. M. Bright, and R. J. Lunn (2006), How thick is a fault? Fault displacement-thickness scaling revisited, in *Earthquakes: Radiated Energy and the Physics of Faulting*, edited by R. Abercrombie, A. McGarr, G. Di Toro, and H. Kanamori, pp. 193–198, American Geophysical Union (AGU).

Spiegelman, M., D. A. May, and C. R. Wilson (2016), On the solvability of incompressible Stokes with viscoplastic rheologies in geodynamics, *Geochemistry, Geophysics, Geosystems*, *17*, 2213–2238.

Stefanou, I., J. Sulem, and H. Rattez (2017), *Cosserat Approach to Localization in Geomaterials*, pp. 1–25, Springer International Publishing, Cham, doi:

10.1007/978-3-319-22977-5_10-1.

Wang, H. (2019), Viscous and second gradient regularization techniques for the description of the behavior of geomaterials, Theses, École centrale de Nantes.

Wang, W. M., L. J. Sluys, and R. de Borst (1997), Viscoplasticity for instabilities due to strain softening and strain-rate softening, *International Journal for Numerical Methods in Engineering*, *40*(20), 3839–3864.

Willett, S. D. (1992), Dynamic and kinematic growth and change of a Coulomb wedge, in *Thrust Tectonics*, edited by K. R. McClay, pp. 19–31, Springer, Dordrecht.

Yamato, P., T. Duretz, and S. Angiboust (2019), Brittle/ductile deformation of eclogites: Insights from numerical models, *Geochemistry, Geophysics, Geosystems*, *20*(7), 3116–3133.

Zhao, X., and M. Cai (2010), A mobilized dilation angle model for rocks, *International Journal of Rock Mechanics and Mining Sciences*, *47*(3), 368 – 384.

Zoback, M. D., R. Apel, J. Baumgärtner, M. Brudy, R. Emmermann, B. Engeser, K. Fuchs, W. Kessels, H. Rischmüller, F. Rummel, and L. Vernik (1993), Upper-crustal strength inferred from stress measurements to 6 km depth in the KTB borehole, *Nature*, *365*(6447), 633–635.

Acknowledgments

The authors acknowledge Boris Kaus and John Naliboff for reviewing an earlier version of this manuscript. The authors acknowledge Rennes Métropole and the French programme NEEDS for their financial supports to the project. The authors thank Carolina Lithgow-Bertelloni for editing the manuscript.The research reported in this article has been partially supported by the European Research Council under grant 664734 "PoroFrac". Author TD thanks Prof. Taras Gerya for exciting discussions. This study is based on forward numerical simulations and does not involve any new data.