Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE
10.1029/2017JB015225

An Invariant Rate- and State-Dependent Friction Formulation for Viscoeastoplastic Earthquake Cycle Simulations

Robert Herrendörfer[1,2], **Taras Gerya**[1], **and Ylona van Dinther**[2]

[1]Geophysical Fluid Dynamics, Institute of Geophysics, Department of Earth Sciences, ETH Zurich, Zurich, Switzerland, [2]Seismology and Wave Physics, Institute of Geophysics, Department of Earth Sciences, ETH Zurich, Zurich, Switzerland

Abstract We present a 2-D numerical modeling approach for simulating a wide slip spectrum in a viscoelastoplastic continuum. The key new model component is an invariant reformulation of the classical rate- and state-dependent friction equations, which is designed for earthquake simulations along spontaneously evolving faults. Here we describe the methodology and demonstrate that it is accurate and stable in a setup consisting of a mature strike-slip fault zone. We show that the nucleation and propagation of an earthquake are well resolved, as supported by a good agreement with various analytical approximations, including those of the nucleation and cohesive zone lengths. Results generally converge with respect to grid size, time step, and other numerical parameters. The convergence rate with respect to grid size depends on the internodal averaging scheme, is influenced by wave reflections, and deteriorates for inclined faults. The simulated slip spectrum, ranging from stable sliding at the loading rate to periodic aseismic slip to periodic seismic slip as a function of nucleation size, is in general agreement with the literature. In this simple setup, dynamic pressure does not play a significant role. By analyzing the role of viscous deformation, we identify and confirm by our simulations a theoretical viscosity threshold below which earthquakes cannot nucleate. This threshold is shown to depend on the reference strength of rate- and state-dependent friction and the loading strain rate, which is in agreement with previous work on the brittle-ductile transition.

Key Points:
- We present an invariant rate- and state-dependent friction formulation for simulations in a continuum mechanics framework
- Numerical accuracy and stability is demonstrated for seismic slip on a straight fault
- Based on our simulations, we derive a viscosity threshold, below which no earthquakes are generated

Supporting Information:
- Supporting Information S1

Correspondence to:
R. Herrendörfer and T. Gerya,
robert.herrendoerfer@gmx.ch;
taras.gerya@erdw.ethz.ch

Citation:
Herrendörfer, R., Gerya, T., & van Dinther, Y. (2018). An invariant rate- and state-dependent friction formulation for viscoelastoplastic earthquake cycle simulations. *Journal of Geophysical Research: Solid Earth*, *123*, 5018–5051. https://doi.org/10.1029/2017JB015225

Received 13 NOV 2017
Accepted 11 MAY 2018
Accepted article online 17 MAY 2018
Published online 16 JUN 2018

1. Introduction

Earthquakes pose a great societal and economic hazard as they cause severe damage in increasingly populated areas. This has been demonstrated by recent large earthquakes in different tectonic settings, such as the 2011 Tohoku earthquake (subduction zone), the 2016 Nepal earthquake (collision zone), the 1999 Izmit earthquake (strike-slip fault), and the 2016–2017 Central Italian earthquakes (normal fault). A crucial part of the seismic hazard assessment is to identify the conditions under which such destructive earthquakes occur. Shortly after the acceptance of plate tectonics, it has been proposed that these earthquake prone conditions are related to a region's tectonic characteristics. However, the popular hypotheses relating slab age and subduction velocity to maximum magnitude were proven to be flawed when the 2004 M9.2 Sumatra (Stein & Okal, 2007) and 2011 M9.0 Tohoku (Kagan & Jackson, 2013) earthquakes occurred in areas forecasted to be very unlikely to host those. Reanalysis of these parameters in global databases have revealed generally weak correlations with seismic observations (i.e., R~0–0.5, Heuret et al., 2011). These correlations are clouded due to at least two main limitations of the available record of observations. First each fault setting is unique in its set of tectonic characteristics. Hence, the observed seismicity is the result of the combined effect of various parameters, whose isolated role remains hidden. Second, we only have decent observational records, while each fault zone is currently at a different stage within its earthquake cycle. Consequently, we need to merge the inherently different snapshots from different fault zones together.

To overcome these two limitations, modeling is a key additional tool, since it can cover a long enough time period with many earthquake cycles and it can investigate the role of a single parameter at a time (e.g., Lapusta et al., 2000; Wang, 2007; van Dinther, Gerya, Dalguer, Mai, et al., 2013). Furthermore, models can provide quantitative insights into the physical processes that are active along the unreachable, deeply buried parts of a fault. It can test a hypothesis based on observations or theoretical considerations and provide testable predictions for future research. To better understand the relationship between tectonic characteristics and the seismogenic potential in realistic settings, an earthquake cycle model should account for a sufficiently

complex fault system embedded in a three-dimensional, viscoelastoplastic medium, which is able to represent the heterogeneity of a strike-slip fault or subduction zone. To obtain insights into the feedback between loading conditions, fault zone characteristics and earthquake activity, it is essential to produce a self-consistent model. In such a model, a fault system and local conditions evolve as a response to tectonic loading and characteristics. In turn, the dynamically evolving fault system and local conditions influence the earthquake potential and slip mode. For this purpose, the model should resolve all relevant physical processes, whose time scales range from million years (geodynamic time scale) to milliseconds (coseismic time scale; e.g., Wang, 2007; van Dinther, Gerya, Dalguer, Mai, et al., 2013). Furthermore, it should incorporate a fault constitutive model that is in agreement with laboratory experiments and that allows for different slip modes. With such an ideal model one could address some fundamental questions. What is the role of earthquakes and their properties for the long-term evolution of a tectonic setting? How important are large-scale tectonic characteristics, which evolve over million years, for earthquakes and their properties. However, a modeling approach that incorporates all these challenging ingredients does not exist yet, despite solid attempts (Lapusta et al., 2000; Sobolev & Muldashev, 2017; van Dinther, Gerya, Dalguer, Corbi, et al., 2013; van Dinther, Gerya, Dalguer, Mai, et al., 2013).

The overall goal of this paper is to bring two modeling approaches closer to profit from their strengths and to overcome their limitations in the process toward developing such a cross-scale, multiphysical model. These two approaches are (i) classical earthquake cycle simulations using rate- and state-dependent friction (RSF) along prescribed fault(s) (e.g., Lapusta et al., 2000) and (ii) the recently developed seismo-thermo-mechanical (STM) approach (van Dinther, Gerya, Dalguer, Corbi, et al., 2013; van Dinther, Gerya, Dalguer, Mai, et al., 2013) based on geodynamic modeling in tectonically realistic settings (Gerya & Yuen, 2007). In the following, we review both approaches, formulate the main objectives, and provide the outline of this study.

1.1. Earthquake Cycle Simulations With RSF

Numerical modeling of earthquakes has contributed in the last decades to our understanding about important processes leading to and resulting from an earthquake (Lapusta & Barbot, 2012). These processes, which are often summarized to the term earthquake cycle, include the following: interseismic loading, the nucleation of an earthquake, dynamic rupture propagation, and postseismic deformation. In classical simulations of earthquake cycles (e.g., Ben-Zion & Rice, 1997; Lapusta & Liu, 2009; Lapusta et al., 2000; Liu & Rice, 2007; Rice, 1993), a fault is treated as a predefined, discrete (i.e., infinitely thin) plane embedded in a homogeneous elastic medium. Slip along that fault is typically assumed to be governed by the RSF formulation (Dieterich, 1978, 1979; Ruina, 1983). This friction formulation is based on "slide-hold slide" and "velocity stepping" rock experiments, initially conducted by Dieterich (1972, 1978) and confirmed in subsequent studies (see review by 1998). In this formulation, friction μ relates shear stress τ_s to normal stress σ_n and depends on the magnitude of slip rate V and the state θ as

$$\tau_s = \mu\,\sigma_n = \left[\mu_0 + a\ln\left(\frac{V}{V_0}\right) + b\ln\left(\frac{\theta V_0}{L}\right)\right]\sigma_n, \tag{1}$$

where L is the characteristic slip distance and μ_0 is the reference friction coefficient defined at an arbitrary steady state slip rate V_0. The term $a\ln\left(\frac{V}{V_0}\right)$ is called the instantaneous "viscosity-like" direct effect (Rice & Ruina, 1983), because it represents the immediate response of μ, and hence τ_s, to a change in V, which is proportional to a. The term $b\ln\left(\frac{\theta V_0}{L}\right)$ is referred to as the evolution effect as it is described by the evolving state variable θ. Different evolution laws have been proposed to parameterize the change of θ as a function of time (Bhattacharya et al., 2015): the aging law (Ruina, 1983), which is based on the observation of time-dependent healing at stationary contact (Dieterich, 1972), the slip law and the Nagata law (Nagata et al., 2012). Earthquake cycle simulations commonly apply the aging (or slowness) law (e.g., Lapusta et al., 2000; Liu & Rice, 2007) defined as

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = 1 - \frac{V\theta}{L}. \tag{2}$$

At constant slip velocity, state evolves toward a steady state $\theta_{ss} = \frac{L}{V}$ (e.g., Rice & Ruina, 1983). The corresponding steady state friction coefficient is defined as

$$\mu_{ss} = \mu_0 + (a-b)\ln\left(\frac{V}{V_0}\right). \tag{3}$$

Friction is rate-weakening if $a - b < 0$, and rate-strengthening if $a - b > 0$.

Despite its long history, RSF is still empirical and lacks a general physical explanation. Nevertheless, some physical interpretations of the terms in equation (1) have been made in the past. The state variable has been related to the age of the asperities along a sliding surface (Dieterich, 1981; Dieterich & Kilgore, 1994) and to porosity (Sleep, 1995). L was shown to correlate with the roughness of the frictional surface and the particle size of the gouge along that surface (Dieterich, 1979, 1981). The direct effect has been interpreted as an Arrhenius type thermally activated rate process of forward and backward disclocation jumps at asperity contacts (e.g., Chester, 1994). This interpretation has led to the regularized version of RSF (Lapusta et al., 2000; Rice et al., 2001):

$$\tau_s = a\,\sigma_n \operatorname{arcsinh}\left[\frac{V}{2V_0}\exp\left(\frac{\mu_0 + b\ln\left(\frac{\theta V_0}{L}\right)}{a}\right)\right] \tag{4}$$

The regularized version overcomes the deficiency that the original version is ill-posed at $V = 0$ and can lead to negative friction for $V \ll V_0$. The difference between the standard and regularized versions becomes negligible for V approaching V_0. Furthermore, recent studies have made progress in deriving a microphysical model that would ultimately allow to scale laboratory results to nature (e.g., Chen & Spiers, 2016).

One limitation of RSF is that it is based on experiments conducted at slip rates much lower than seismic rates. The logarithmic weakening of RSF with slip velocity is weaker than the weakening observed in high slip rate experiments, such additional weakening mechanism have been explored in earthquake cycle simulations (see review by Lapusta & Barbot, 2012). Nevertheless, results from earthquake cycle simulations with RSF equation (4) agree with observations in terms of slip rate, stress drop, amount of slip, rupture speed, and accelerating postseismic slip after an earthquake (see review by Lapusta & Barbot, 2012). Not only earthquake slip but also different parts of the slip spectrum including slow slip transients observed in nature (Peng & Gomberg, 2010) may be explained using the rate-and-state fault model, as confirmed by laboratory experiments (Leeman et al., 2016) and numerical simulations (Liu & Rice, 2007).

Past studies have identified length and time scales, which are crucial to numerically resolve the earthquake cycle in space and time. The first length scale is the minimum size of a slipping patch required for the transition from stable sliding to instable slip along a rate-weakening fault, which is called the nucleation length (Ampuero & Rubin, 2008; Dieterich, 1992; Kaneko et al., 2008; Lapusta, 2003; Rice, 1993; Rice & Ruina, 1983; Rice et al., 2001; Rubin & Ampuero, 2005; Ruina, 1983). The second length scale, which is smaller than the first one, is the cohesive zone size at the dynamically propagating rupture front, along which the stress linearly drops with slip from its maximum to dynamic value (Cocco & Bizzarri, 2002; Cocco et al., 2004; Day et al., 2005; Lapusta & Liu, 2009). Resolving these length scales with a fine enough spatial resolution allows to reach a continuum limit as opposed to inherently discrete systems (e.g., Ben-Zion & Rice, 1993, 1997; Lapusta & Liu, 2009; Lapusta et al., 2000; Rice, 1993). In a typical earthquake cycle, the slip velocities change from less than 10^{-9} m/s in the interseismic period, and 10^{-8}–10^{-3} m/s during the accelerating nucleation and decelerating postseismic phases, to 10^{-2}–10^{0} m/s during the dynamic rupture propagation. Lapusta et al. (2000) showed that this range of slip rate is resolvable in time due to the existence of the direct effect of RSF. To capture and resolve this more than 9 orders of magnitude range in slip velocities in time, they derived an adaptive time step, which is inversely proportional to the slip rate and is a function constitutive parameters.

1.2. Seismo-Thermo-Mechanical Modeling Approach

The benefits of the classical earthquake cycle modeling approach lie in its simplicity, accuracy, and computational speed. However, the assumption of a predefined fault plane embedded in an elastic homogeneous medium makes it impossible to investigate the interaction between earthquakes with tectonic loading and the long-term evolution of fault networks in a rheologically complex plate boundary environment. Several key aspects with respect to faults and their surrounding medium are not taken into account, including, amongst others, the structure of wide fault zones with a damage zone, evolving fault properties and geometry, and a heterogeneous off-fault viscoelastoplastic medium. Recent studies have tackled some of these limitations by including material heterogeneities in the off-fault medium (Erickson & Dunham, 2014; Kaneko et al., 2011), viscous deformation (Lambert & Barbot, 2016), off-fault plasticity (Erickson et al., 2017), and evolving fault zone structures using a damage rheology model (Lyakhovsky & Ben-Zion, 2009; Lyakhovsky et al., 2016). However, none of these studies include loading of a fault system (e.g., slab pull forces in subduction zones) in a realistic tectonic environment, which has evolved over million years.

To address some of these key aspects, the "seismo-thermo-mechanical" (STM) modeling approach has been recently developed to combine the advantages of long-term geodynamic modeling and short-term earthquake cycle simulations (van Dinther, Gerya, Dalguer, Mai, et al., 2013). This approach is based on continuum mechanics (Gerya, 2010; Gerya & Yuen, 2003, 2007) and allows for the simulation of viscoelastoplastic deformation in response to and interaction with evolving tectonic forces, temperatures, and (pore fluid) pressures. The deformation spectrum includes, amongst others, elastic bending of the lithosphere, the spontaneous generation and evolution of fault systems, and linear and nonlinear viscous deformation of the mantle and lower crust. Furthermore, other long-term processes can be accounted for such as erosion, sedimentation, melting, and phase transitions (e.g., serpentinization and eclogitization). van Dinther, Gerya, Dalguer, Corbi, et al. (2013) extended this geodynamic approach by implementing inertia and a strongly rate-dependent friction formulation to simulate slip transients along spontaneous rupture paths. This STM approach has first been applied to reproduce earthquake cycles in an analogue experiment (Corbi et al., 2013; van Dinther, Gerya, Dalguer, Corbi, et al., 2013). The same setup was used to investigate the role of the seismogenic zone width and loading velocity in subduction zones (Corbi et al., 2017; Herrendörfer et al., 2015). van Dinther, Gerya, Dalguer, Corbi, et al. (2013) demonstrated in a realistic subduction zone setting that the STM approach can simulate a spontaneous brittle-ductile transition along the megathrust and discussed the weakness of faults due to pore fluids with respect to the long-term subduction process and the short-term seismicity. In addition, van Dinther et al. (2014) simulated seismicity on splay and outer-rise normal faults and its interaction with megathrust seismicity. More recently, both Gutenberg-Richter and characteristic earthquake statistics have been simulated in a continental-collision setting (Dal Zilio et al., 2018).

A distinct limitation of the STM modeling approach so far is that the coseismic duration (years), slip velocity (10^{-8} m/s), and rupture velocity (10^{-5} m/s) of these slip transients are many orders of magnitude too slow in comparison to earthquakes in nature. This resulted mainly from the absence of an adaptive time stepping and the lack of appropriate time and length scales within the rate-dependent friction formulation.

1.3. Objectives and Outline

The core of the paper is the presentation of new implementations in the STM approach, which allows us to overcome this limitation and to simulate realistic earthquake cycles in a continuum. The goal of this paper is to explore the merits and limitations of our new approach by testing its numerical accuracy and stability. To facilitate rigorous testing, we choose an idealized model setup of mature strike-slip fault zone, for which we reduced the model complexity with respect to van Dinther, Gerya, Dalguer, Corbi, et al. (2013) by excluding the evolution of temperature and other features of STM. In section 2, we begin with describing the physical model of our continuum mechanics framework, including the newly implemented compressibility for simulating pressure waves. This is followed by the presentation of the key improvement in our STM-RSF approach—that is the development and implementation of an invariant reformulation of the classical RSF equations. We discuss the key strength of this invariant formulation to simulate spontaneously evolving faults and rupture paths in a continuum, while keeping the ability to resolve earthquakes under consideration of the length and time scales inherent to RSF. Then, after providing an overview of the numerical implementation, we focus our discussion on (i) the numerical representation of a fault zone in our continuum approach, (ii) the internodal interpolation of stress and viscosity, (iii) the adaptive time step, and (iv) and the improved treatment of plasticity using grid-based iterations. Section 3 describes the reference model setup and contains a thorough analysis of all stages of a simulated earthquake cycle, including the initial loading phase, nucleation process, dynamic rupture propagation, seismic wave propagation, postseismic relaxation, and interseismic healing. Section 3 is concluded with a description of the evolution of the adaptive time step and other numerical parameters. Section 4 investigates the numerical stability of our calculations by conducting a convergence study with respect to several numerical parameters (e.g., grid size and time step) and procedures (e.g., internodal interpolation). In section 5, we demonstrate the capability of our approach to simulate a wide slip spectrum by analyzing the role of the nucleation size similar to Liu and Rice (2007). Furthermore, we analyze the role of dynamic pressure and compressibility. In section 6, we demonstrate the applicability of the STM-RSF approach to study the impact of viscous deformation on the earthquake cycle. The main results are discussed and summarized in sections 7 and 8, respectively.

2. Methodology

This section presents our STM-RSF modeling approach. It starts with the description of the governing equations in continuum mechanics including the conservation equations for mass and momentum and the

viscoelastoplastic constitutive relationship. Then we discuss the necessity for and the ingredients of an invariant reformulation of RSF. This is followed by an overview of the numerical implementation of the physical model with reference to details given in the supporting information. Particularly, we describe the numerical representation of the fault zone, the adaptive computational time step, and the flow of calculations.

2.1. Physical and Mathematical Model

2.1.1. Continuum Mechanics Approach

For the simulation of geodynamic processes, we consider the geological medium as continuous, which means that any variations of a property on the microscopical scale can be effectively represented by one value on the macroscopical scale. The continuity equation (5) requires that mass is conserved in all points as

$$\rho \frac{\partial v_i}{\partial x_i} = -\frac{D\rho}{Dt}, \tag{5}$$

where ρ is density, $\frac{D}{Dt}$ denotes the material time derivative, i and j are coordinate indices, x_i and x_j are spatial coordinates, and v_i is velocity. For the conservation of momentum in the gravity field and under the plain-strain assumption, we solve the momentum equation (6), which is given as

$$\frac{\partial \tau_{ij}}{\partial x_j} - \frac{\partial P}{\partial x_i} = \rho \frac{Dv_i}{Dt} - \rho g_i, \tag{6}$$

where g_i is gravity. τ_{ij} is the deviatoric stress tensor defined as

$$\tau_{ij} = \sigma_{ij} + \delta_{ij} P, \tag{7}$$

where σ_{ij} is the Cauchy stress tensor and δ_{ij} is the Kronecker delta. Pressure P is defined as the mean stress

$$P = -\frac{\sigma_{kk}}{3}, \tag{8}$$

where k is a coordinate index including x, y, and z. The negative sign indicates that pressure is positive under compression whereas normal stress is positive under extension. We assume a compressible material with the bulk modulus K defined as $\frac{1}{K} = \frac{1}{\rho}\frac{D\rho}{DP}$, such that

$$\frac{D\rho}{Dt} = \frac{\rho}{K}\frac{DP}{Dt} \tag{9}$$

To solve equations (5)–(6), we define a constitutive relationship, which relates stresses to velocities via strain rates. Strain rates $\dot{\varepsilon} = \frac{d\varepsilon}{dt}$ are defined under the assumption of infinitesimal strain ε during a time step as follows:

$$\dot{\varepsilon}_{ij} = \frac{1}{2}\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right) \tag{10}$$

A Maxwell viscoelastic body is set in series with a plastic-frictional slider. Hence, strain rate is decomposed in its elastic, viscous, and plastic components as

$$\dot{\varepsilon}'_{ij} = \dot{\varepsilon}'_{ij(\text{elastic})} + \dot{\varepsilon}'_{ij(\text{viscous})} + \dot{\varepsilon}'_{ij(\text{plastic})} \tag{11}$$

where $\dot{\varepsilon}'_{ij} = \dot{\varepsilon}_{ij} - \delta_{ij}\frac{\dot{\varepsilon}_{kk}}{3}$ is the deviatoric strain rate tensor. The relationship between total stresses and elastic strains follows Hooke's law under the assumption of an isotropic material and symmetry of the Cauchy stress and infinitesimal strain tensors:

$$\sigma_{ij} = \left(K - \frac{2G}{3}\right)\delta_{ij}\varepsilon_{kk(\text{elastic})} + 2G\varepsilon_{ij(\text{elastic})} \tag{12}$$

where G is the shear modulus. The relationship between deviatoric stresses and elastic strain rates is obtained by inserting (12) into (7) (and using the definition of P in (8)), using $\varepsilon'_{ij(\text{elastic})} = \varepsilon_{ij(\text{elastic})} - \delta_{ij}\frac{\varepsilon_{kk(\text{elastic})}}{3}$ and taking the time derivative:

$$\dot{\varepsilon}'_{ij(\text{elastic})} = \frac{1}{2G}\overset{\nabla}{\frac{D\tau_{ij}}{Dt}} \tag{13}$$

where $\overset{\nabla}{\frac{D}{Dt}}$ denotes the corotational time derivative.

The viscous component of the strain rate is represented by

$$\dot{\varepsilon}'_{ij(\text{viscous})} = \frac{1}{2\eta}\tau_{ij} \tag{14}$$

where η is the effective ductile viscosity. For plastic deformation, we define the yield function F:

$$F = \tau_{II} - \sigma_{\text{yield}}, \quad \sigma_{\text{yield}} = \mu P + C \tag{15}$$

where $\tau_{II} = \sqrt{{\tau_{xx}}^2 + {\tau_{xy}}^2}$ is the square root of the second invariant of the deviatoric stress tensor. σ_{yield} is the pressure-dependent yield strength defined following Drucker and Prager (1952), μ is friction, and C is cohesion. Plastic strain rates $\dot{\varepsilon}'_{(\text{plastic})} > 0$ if $F = 0$. Classically, $\dot{\varepsilon}'_{(\text{plastic})} = 0$ for $F < 0$. As we explain in Section 2.1.2, $F = 0$ and some plastic deformation always occur in case of rate-and state-dependent friction. We apply a nonassociated plastic flow law (e.g., Vermeer, 1998) by defining the plastic potential G, which is different from the yield function F:

$$G = \tau_{II} - \sin(\psi)\, P - \cos(\psi) C. \tag{16}$$

In a nonassociative plastic flow, the plastic strain increment occurs perpendicular to the plastic potential G. The dilation angle ψ is assumed to be zero, which means that the volume does not change during plastic yielding. Thus, deviatoric plastic strain is defined as

$$\dot{\varepsilon}'_{ij(\text{plastic})} = \chi \frac{\partial G}{\partial \tau_{ij}} = \chi \frac{\tau_{ij}}{2\tau_{II}}, \tag{17}$$

where χ is the unknown plastic multiplier to be solved for to fulfill the constitutive relationship (equation (11)) and the yield condition (equation (15)) in agreement with the continuity (5) and momentum equation (6). The square root of the second invariant of deviatoric plastic strain rate is defined as

$$\dot{\varepsilon}'_{II(\text{plastic})} = \sqrt{\left(\dot{\varepsilon}'_{xy(\text{plastic})}\right)^2 + \left(\dot{\varepsilon}'_{xx(\text{plastic})}\right)^2} = \frac{\chi}{2}. \tag{18}$$

2.1.2. Invariant Reformulation of RSF

There are key differences between the concepts that underly the RSF equation (equation (4)) used in classical seismic cycle simulations and the yield function used in our continuum approach (equation (15)). The first difference concerns the definition of deformation for slip on the one hand and for plastic strain on the other hand. In classical seismic cycle simulations, discontinuous brittle deformation occurs in form of slip along a predefined fault plane (i.e., within an infinitely thin deformation zone). Slip or slip velocity is a vector quantity and therefore occurs in a certain direction, which is constrained by the predefined fault plane. The magnitude of slip velocity enters into the RSF equations. In contrast, in continuum mechanics, plastic deformation is treated as strain, which is a volumetric deformation and represented by a tensor. Plastic deformation can occur everywhere and spontaneously localize into a shear zone or fault zone of finite thickness. The location and orientation of faulting can change through time and is thus an outcome of the simulation. In our formulation, plastic strain rates are used and the magnitude of plastic strain rated is given by the second plastic strain rate invariant. The difference in the concepts of slip and strain requires that the slip rate magnitude in the classical RSF formulation is related to the second plastic strain rate invariant in the continuum mechanics description. Here we scale the magnitude of the slip rate to the second invariant of plastic strain rate by following the same approach as Chester (1994), Sleep (1997), and Noda and Shimamoto (2012). They used the thickness of the fault zone D as a scaling factor to obtain the scalar plastic slip rate V_p:

$$V_p = 2\dot{\varepsilon}'_{II(\text{p})} D \tag{19}$$

In our general approach with no predefined faults, we assume that slip rates are nonzero everywhere. Note that this is different in the approach by Barbot and Fialko (2010) who relate slip rates to inelastic strain rates only on predefined fault surfaces. The numerical interpretation of D is discussed in section 2.2.1.

Second, the fault strength is evaluated differently. Normal stress σ_n and shear stress in RSF τ_s are related to the orientation of a predefined fault. In contrast, in the continuum approach it is common to use quantities that are invariant of the coordinate system and are able to adapt to spontaneous fault evolution. The yielding condition, such as equation (15), is defined as a function of the mean stress (P) and second stress invariant (τ_{II}). We therefore propose to use mean stress and second stress invariant instead of normal and shear stress.

For some fault orientations and loading conditions, the mean stress and second stress invariant might be different to the shear and normal stress along the evolved fault zones. However, they are equal in the restricted setting of laboratory experiments, on which the classical RSF formulation is based. Thus, to our knowledge, there is no experimental evidence to favor the one over the other.

In addition, in continuum mechanics, material has a certain cohesion following the observation of residual strength at zero pressure. In RSF, in contrast, the fault is treated as broken material with zero residual strength and hence zero cohesion. Here we follow this general line of reasoning that fault material has zero cohesion. Alternatively, Marone et al. (1992) discussed the option that cohesion can be included as a part of the state variable.

The third key difference lies in the question when plastic deformation or slip becomes active. In equation (15), the yield strength is usually treated as a threshold value of stress. Plastic deformation begins only if the second invariant of the deviatoric stress tensor is equal to the yield strength, and it stops as soon as the stress falls below the strength. Thus, plastic slip rate changes discontinuously around the yield strength. In contrast, a key assumption underlying the RSF equation (1) is that some slip always occurs if shear stress is larger than zero. This difference between these concepts was already noted by Nakatani (2001). He interprets the term $\sigma_n \left[\mu_0 + b \ln\left(\frac{\theta V_0}{L}\right)\right]$ in equation (1) as an interface strength. This interface strength is similar to a threshold value in the sense that slip rates become noticeable only when stress approaches it. Following Nakatani (2001), RSF can be regarded as a smooth version of the classical yield strength. The smoothness of the rate- and state-dependent yield strength is proportional to a. Consequently, if a tends to zero, the RSF framework approaches the classical notion of yield strength.

By taken all these differences between RSF and the usual definition of the yield strength into account, we arrive at the invariant regularized formulation of RSF after replacing (i) V by V_p, (ii) σ_n by P, and (iii) τ_s by τ_{II}:

$$\tau_{II} = \sigma_{\text{yield}} = a P \operatorname{arcsinh}\left[\frac{V_p}{2V_0}\exp\left(\frac{\mu_0 + b \ln\left(\frac{\theta V_0}{L}\right)}{a}\right)\right], \tag{20}$$

with the aging evolution law

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = 1 - \frac{V_p \theta}{L}. \tag{21}$$

Equation (20) is the yielding condition we apply, which implies that $\tau_{II} = \sigma_{\text{yield}}$, and hence, the yielding condition (15) is always fulfilled ($F = 0$), such that some plastic deformation always occurs. We refer to $\sigma_{\text{yield}}^{\text{RSF}} = P\left[\mu_0 + b \ln\left(\frac{\theta V_0}{L}\right)\right]$ as the RSF interface yield strength, following Nakatani (2001).

2.2. Numerical Implementation

2.2.1. Overview

For the numerical implementation of the governing equation, we use a modified version of the code I2ELVIS (Gerya & Yuen, 2007; van Dinther, Gerya, Dalguer, Corbi, et al., 2013). The governing mechanical equations are discretized in time with the first-order Euler-backward discretization scheme (Text S1). After time discretization and rewriting of equation (10), the constitutive relationship becomes

$$\tau_{ij} = 2\eta_{\text{vp}} Z \dot{\varepsilon}'_{ij} + \tau^0_{ij} \cdot (1 - Z), \quad Z = \frac{G\Delta t}{G\Delta t + \eta_{\text{vp}}} \tag{22}$$

η_{vp} is the so called viscoplastic viscosity defined as

$$\eta_{\text{vp}} = \eta \frac{\tau_{II}}{\eta\chi + \tau_{II}} = \eta \frac{\tau_{II}}{2\eta\dot{\varepsilon}_{II(\text{plastic})} + \tau_{II}} \tag{23}$$

The 2-D numerical grid is fully staggered, which means that pressure, velocities, and stresses are defined in different geometric points (Figure 1, see Text S2 for a detailed discussion of the grid structure). This staggered grid is a natural choice to discretize the equations for mass and momentum in space with stress-conservative finite differenes (Gerya, 2010, see Text S2 for discretization formulas and boundary conditions). The system of equations is solved for velocities and pressure in an implicit way to computer accuracy using the direct inversion method based on the Pardiso solver (Schenk & Gärtner, 2004, 2006). The flow of calculations to solve the equations for the conservation of mass and momentum coupled to RSF is described in section 2.2.5.

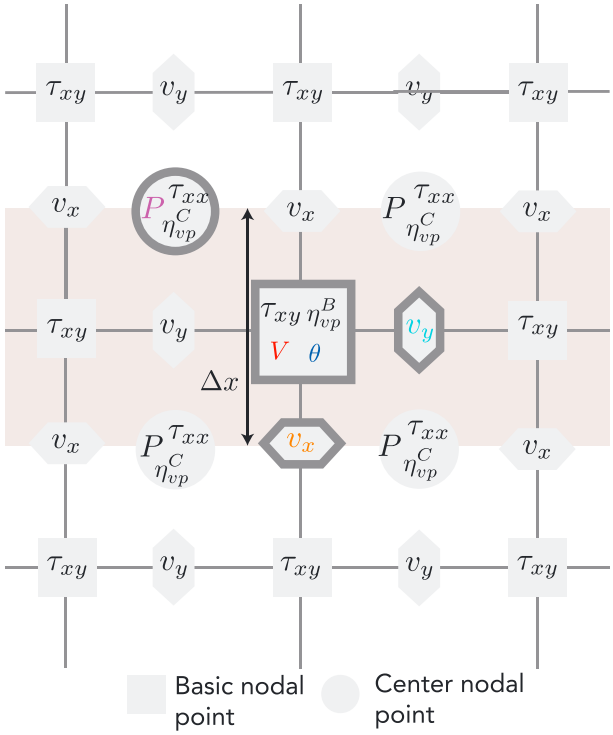

Figure 1. Numerical fault zone representation. Staggered grid along the fault zone, while the brown area denotes the fault zone thickness. τ_{xy}, V_p, and θ are defined in the basic nodal points. Pressure P and τ_{xx} are defined in center nodal points. Viscoplastic viscosity is defined in basic and center nodal points.

As in many numerical codes used in the geodynamical community, the Eulerian spatial discretization is combined with a Lagrangian marker-in-cell advection scheme (Figure 2 and Text S3). The usage of markers is important in case of large deformation compared to the grid to avoid the need of remeshing (Gerya, 2010). Although such large deformation does not occur in the present study, it is a key aspect of the STM modeling approach and in general in the geodynamic modeling community (e.g., Gerya & Yuen, 2003, 2007; Moresi et al., 2003). Markers are initially randomly distributed within the regular Eulerian grid. One parameter that influences the initial marker distribution is R_n, which is the amount of pseudo random numbers that is used to determine the random part of the marker location (see equation (S27)). The markers are assigned to rock properties (e.g., density, viscosity, shear modulus, and RSF properties). Furthermore, markers store deviatoric stresses, pressure, velocities, and state for evaluating the respective time derivates (Text S1). We have optimized the advection scheme to minimize the interpolation error between markers and nodes for the condition of small deformation. The updated advection scheme consists of the five steps shown in Figure 2 and described in detail in Text S3.

2.2.2. Numerical Fault Zone Representation

In classical continuum models without internal length scale, plastic deformation localizes to within 1–2 grid cells, and hence, the thickness of a fault zone depends on the grid size (e.g., de Borst & Sluys, 1991; Lavier et al., 2000). In this paper we therefore consider a localized zone as a mature fault zone with a thickness D equal to the grid size Δx and equation (19) becomes

$$V_p = 2\dot{\varepsilon}'_{II(\text{plastic})}\Delta x. \tag{24}$$

This definition of slip rate is similar to van Dinther, Gerya, Dalguer, Corbi, et al. (2013), who, however, used the viscoplastic strain rate. We want to point out that this relationship between slip rate and plastic strain rate may need a physics-based redefinition if deformation is distributed within more than a grid cell during the localization process toward a mature fault zone. We address this issue in more detail in a future study, where evolving faults will be studied.

The discretization of the mature fault zone using a staggered grid (Figure 1a) is very similar to the stress glut model, which assumes an inelastic zone with a thickness of Δx (Andrews, 1999; Dalguer & Day, 2006).

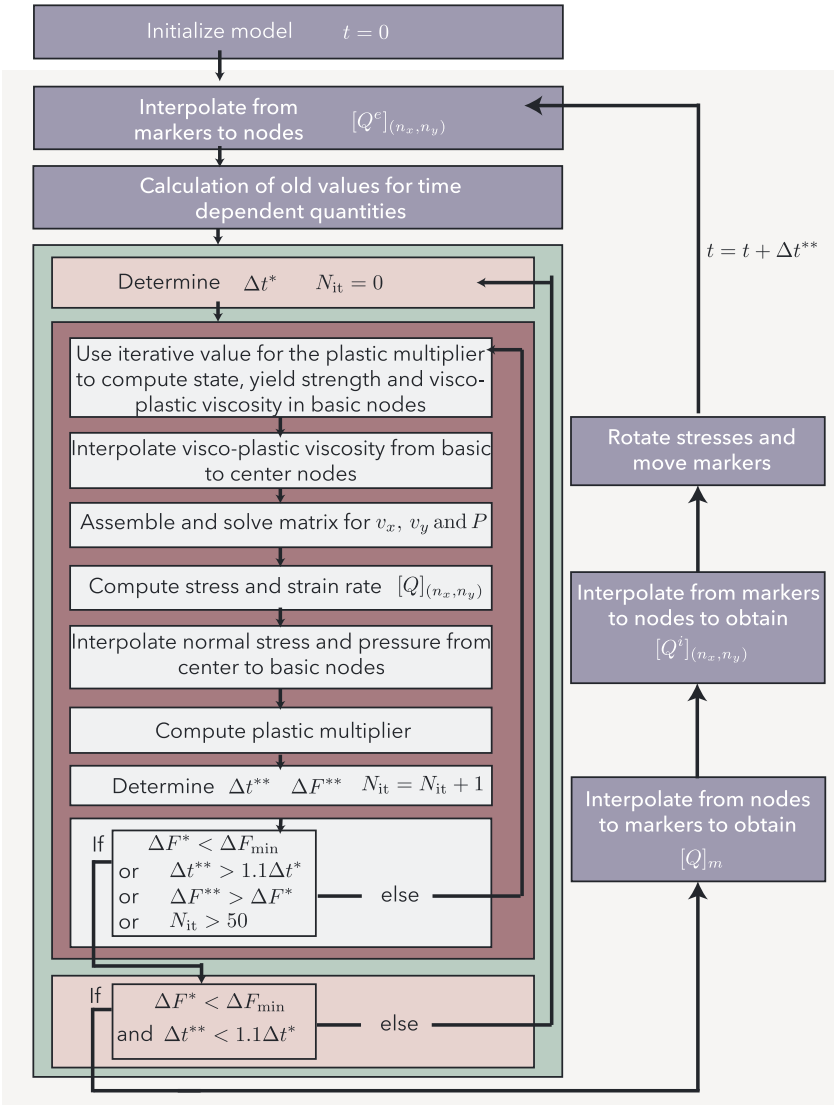

Figure 2. Flow chart for the iterative solution algorithm of the momentum and continuity equations. Steps at each time step are shown in the gray box. Marker treatment steps are shown in purple boxes. Picard iteration steps are located in the brown box. Modifications of the time step and the control over the convergence during the iterations are located in the green box.

2.2.3. Internodal Interpolation of Viscosity and Stress

To conduct iterations for solving the nonlinear problem in an accurate manner (see next subsection), plasticity is evaluated in basic nodal points in contrast to the previous STM approach, where plasticity was treated on markers (e.g., Gerya & Yuen, 2007; van Dinther, Gerya, Dalguer, Corbi, et al., 2013). Plasticity is only treated in the basic nodal points to ensure that the number of unknowns is equal to the number of equations. To calculate the square root of the second deviatoric stress invariant τ_{II} and yield strength $\sigma_{\rm yield}$ in the basic nodal points, it is necessary to interpolate normal stress and dynamic pressure from center to basic points of a grid. For both quantities, we use arithmetic averaging (Text S4). The goal of the plasticity evaluation based on the invariant formulation of RSF is to accurately calculate the viscoplastic viscosity in each basic nodal point. A stress conservative discretization of the momentum equations using the staggered grid requires that the viscoplastic viscosity is defined at both nodal and center points (Gerya, 2010, Figure 1). Thus, interpolation of the viscoplastic viscosity from basic to center points is necessary. In the geodynamic literature concerned about solving Stokes-like equations, the issue of interpolating viscosity has been addressed in a few studies (e.g., Deubelbeiss & Kaus, 2008; Duretz et al., 2011; Moresi et al., 1996; Thielmann et al., 2014). In the presence of sharp gradients in (viscoplastic) viscosity, it is not straightforward what interpolation method should be

used to achieve an accurate solution, as it also depends on the problem under consideration and where the viscosity contrast passes the numerical grid (Deubelbeiss & Kaus, 2008). Deubelbeiss and Kaus (2008) analyzed the accuracy for harmonic, arithmetic, and geometric averaging. They found that harmonic averaging between nodes yields the most accurate finite-difference results.

We deal with a similar problem, because a fault in our approach represents a layer in which the viscoplastic viscosity is many orders of magnitude lower than in the surroundings. Therefore, we evaluate the dependence of our results on the interpolation method by testing harmonic and arithmetic averaging (Text S4). For most parts of the paper we use harmonic averaging based on the suggestions of Deubelbeiss and Kaus (2008) and our experiences in simulations with evolving faults that are not parallel to the grid. In section 4, we investigate the difference between harmonic and arithmetic averaging of the viscoplastic viscosity for the convergence with respect to grid size and other numerical parameters introduced in the following sections.

2.2.4. Adaptive Time Step

We require that the time step length is the minimum of the time steps needed to resolve the state weakening (w) and healing (h), to limit displacement per grid cell (d) and the viscoelastoplastic relaxation time scale (vep) as

$$\Delta t = \zeta \min_{n_x=1,n_y=1}^{N_x,N_y} \left[\min\left[\Delta t_w, \Delta t_h, \Delta t_d, \Delta t_{\text{vep}}\right]\right] \tag{25}$$

with ζ is the time step factor, whose size is investigated in section 4. This time step is taken to be the minimum over the entire model domain. To capture the weakening of the state variable, we use the adaptive time step developed by Lapusta et al. (2000), which is inversely proportional to the slip rate:

$$\Delta t_w = \Delta\theta_{\max}\frac{L}{V_p} \tag{26}$$

The coefficient $\Delta\theta_{\max}$ is determined following Lapusta et al. (2000) and Lapusta and Liu (2009) as a function of a, b, L, P, G, and the grid size Δx and is not allowed to be larger than 0.2.

$$\xi = \frac{1}{4}\left[\frac{kL}{aP} - \frac{b-a}{a}\right]^2 - \frac{kL}{aP} \tag{27}$$

if $\xi > 0$:

$$\Delta\theta_{\max} = \min\left[\frac{aP}{kL-(b-a)P}, 0.2\right] \tag{28}$$

if $\xi < 0$:

$$\Delta\theta_{\max} = \min\left[1 - \frac{(b-a)P}{kL}, 0.2\right] \tag{29}$$

with

$$k = \frac{2}{\pi}\frac{G^*}{\Delta x} \tag{30}$$

being the stiffness of the system under the assumption of a cellular basis of slip (i.e., slip is the same throughout a single cell ; Lapusta et al., 2000) with $G^* = \frac{G}{1-\nu}$. $\Delta\theta_{\max}$ is not allowed to be larger than 0.2 following Lapusta and Liu (2009). In addition, we require a constraint on the time step during the healing phase. In principle, the state increase is linear with time for very low slip velocities such that the time step could be arbitrarily large. However, we justify in section 3.4 that for the initial loading phase it is important to limit the time step as

$$\Delta t_h = 0.2\,\theta. \tag{31}$$

The fraction is chosen according to $\Delta\theta_{\max}$ for the state weakening time step. We further require that the displacement per time step is limited to a fraction of the grid size Δx as

$$\Delta t_d = \Delta d_{\max}\min\left[\left|\frac{\Delta x}{v_x}\right|, \left|\frac{\Delta x}{v_y}\right|\right], \tag{32}$$

where $\Delta d_{\max} = 10^{-3}$ in this study. If the state does not influence friction (i.e., if $b = 0$), the adaptive time step developed by Lapusta et al. (2000) is not sufficient. Here we aim to capture the increasing slip rate

in case of a purely rate-dependent friction by resolving the viscoelastoplastic relaxation time scale $\frac{\eta_{vp}}{G}$ by the fraction f_{max}:

$$\Delta t_{vep} = f_{max} \frac{\eta_{vep}}{G}. \tag{33}$$

with $f_{max} = 0.2$ in this study. This viscoelastoplastic relaxation time scale combines the viscoelastic relaxation time scale $\frac{\eta}{G}$ and the elastoplastic time scale $\frac{\tau_{II}}{G\chi} = \frac{\tau_{II}}{2G\dot{\varepsilon}'_{II(plastic)}}$. Therefore, this also provides a constraint on the time step length if viscous deformation dominates such that the relaxation time is sufficiently well resolved.

Finally, we note that we do not apply a minimum time step cutoff. This is in contrast to Lapusta et al. (2000) and Lapusta and Liu (2009), in which the time step is only adapted until it reaches $\frac{1}{3}\frac{\Delta x}{c_s}$, where c_s is the shear wave speed. We justify this decision in section 4.1.

2.2.5. Solution Cycle

To solve for the conservation of mass and momentum, we use the iterative solution cycle highlighted in green in Figure 2. The treatment of plasticity on the nodes is the prerequisite for conducting iterations with the goal to minimize the error in the yield strength function F.

We use predictor-corrector Picard iterations. Before entering into the iteration cycle, an initial time step length Δt^* is determined using equation (25) based on the solution from the previous time step. At the beginning of each iteration, an estimate of the plastic multiplier (after applying scaling equation (24) transferred to slip rate) is used to determine the viscoplastic viscosity by evaluating the RSF equations (see Text S4 for details on this and the subsequent steps of the iteration cycle). Then, the viscoplastic viscosity is interpolated from basic to center points using either the arithmetic or harmonic average. After the solution is obtained, pressure and stresses are interpolated from center to basic points using the arithmetic average. RSF equations are evaluated in the basic points to obtain a new estimate of the plastic multiplier. This is used for the next iteration and so forth. The error in the yield strength function is defined in each basic point as

$$\Delta F = \left(\frac{\sigma_{yield} - \tau_{II}}{P} \right)^2. \tag{34}$$

This is summed up for all basic nodes and divided by the number of basic nodes to obtain a global error ΔF^* (see equation (S48) for the exact definition of ΔF). Iterations continue until $\Delta F^* <= \Delta F_{min}$. Dividing the difference between the yield strength and square root of second deviatoric stress invariant by pressure provides an error in friction. Consequently, this definition leads to the same accuracy in the determination of friction in models with different pressure. During the iterations we ensure that the requirement for the adaptive time step length is also fulfilled at the end of the time step. For this purpose, we calculate at the end of each iteration the time step length Δt^{**} resulting from the solution of the equations. If $\Delta t^{**} > 1.1\Delta t^*$, iterations are restarted from the initial conditions with the time step length $\Delta t^* = 0.9\Delta t^*$. Furthermore, we check if and how fast the solution converges. If ΔF of the current iteration becomes larger than the one from the previous iteration, iterations are restarted from the initial conditions with a new time step $\Delta t^* = 0.5\Delta t^*$. The same time step correction applies in case of slow convergence such that the number of iterations exceeds 50. The values for the time step correction and the limit in the number of iterations are empirically determined and lead to convergence in all simulations presented in this study, also if the time step factor ζ is not appropriate.

3. Reference Model

This section presents an example of an earthquake cycle simulated with the presented STM-RSF methodology. We begin with the description of the model setup of a mature strik-slip fault zone. This is followed by an overview of the results. Then a detailed analysis of the earthquake cycle and seismic wave propagation is conducted. To validate the implementation of compressibility and the invariant reformulation of RSF, we check the accuracy of our calculations, where analytical approximations and theoretical estimates are available.

3.1. Model Setup

Here we describe the 2-D reference model setup, which consists of a dextral strike-slip plate boundary zone (Figure 3). It is designed to enable comparison to both classical earthquake cycle simulations and theoretical estimates of the nucleation and cohesive zone lengths. Both classical simulation and theoretical estimates are mostly based on the assumption of an infinitely thin fault in a purely elastic medium with constant normal

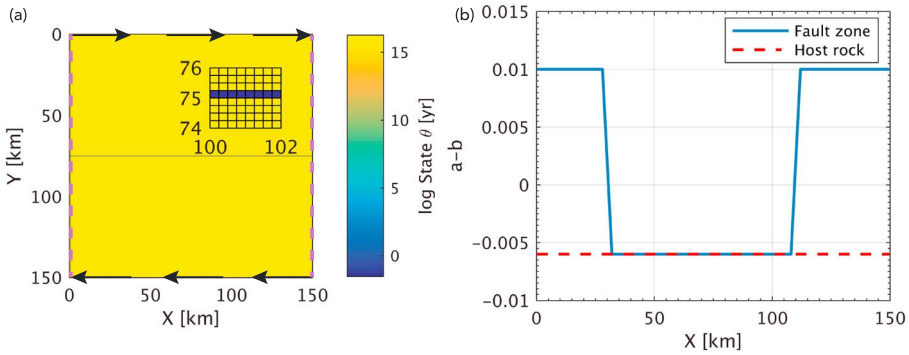

Figure 3. Reference 2-D model setup of a dextral strike-slip fault zone. (a) The fault (top view) is defined as a central zone with a thickness of Δx. In that zone, the initial state is distinctly lower than in the host rock (see inlet, numerical grid is indicated). The fault is loaded by applying Dirichlet boundary conditions for horizontal velocity v_x ($v_x = \pm 2.0 \cdot 10^{-9}$ m/s $= \pm 6.3$ cm/year) at the top and bottom boundary. Neumann boundary conditions for v_x act at the sides (magenta dashed line). Vertical velocity v_y is set to zero at all boundaries. (b) Profile of (a) and (b) along the fault zone and the homogeneous host rock.

stress and zero cohesion. Therefore, (i) the fault is essentially predefined as an initially weak thin layer; (ii) the off-fault medium is elastic with negligible contributions from viscous and plastic deformation; (iii) the value for pressure in the RSF equation is fixed at the initial background pressure P_B and gravity is set to zero to avoid vertical pressure gradients; and (iv) cohesion is set to zero. Parameters for the reference model setup are given in Table 1.

Table 1
Parameters of the Reference Model

Parameter	Symbol	Value
Shear modulus	G	30 GPa
Bulk modulus	K	50 GPa
Poisson ratio	ν	0.25
Density	ρ	2,700 kg/m^3
Shear wave speed	c_s	3.3 km/s
Effective viscosity	η	$5 \cdot 10^{26}$ Pa s
Seismogenic zone width	W_s	76 km
Initial mean stress	P_B	5 MPa
Gravity	g	0 m/s^2
Reference friction	μ_0	0.2
Reference slip velocity	V_0	$4 \cdot 10^{-9}$ m/s
Characteristic slip distance	L	0.01 m
RSF direct effect	a	0.011
RSF evolution effect	b	
in VS region		0.001
in VW region		0.017
Initial state	θ_i	
Host rock		$\frac{L}{V_0}\exp(40)$ s
Fault zone		$\frac{L}{V_0}\exp(-1)$ s
Grid size	$\Delta x = \Delta y$	250 m
Time step factor	ζ	1.0
Error norm	$\Delta F_{\max}$	$2 \cdot 10^{-6}$
Number of marker per grid cell	N_m	16
Amount of random numbers for initial marker distribution	R_n	100

The initial condition is a mature fault, which is defined as a horizontal weak layer with a thickness of Δx in the middle of a strong homogeneous host rock with dimensions of H_x =150 km and H_y =150 km (Figure 3a). The difference in strength is realized by a distinctly lower initial state in the layer than in the host rock. In the weak layer, a steady state rate-weakening zone (i.e., $a-b<0$) with a length $W=76$ km lies in the center (from $X=32$ to $X=108$ km) and rate-strengthening zones (i.e., $a-b>0$) at the sides (Figure 3b). These zones are separated by 4-km wide transition zones. The host rock has the same RSF properties as the rate-weakening zone. Due to the high initial state, plastic deformation is ineffective in the host rock with slip velocities lower than $<10^{-30}$ m/s. The high initial state can be considered to be equivalent to high cohesion. Viscous deformation is essentially absent because the viscoelastic relaxation time is on the order of million of years due to the high viscosity of $5{\cdot}10^{26}$ Pa s and thus much longer than the hundred of years long recurrence times considered in this study. The values for G and K and RSF parameters a, b, and L are within the range adopted in classical earthquake cycle simulations. The relatively low initial background pressure P_B of 5 MPa represents the potential influence of fluids.

The plate boundary zone is loaded by applying velocity boundary conditions $v_x^B=-2\cdot10^{-9}$ m/s and $v_x^T=2\cdot10^{-9}$ m/s at the top and bottom of the model domain, respectively. This imposes a loading strain rate in case of full locking.

$$\dot{\varepsilon}_L=\frac{v_x^B-v_x^T}{2H_y}. \tag{35}$$

In case of stable sliding of fault, the loading conditions lead to a slip rate of $V_l=4.00\cdot10^{-9}$ m/s (12.6 cm/year) along the fault. The reference slip velocity V_0 is set equal to this slip rate such that stress becomes equal to the reference strength $\tau_0=\mu_0 P_B$ when the steady state slip velocity is equal to V_0.

3.2. Initial Loading

The imposed boundary conditions increase stress linearly with time in the rate-strengthening zone at the beginning of the simulation (Figure 4a). The square root of the second deviatoric stress invariant follows the equation for viscoelastic deformation under constant loading strain rate $\dot{\varepsilon}_L$, which can be further simplified to the equation for purely elastic deformation owing to the high viscosity:

$$\tau_L(t)=\left(\tau_L(0)-2\dot{\varepsilon}_L\eta\right)\exp\left(-\frac{G}{\eta}t\right)+2\dot{\varepsilon}_L\eta\approx 2G\dot{\varepsilon}_L t+\tau_L(0). \tag{36}$$

A key property of RSF is that the fault always slips (Figure 4b), and the square root of the second deviatoric stress invariant is always equal to the yield strength (Figure 4a). The elastic loading of the fault causes an exponential increase in slip rate with time in the rate-strengthening zone. This is due to the viscosity-like direct effect of RSF, for which a stress increase is directly transferred into an increase in slip rate. As the square root of the second deviatoric stress invariant approaches the static strength τ_0, slip velocity reaches a noticeable level while approaching the loading rate V_l. At these rates, plastic deformation slows the elastic stress build up down (Figures 4a and 4b).

3.3. Earthquake Cycle

After this initial loading phase, earthquake cycles are generated in form of periodically recurring seismic slip events (Figures 4b and 4c). Slip propagates from the rate-strengthening zones into the central rate-weakening region (Figure 4c). An earthquake nucleates (section 3.3.1) and accelerates during the rupture propagation (section 3.3.2) to a maximum slip velocity of $\sim$ 0.83 m/s during the first earthquake (Figure 4b). This is larger than the seismic threshold of $V_{\text{thres}}=0.012$ m/s, which is defined by Rubin and Ampuero (2005) as

$$V_{\text{thres}}=\frac{2\,a\,Pc_s}{G}. \tag{37}$$

During the dynamic rupture propagation, seismic waves are generated (section 3.3.3). Once the rupture reaches the rate-strengthening regions, slip is decelerated (section 3.3.4) and the rate-weakening fault heals (section 3.3.5). In the initial sequence of the first three earthquakes, rupture properties vary slightly, after which they become more similar, particularly in terms of the hypocenter location (Figure 4c). This initial stabilization period is typical for earthquake cycle simulations and usually excluded in the analysis of the results. Slip reaches a maximum of $\sim$1.6 m in the center of the rate-weakening zone. The release of this amount of slip agrees with the slip deficit that is accumulated at the loading rate V_l during the interseismic period with a duration of $\sim$12.7 years. In the following, we analyze the earthquake cycle for the third event in detail.

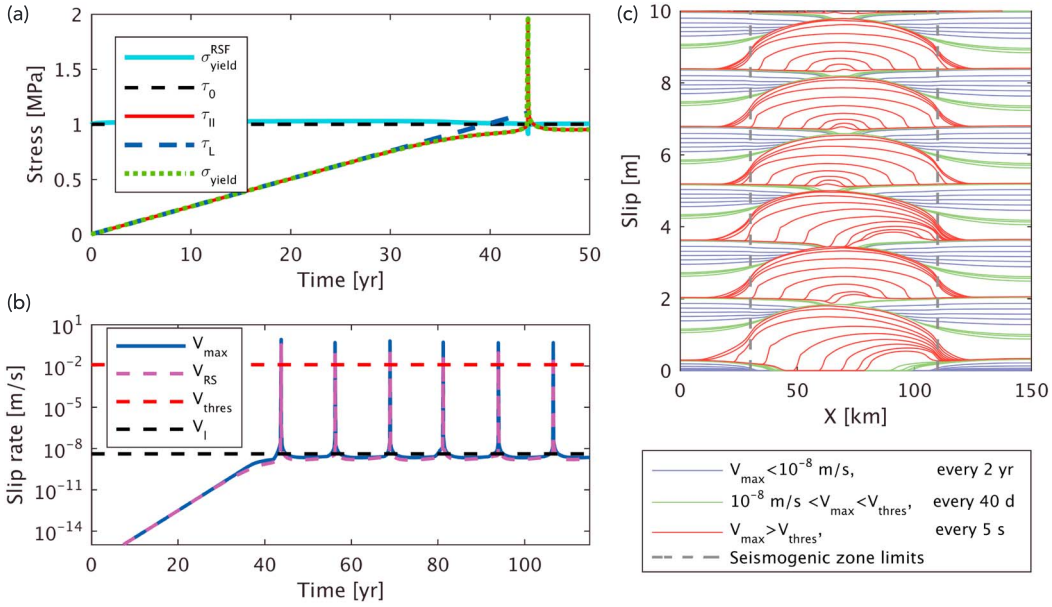

Figure 4. Overview of the results in the reference model. (a) Interseismic loading prior to the first event in the rate-strengthening region at $X = 25$ km. The square root of the second deviatoric stress invariant follows the elastic loading stress $\tau_L \approx 2G\dot{\epsilon}_L t$ and is equal to the yield strength σ_{yield}. Before it reaches the interface strength, the square root of the second deviatoric stress invariant deviates from τ_L. (b) Slip velocity as a function of time. V_{max} is the maximum slip velocity measured along the fault, V_{RS} is the slip velocity at $X = 25$ km, V_l is the loading slip velocity and equal to the reference slip velocity V_0. The treshold slip velocity V_{thres} (equation (37)) is used to identify seismic rates. (c) Slip contours, plotted at regular times, whose interval depends on V_{max} as indicated in the legend.

3.3.1. Nucleation

At the beginning of the earthquake cycle, the rate-strengthening zones slip at steady state rate close to the loading rate (Figures 5a and 5b). This continuous slip produces stress concentrations at the two transitions to the locked rate-weakening zone. These stress concentrations build nucleation fronts, which slowly propagate due to the stable sliding parts of the fault toward the center. This causes an increase in stress and acceleration of slip toward the loading rate in an increasingly larger portion of the rate-weakening zone (Figures 5a and 5b). When the stably slipping patch in the right part reaches a critical size, which is called the nucleation length, it becomes unstable and the rupture nucleates. In the following, we describe the nucleation process in more detail by dividing it into four stages (A–D). In stage A, stress is increased while the leftward propagating nucleation front approaches (Figure 5c). Slip velocity increases with stress, but it remains at a negligible level so that state can increase with time and almost no slip accumulates (Figures 5c and 5d). At the transition between stages A and B, state begins to decrease and slip to accumulate while stress reaches its maximum (Figures 5d and 5e). During stage B, state decreases in interaction with an increasing slip velocity toward the reference slip velocity V_0 (Figure 5d). At the same time, stress drops to its static reference value $\tau_0 = \mu_0 P_\beta = 1$ MPa (Figure 6a). We interpret the length along which this stress drop occurs (Figure 6a) as a quasi-static cohesive zone length, because the rupture speed at this stage goes to zero. The length is around 4.25–4.5 km along both nucleation fronts, while they propagate toward the center (Figures 6a and 6b). This length can be compared to the theoretical estimate of the quasi-static cohesive zone size defined at zero rupture speed v_r (the rupture is barely moving according to Day et al., 2005), which has been derived for linear slip weakening friction (Palmer & Rice, 1973) and was adapted to RSF (Lapusta & Liu, 2009) as

$$\Lambda_0 = \frac{9\pi}{32}\frac{GL}{bP(1-\nu)} \quad \Lambda = \Lambda_0 / f(v_r), \tag{38}$$

where Λ is the dynamic cohesive zone length, which decreases as some function f of rupture speed during the dynamic rupture propagation (Day et al., 2005). Λ_0 equals to 4.16 km for the parameter set in the reference model, which agrees very well with our simulated quasi-static cohesive zone length during the nucleation process.

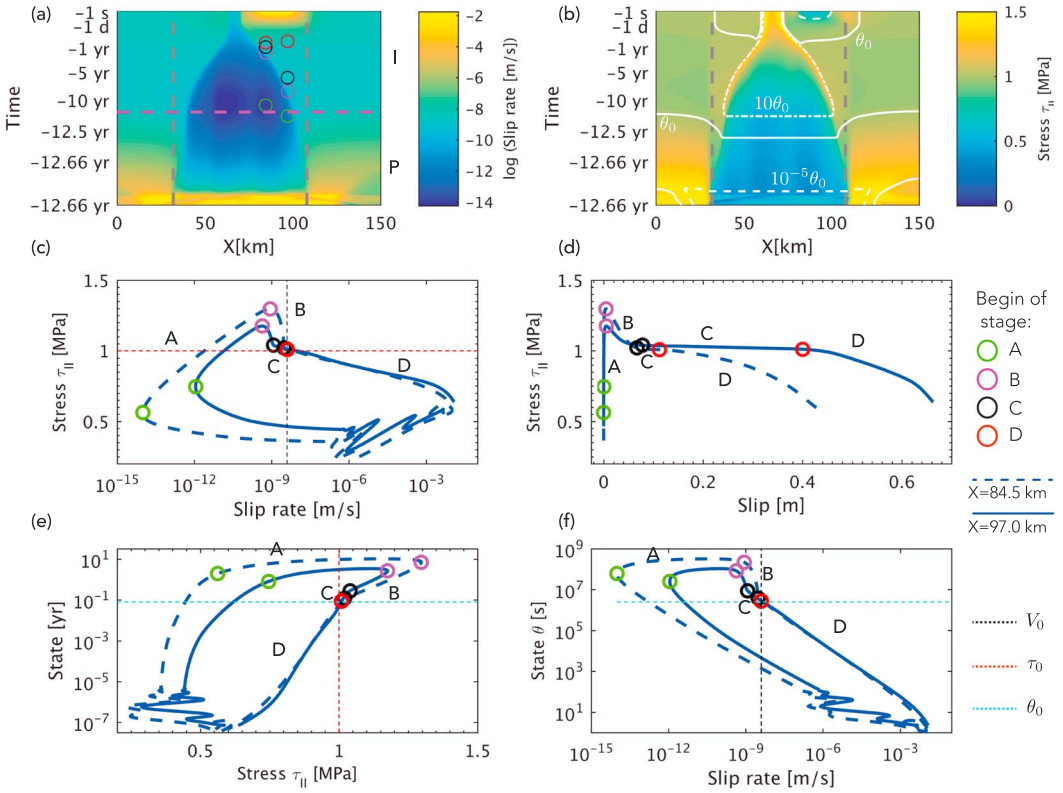

Figure 5. Nucleation process. Spatiotemporal evolution of (a) slip rate V_p and (b) the square root of the second deviatoric stress invariant σ_{II} and state θ (contour lines at three levels with respect to the reference state θ_0; b) during the postseismic deformation after the second event (P) and for the interseismic phase (I) including the nucleation process of the third event (stages A–D) in the reference model. Time zero is defined as the time when slip rate becomes larger than the threshold value V_{thres}. Magenta line in (a) shows the time at which the maximum slip velocity in the rate-strengthening zones drop below the loading rate $V_l = 4.00 \cdot 10^{-9}$ m/s. Differently colored circles represent the begin of stages A–D of the nucleation process in two locations along the rate-weakening zone of the fault. Phase diagrams in these two locations for (c) square root of the second deviatoric stress invariant τ_{II} versus slip rate V_p, (d) square root of the second deviatoric stress invariant τ_{II} versus slip, (e) state θ versus square root of the second deviatoric stress invariant τ_{II} (e) and state θ versus slip rate V_p (f). Reference values for slip velocity, square root of the second deviatoric stress invariant, and state are shown.

During stage C, after the nucleation front has passed, square root of the second deviatoric stress invariant and state are close to, but still larger than, their reference values τ_0 and θ_0, respectively (Figures 5c–5f). Before the slipping area behind the rupture front reaches the size of the nucleation length, slip is accumulated in a stable way (Figure 5d), while square root of the second deviatoric stress invariant and state decrease and slip rate increase toward their reference values (stage C is hardly visible in the phase diagram of square root of the second deviatoric stress invariant, slip velocity, and state, Figures 5c–5f).

The transitions between the nucleation stages A–C are shifted in time (or slip) and/or in phase space (Figures 5c–5f) as the nucleation front propagates and the area of stable slip becomes larger (Figures 5a and 5b): For those points further to center, (i) slip velocity is lower and state increases to higher values at the beginning of stage A; (ii) the peak stress at the end of stage A is higher at larger slip velocities and higher state; (iii) state, stress, and slip velocity are closer to their reference values at the end of stage B; and (iv) the time interval of the stable sliding and, hence, the amount of stable slip are reduced during stage C. Consequently, the quasi-static stress drop becomes larger, the fault slips faster, and the propagation speed of the nucleation front increases. Also, the time that is necessary to reduce the state from $10\theta_0$ to θ_0 becomes shorter (Figure 5b). At some point, the beginning and end of stage C coalesce into the same point in time (Figure 5a) and phase space (Figure 5c). This means that stable sliding behind the rupture front is not possible anymore. This is transferred from the nucleation front to the rest of the so far stably sliding patch behind the nucleation front. This is the onset of instability during stage D, which is driven by the feedback between decreasing state and increasing slip velocities (Figure 5). Due to the slightly asymmetric model setup, the nucleation front

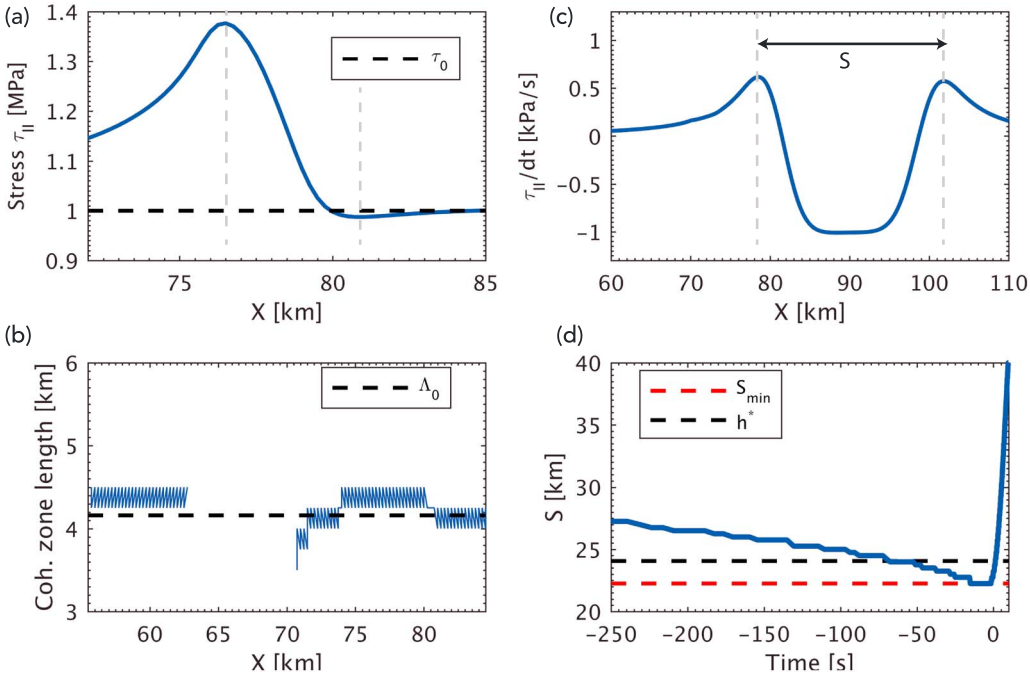

Figure 6. Quasi-static cohesive zone, and nucleation zone. (a) Quasi-static cohesive zone along the leftward propagating nucleation front when the peak of the square root of the second deviatoric stress invariant is located at $X = 76.5$ km. The length of this zone is defined as the distance between the local maximum and minimum stress as indicated. (b) Measured quasi-static cohesive zone length as a function of X for both nucleation fronts starting from approximately 1.7 years before slip velocity reaches the coseismic threshold V_{thres}. Λ_0 is the theoretical estimate of this length (Lapusta & Liu, 2009). (c) Time derivative of the square root of the second deviatoric stress invariant. S denotes the distance between the peaks. (d) Evolution of the interpeak distance S (c) during the last part of the nucleation process, during which S first decreases until it reaches the minimum. This minimum can be compared against the theoretical nucleation size h^* (Rubin & Ampuero, 2005), as discussed in the main text.

from the right side propagates faster than the one from the left side (Figures 5a and 5b). Therefore, the right stable sliding patch becomes unstable first and, hence, the region where the earthquake nucleates.

The evolving instability is characterized by a stress drop in the central portion of this nucleation region, whereby the surroundings are dynamically loaded (Figures 5a and 5b). In other words, the stress rate profile in the nucleation area consists of a minimum in between two peaks. The distance S between the stress rate peaks shrinks with time toward $S_{min} = 22.25$ km. Two dynamic rupture fronts are created in this process, which rapidly propagate outward during the subsequent dynamic rupture propagation such that S increases after reaching S_{min} (Figure 6d). This shrinking of the rapidly accelerating slip zone during nucleation of an earthquake has been noted before in RSF simulations by Dieterich (1992) and Rubin and Ampuero (2005). Rubin and Ampuero (2005) took the minimum S_{min} as a measure to determine the nucleation length. They derived a theoretical estimate of the nucleation size h^* on the basis of energy balance for a quasi-statically expanding crack, which fits their simulations for $0.5 > a/b > 1$:

$$h^* = \frac{2}{\pi} \frac{GbL}{(b-a)^2 P(1-\nu)} \tag{39}$$

Using the parameters in our simulation ($a/b \sim 0.65$), $h^* = 24.1$ km, which is close to the simulated S_{min}. For earthquakes 4–6, nucleation occurs in the center of the rate-weakening zone (Figure 4c). Since it is triggered by the two inward propagating nucleation fronts at the same time in these cases, the nucleation size is approximately half as big.

3.3.2. Dynamic Rupture Propagation

After the nucleation size is reached, the third event in the reference model begins to propagate dynamically (Figures 7a and 7b). It propagates bilaterally with a longer propagation distance to the left due to the hypocenter location being closer to the right rate-strengthening zone. At the rupture front within the cohesive zone,

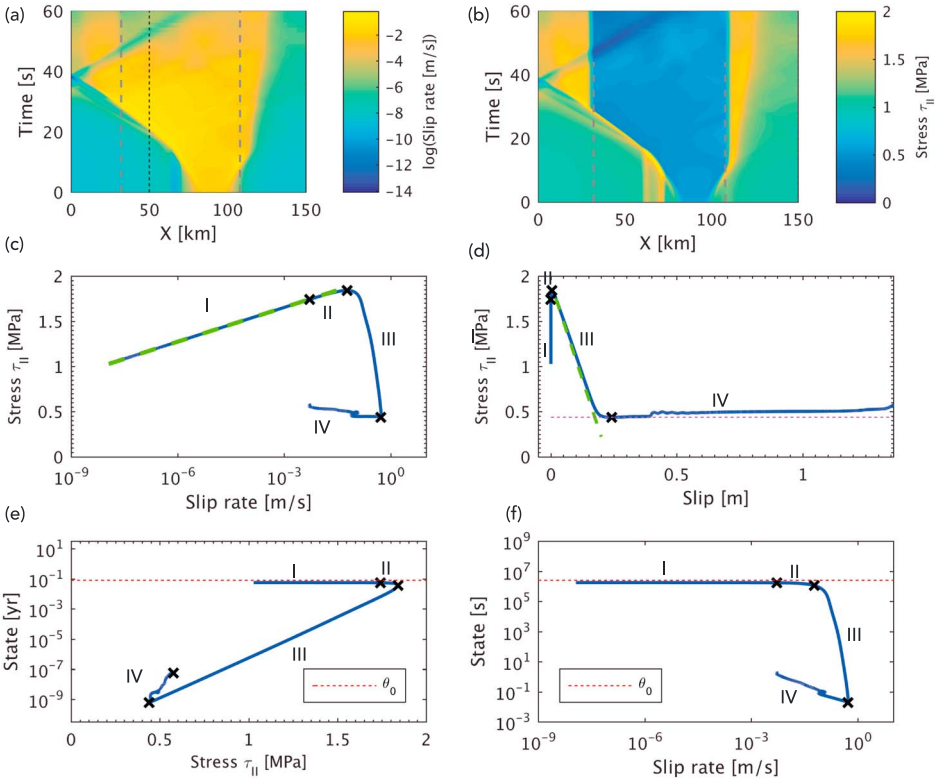

Figure 7. Dynamic rupture propagation. Spatiotemporal evolution of slip rate V_p (a) and the square root of the second deviatoric stress invariant σ_{II} (b) during the third earthquake in the simulation example. Time zero is defined as the time when maximum slip rate reaches the coseismic threshold. I–VI indicate the stages discussed in detail in the main text. Gray dashed lines indicate the transition zone from the central rate-weakening to the outer rate-strengthening zones (Figure 3 b). Phase diagrams of (a) the square root of the second deviatoric stress invariant τ_{II} versus slip rate V_p, (b) square root of the second deviatoric stress invariant τ_{II} versus slip, (c) state θ versus square root of the second deviatoric stress invariant τ_{II} and (d) state θ versus slip rate V_p in fault location $X = 50$ km from the end of the second event until the end of the third event. I–VI along the black solid line indicate stages during dynamic rupture propagation (Figures 7 and 7b). Green dashed lines represent analytical expressions of parts of the seismic cycle as discussed in the main text: The direct effect (c) and linear slip weakening (d).

the square root of the second deviatoric stress invariant drops from the maximum to the dynamic level (Figure 8a), which is often described as the breakdown process.

To illustrate the breakdown process at the rupture front in our simulations, we describe it in detail for location $X = 50$ km (Figures 7c–7f). Similarly to Bizzarri and Cocco (2003), we subdivide this process into four distinct stages (I–IV). In stage I, during which the rupture front approaches (Figures 7a and 7b), the square root of the second deviatoric stress invariant increases due to dynamic elastic loading from stress drops along parts of the fault that already slip at seismic slip rates (Figures 7a–7c). While state decreases only very slightly from its steady state value θ^*_{ss}, slip velocity increases exponentially with the square root of the second deviatoric stress invariant (Figure 7c). This simulated rate-strengthening process is due to the "viscosity-like" direct effect of RSF, because an increase in stress results directly in an increase in slip rate. It agrees with equation (1) expressed as $V = f(\tau_{II})$ (green dashed line in Figure 7c)

$$V_p = V_0 \exp\left(\frac{\tau_{II}}{aP} + \frac{-\left(\mu_0 + b \ln\left(\frac{\theta^*_{ss} V_0}{L}\right)\right)}{a}\right), \tag{40}$$

where the second term in the exponent consists of only constants assuming that θ remains at θ^*_{ss}.

Stage II is characterized by the onset of noticeable slip (Figure 7d) and decreasing state (Figure 7e), which leads to a deviation from the direct effect (Figure 7c). At the end of this stage, the square root of the second deviatoric stress invariant reaches its peak value. In the weakening stage III during the passage of the rupture

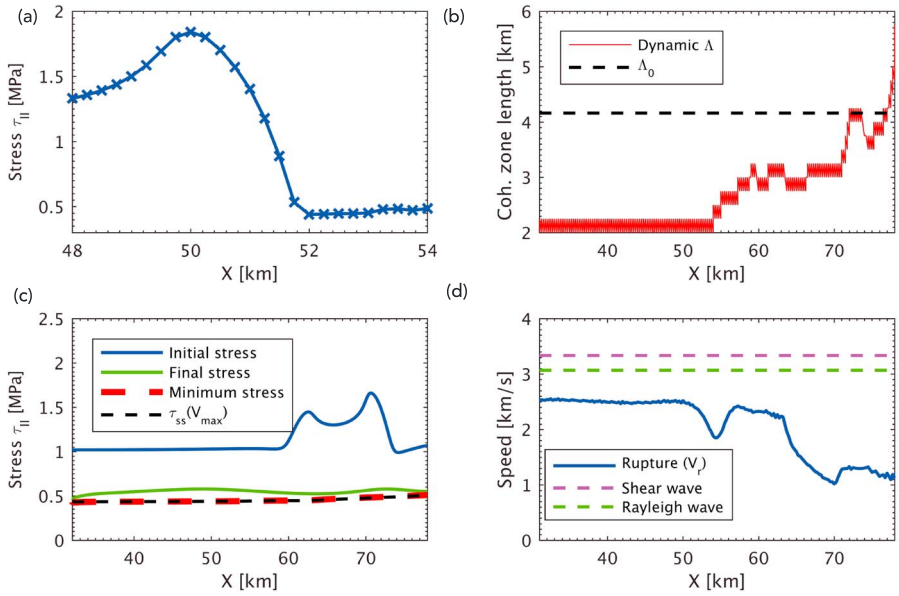

Figure 8. Dynamic cohesive zone size, stress profile, and rupture speed for the leftward propagating rupture front. (a) Zoom into the stress profile at the time when the rupture front arrives at $X = 50$ km. Crosses indicate basic nodal points, indicating a sufficient resolution of the minimum cohesive zone length using a grid size of 250 m. (b) Cohesive zone length as a function of the rupture path. Λ_0 is the quasi-static cohesive zone length. Note that the cohesive zone length is larger than Λ_0 close to the nucleation region but quickly decreases below Λ_0. (c) Profile of the square root of the second deviatoric stress invariant at the beginning and end of the rupture propagation. The minimum of the square root of the second deviatoric stress invariant agrees with the steady state strength calculated using the maximum slip velocity reached in each point. (d) Speed of the rupture front. For the calculation of the rupture speed, we define in each point along the rupture propagation the arrival time of the rupture front when slip rate becomes larger than V_{thres}. Shear and Rayleigh wave speeds are shown for comparison.

front, state drops by many orders of magnitude (Figure 7e) and the square root of the second deviatoric stress invariant decreases from the peak to the dynamic value linearly with slip d (Figure 7d). This simulated linear slip weakening process follows the corresponding approximation given by Lapusta and Liu (2009; green dashed line in Figure 7d) as

$$\tau_{II}(t - t0) = \frac{bP}{L}\delta(t) + \delta(t_0). \tag{41}$$

As in Bizzarri and Cocco (2003), the peak slip velocity occurs in between the peak and dynamic stress levels (Figure 7e). More precisely, it occurs very close to the end of slip weakening process when the slip weakening rates are reduced toward the dynamic stress level. In stage IV, slip continues at the dynamic stress level (Figure 7d). The fault restrengthens with decreasing slip rate and increasing state (Figure 7f).

The cohesive zone length shrinks from Λ_0 to 2 km during rupture propagation (Figure 8c), while rupture speed increases from less than 1 to 2.5 km/s (Figure 8d). Variations of the rupture speed as well as the final rupture speed can be related to variable prestress conditions (Figure 8b). The final rupture speed is ~0.81 times the Rayleigh and ~0.75 times the shear wave speed. As soon as the rupture reaches the rate-strengthening zones, it gets decelerated with rupture speed and slip rate dropping down. Since the rupture gained more energy on its longer propagation path to the left, it is able to propagate a longer distance into the rate-strengthening zone.

3.3.3. Seismic Waves

Variations in rupture speed during the dynamic rupture propagation cause the generation of shear and pressure waves. Their radiation pattern is in agreement with a dextral strike-slip event (Figures 9a and 9b). Shear waves produce maximum velocity changes perpendicular and parallel to the fault, while those of pressure waves occur at an angle of around 45° to the fault. The measured propagation speeds of these shear and pressure waves agree with theoretical estimates (Figure S8). Different waveforms can be identified (Figure 8d), which are described in relation to the dynamic rupture propagation in Text S5. Those waveforms, which propagate along the fault, increase slip rate and the square root of the second deviatoric stress invariant in front

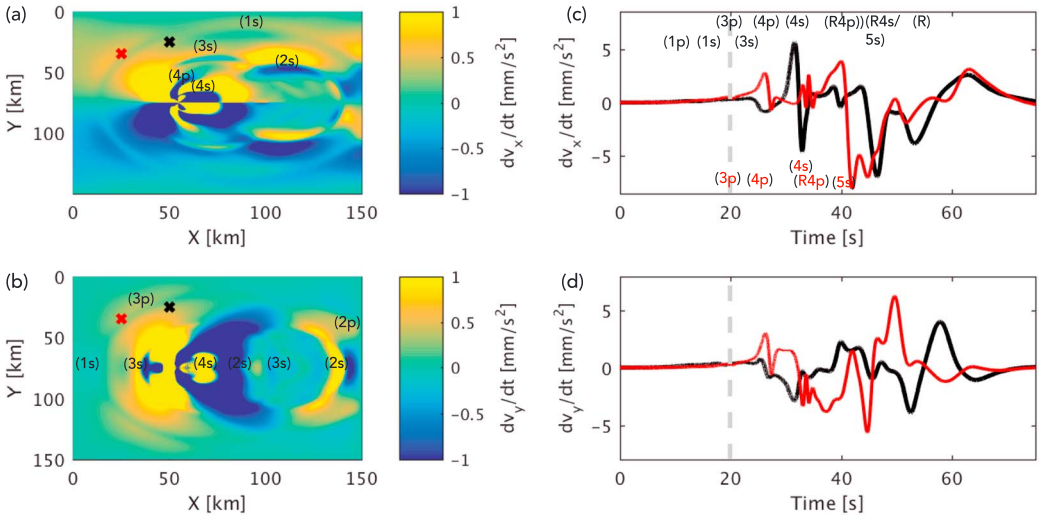

Figure 9. Seismic waves. Snapshots of the time derivates of (a) horizontal and (b) vertical velocity during the third event of the simulation example. Letters denote different waveforms, which are described in Text S5. Time series of time derivatives of (c) horizontal and (d) vertical velocities at two locations as indicated with the correspondingly colored crosses in (a) and (b).

of the rupture (Figures 7a and 7b). Figures 9c and 9d show the arrival of these waveforms at two stations off the fault. Toward the end of the rupture, the generated waves reach the boundary, at which they are reflected back toward the center due to the absence of absorbing boundary conditions (Figures 7a and 7b). They reach the rate-weakening zone again and lead to variations in stress (Figure 7b). These wave reflections are noticeable at the end of the station recordings (indicated by R in Figures 9c and 9d).

3.3.4. Postseismic Relaxation

The propagation of a into the rate-strengthening region leads to an increase in slip rate and the square root of the second deviatoric stress invariant (Figures 5a and 5b). During the postseismic period, the elevated stress and slip velocity decrease over time. This postseismic relaxation process at fault location $X = 25$ km follows the analytical approximation derived by Perfettini and Ampuero (2008; Figure 10a):

$$V_a(t-t_0) = \frac{V_a(t_0)\mathrm{e}^{\frac{t}{t_r}}}{1+\frac{V_a(t_0)}{V_l}\left(\mathrm{e}^{\frac{t}{t_r}}-1\right)}, \quad t_r = \frac{(a-b)P}{KV_l}, \quad K = \frac{2}{\pi}\frac{G^*}{W_s} \tag{42}$$

with $W_s = 32$ km is the length of the left rate-strengthening zone. Time t_0 is taken after slip rate reaches its maximum value. Note that we take V_l, which Perfettini and Ampuero (2008) define as the loading slip velocity, to be the minimum slip velocity to which V_p relaxes at the end of the postseismic phase ($V_l = 1.54 \cdot 10^{-9}$ m/s = 4.9 cm/year).

3.3.5. Interseismic Healing

In the rate-weakening zone, slip rate drops down to less than one thousands of the loading rate after the earthquake (Figure 5 a). At these very low slip velocities, the state increases linearly with time (Figure 10b). This simulated linear healing follows the approximation of the aging state evolution equation (2), which is integrated assuming negligible slip rates (Figure 10b):

$$\theta = t \tag{43}$$

Deviations from the linear trend arise due to the arrival of the nucleation front at the end of stage B.

3.4. Time Step Length and Picard Iterations

The adaptive time step used in this study is mainly based on the work by Lapusta et al. (2000). Additionally, we introduced further constraints on the time step length, whose role during the simulation of earthquake cycles is investigated in the following.

During the initial loading phase described in section 3.2, the adaptive time step increases with time to a maximum of ~4 years, which is controlled by the healing time step Δt_h (Figure 11a). The weakening time step Δt_w and viscoelastoplastic time step Δt_{vep} are several orders of magnitude higher and reduce with time

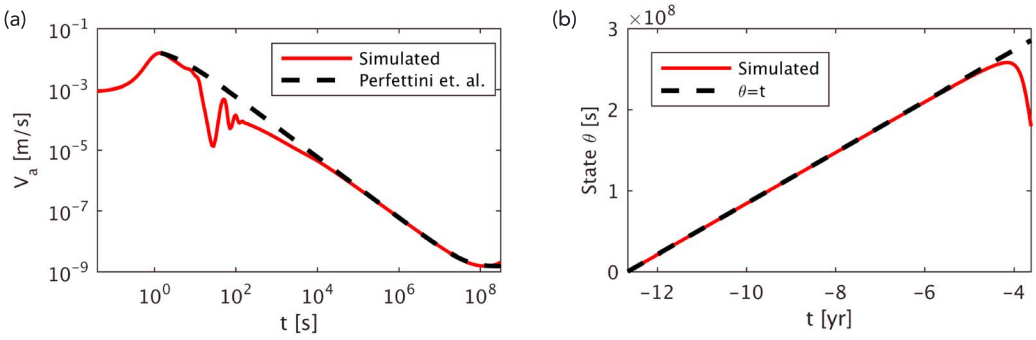

Figure 10. Postseismic relaxation and interseismic healing. (a) Comparison between postseismic relaxation in the rate-strengthening zone (at $X = 25$ km) following the second rupture with analytical expression for postseismic relaxation given in Perfettini and Ampuero (2008). Note that the beginning of the postseismic process is interfered by those seismic waves, which are generated by the second event and reflected from the model boundaries. (b) Comparison of simulated (following the second rupture) and analytical healing at fault location $X = 50$ km during the interseismic period.

and increasing slip rates until they become smaller than the healing time step. Around that time the contributions from healing and weakening of the state become equal and the number of Picard iterations and time step corrections increase (Figures 11c and 11d). The introduction of the healing time step and the displacement time step (if the initial loading phase is longer) ensure to capture the time after which the time step length decreases. Test runs without the healing time step results in jumping over that time, which considerably increases the number of Picard iterations and time step corrections during the initial loading phase.

During earthquake cycles, the time step length varies from ~3 days during the interseismic period to ~0.6 ms during the coseismic period (Figures 11a and 11b). The adapting time step is mainly determined by the healing

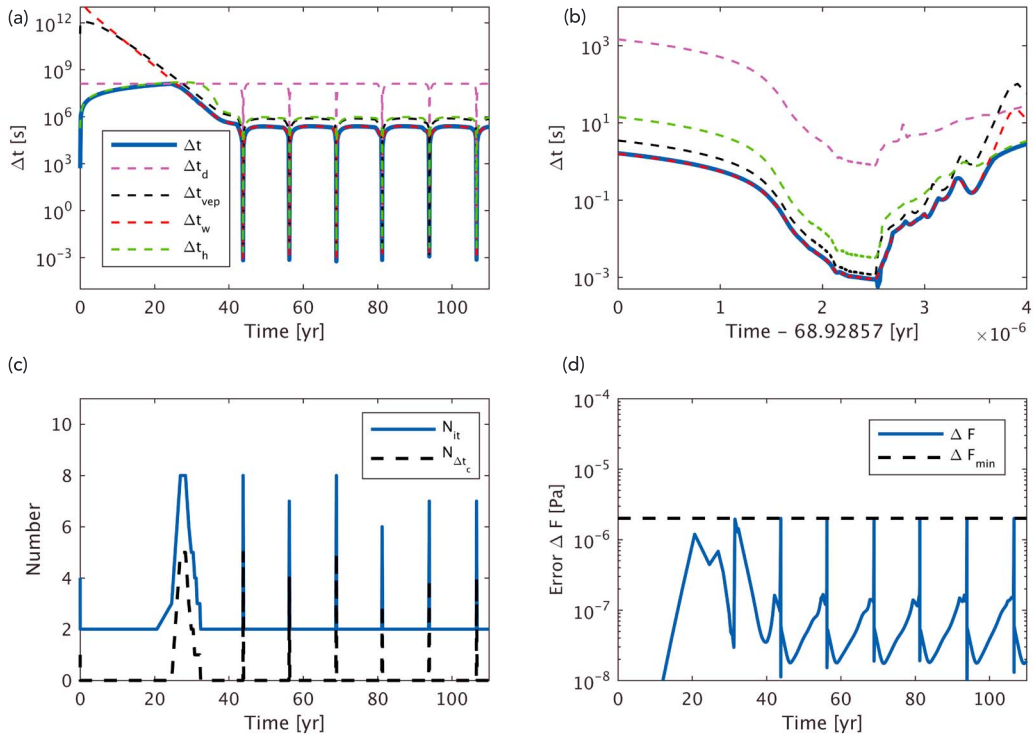

Figure 11. Overview of numerical parameters. (a) Adapted computational time step length Δt, displacement time step Δt_d, relaxation time step Δt_{vep}, weakening Δt_w and healing time step Δt_h. (b) Zoom in (a) during the third event. (c) Number of Picard iterations N_{it} and time step corrections $N_{\Delta t_c}$ during Picard iterations. (d) Error ΔF in determining the yield function F at the last global iteration is reduced below the error limit ΔF_{min}.

Table 2
Numerical Input Parameters for Convergence Analysis

			Parameter values		
Convergence w.r.t.	Δx (m)	ζ	$\Delta F_{min} \cdot 10^6$	N_m	R_n
1. Grid size[a]	187.5, 250, 500, 1,000	1	2	16	100
2. Time step factor	500	0.25, 0.5, 1, 2	2	16	100
3. Error limit	500	1	0.5, 1, 2	16	100
4. Number of markers per grid cell	500	1	2	16, 64, 256	100
5. Randomness of initial marker distribution	500	1	2	16	100, 200, 400

[a] Approximate computational run times (on Swiss National Supercomputing Centre cluster Mönch, with 14 open MP threads) from the beginning until the end of the first simulated earthquake: ~400 hr at $\Delta x = 187.5$ m, ~120 hr at $\Delta x = 250$ m, ~10 hr at $\Delta x = 500$ m, and ~1.5 hr at $\Delta x = 1{,}000$ m. Note that with increasing spatial resolution, the computational time step requirement becomes finer (equations (28)–(30)). A grid resolution of 125 m leads to unfeasible computational times.

time step Δt_w (Figure 11b). The larger viscoelastoplastic time step Δt_{vep} closely follows Δt_w, which indicates a potential similarity of these two time step requirements. A few Picard iterations are sufficient to converge below the error limit ΔF_{min} (Figures 11c and 11d). During the coseismic period, more iterations and more time step corrections are necessary than during the interseismic period.

4. Convergence Analysis

We investigate the numerical stability of our STM-RSF approach by analyzing the convergence of the results with respect to the following numerical input parameters, which we have identified as the most important for our method: grid size Δx, time step factor ζ (equation (25)), error limit of the Picard iterations ΔF_{min} (equation (34)), amount of markers N_m per grid cell, and degree of randomness R_n in the initial marker distribution (Table 2). To evaluate convergence with respect to each parameter, we use the rupture arrival time (see Table 3 for exact definition) as a primary parameter for comparison, because it is a very sensitive parameter for numerical precision (Day et al., 2005). Other coseismic and interseismic parameters are analyzed in Text S5. For Δx, we adapt the thickness of the zone with a low initial state to Δx to ensure that the deformation is localized within one grid size along the entire fault, including the rate-strengthening zones. To assess the role of the internodal interpolation of viscoplastic viscosity (section 2.2.3), we conducted the convergence analysis for arithmetic and harmonic averaging.

4.1. First Earthquake

The grid size is expected to have a large impact on the results, because it needs to be small enough to resolve the cohesive zone length during the entire rupture simulation (Day et al., 2005). Figure 12a shows the influence of the grid size and the internodal interpolation of viscoplastic viscosity on the time evolution of slip rate at $X = 50$ km during the passage of the rupture front of the first earthquake. The difference in shape and timing of the slip rate function becomes smaller with decreasing grid size. Oscillations behind the rupture front distinctly decrease with increasing spatial resolution. This indicates that these oscillations have a numerical rather than a physical origin. These oscillations have also been observed in classical earthquake cycle simulations and attributed to insufficient spatial resolution of the cohesive zone (Lapusta & Liu, 2009). The convergence with respect to grid size is faster for arithmetic than for harmonic interpolation. This difference in convergence rate is shown for the rupture arrival time, which for harmonic interpolation decreases linearly

Table 3
Definition of Output Parameters

Symbol	Short description	Definition
T_a	Rupture arrival time	Time after nucleation at which the rupture front arrives at $X = 50$ km
X_h	Hypocenter location	The central point in the nucleation region defined by S_{min}
$\bar{d}$	Average slip	Average slip inside the rate-weakening zone during an earthquake
T_i	Interseismic duration	Time interval between two earthquakes for $\max[V] < V_{thres}$

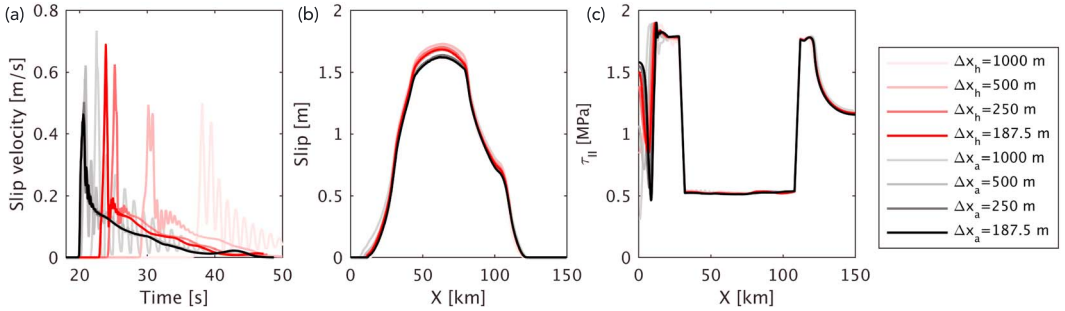

Figure 12. Influence of grid size. Influence of Δx on the solution of the first earthquake in the simulation example: (a) Time evolution of slip velocity V_p in $X = 50$ km, profiles of (b) slip and (c) the square root of the second deviatoric stress invariant when V_{max} becomes smaller than 0.05 m/s at the end of the first earthquake. For each grid size, results are shown for arithmetic (Δx_a) and harmonic (Δx_h) internodal interpolation of the viscoplastic viscosity.

with grid size and for arithmetic interpolation asymptotically approaches around 20 s (Figure 13a). The difference between results from models with harmonic and arithmetic averaging becomes smaller with decreasing grid size, indicating that, at fine grid size, choice of internodal interpolation becomes less important. A similar convergence with respect to grid size is obtained in terms of interseismic duration, while nucleation length, cohesive zone length, and hypocenter location of the first earthquake are essentially not affected by the grid size or intergrid interpolation (Text S5).

The impact of the time step factor ζ and the error limit for the Picard iterations ΔF on the rupture properties of the first earthquake is small (less than 1%) compared to the role of Δx and the interpolation of the plastic multiplier (Figures 13b and 13c). A smaller time step factor than 1 leads to larger computational costs due to the larger amount of time steps. Larger time step factors than 1 leads to a higher number of Picard iterations and time step corrections. The time step is corrected to ensure convergence with Picard iterations below the error limit. Hence, a similar, but computationally more expensive solution is obtained. Similarly, defining a minimum time step as in Lapusta and Liu (2009) leads to a higher number of Picard iterations and time step length corrections per time step during the phase of the highest slip rates. This indicates that defining a minimum time step, which is higher than the one required by equation (25), causes problems in the accurate determination of the yield strength in that phase.

A 16 times higher number of markers per grid cell leads to essentially the same solution as in the reference model in terms of the arrival time (Figure 13d) and other rupture properties (Text S5). Also, different initial marker distributions, which are determined by R_n, do not affect the numerical results (Figure 13e).

At the end of the first earthquake (i.e., V_{max} has dropped below 0.05 m/s), the profiles of slip and the square root of the second deviatoric stress invariant along the fault show the same characteristic pattern for different grid sizes, the two internodal interpolation schemes (Figure 12b) and for different time step factors (Figure S14). The smooth and overlapping profiles of the square root of the second deviatoric stress invariant at the end of the rupture (Figure 12c) are modified by reflected waves (Figures S14 and 15). This modification shows a high sensitivity to grid size, time step factor, and internodal interpolation. Consequently, the stress profiles in the following interseismic period are affected (Figures S14 and 15), which is taken into account in the convergence analysis for the subsequent earthquakes.

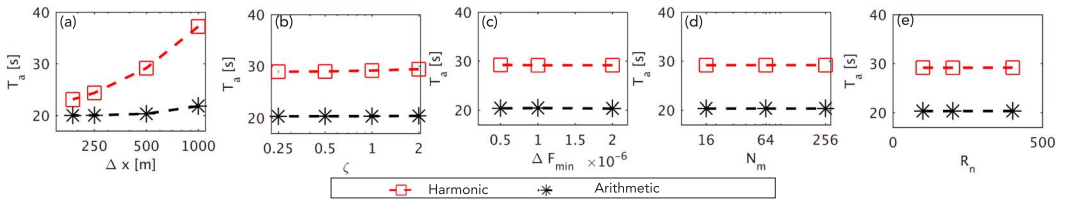

Figure 13. Convergence of rupture arrival time with respect to several numerical parameters. (a and b) grid size Δx, (c and d) time step adaptivity factor ζ, (e and f) error limit ΔF, (g and h) amount of markers N_m, and the marker distribution parameter R_n for the rupture arrival time T_a at $X = 50$ km for the first earthquake in the simulation example. Results are shown for arithmetic and harmonic averaging of the viscoplastic viscosity.

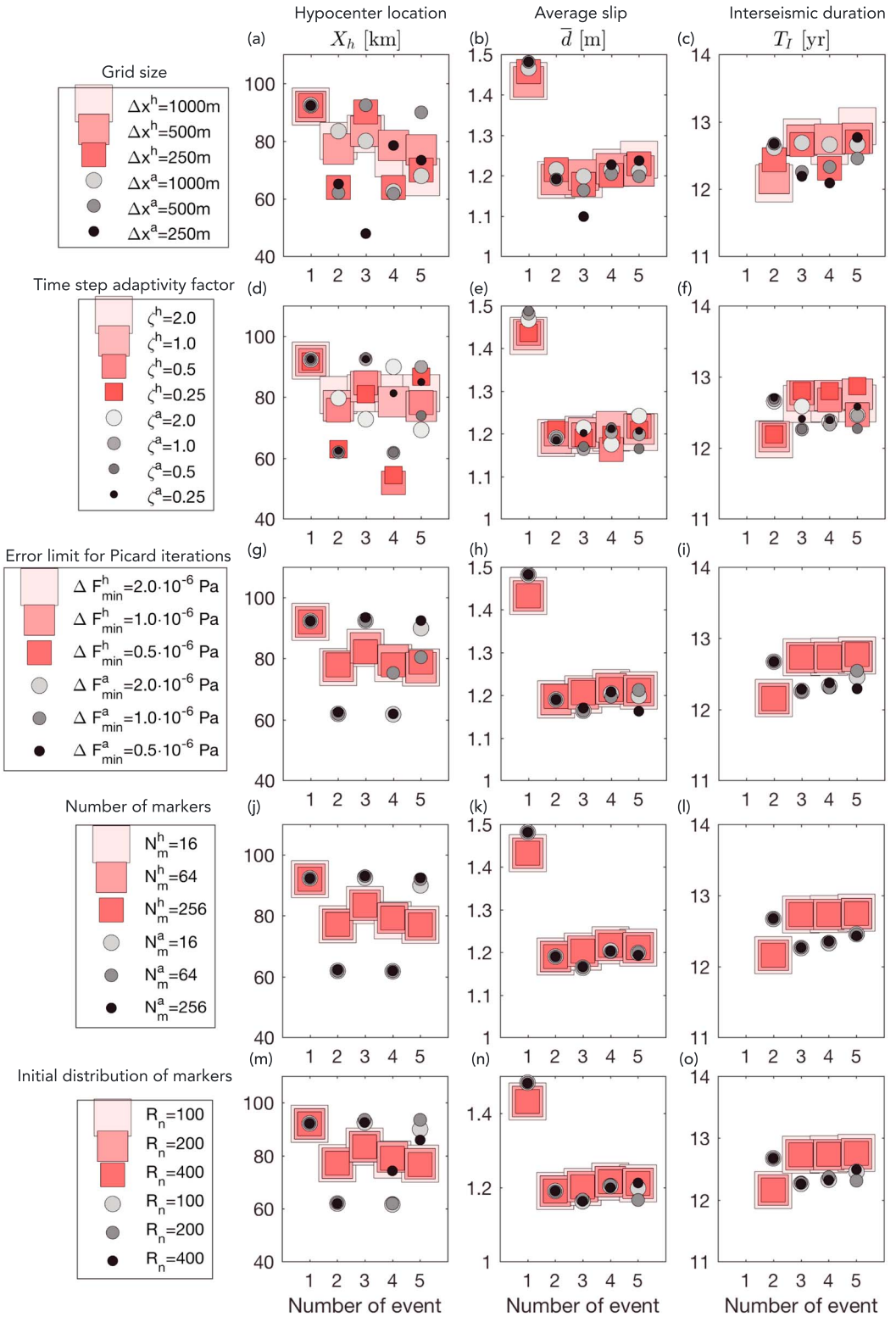

Figure 14. Convergence of earthquake cycle results with respect to investigated numerical parameters. Role of (a–c) grid size Δx, (d–f) time step adaptivity factor Λ, (g–i) error limit ΔF, (j–l) amount of markers N_m, and randomness of the initial distribution R_n for hypocenter location (left), average slip (middle) and interseismic duration (right) for the first five earthquakes. Subscripts *a* and *h* denote models run with arithmetic and harmonic averaging of the viscoplastic viscosity.

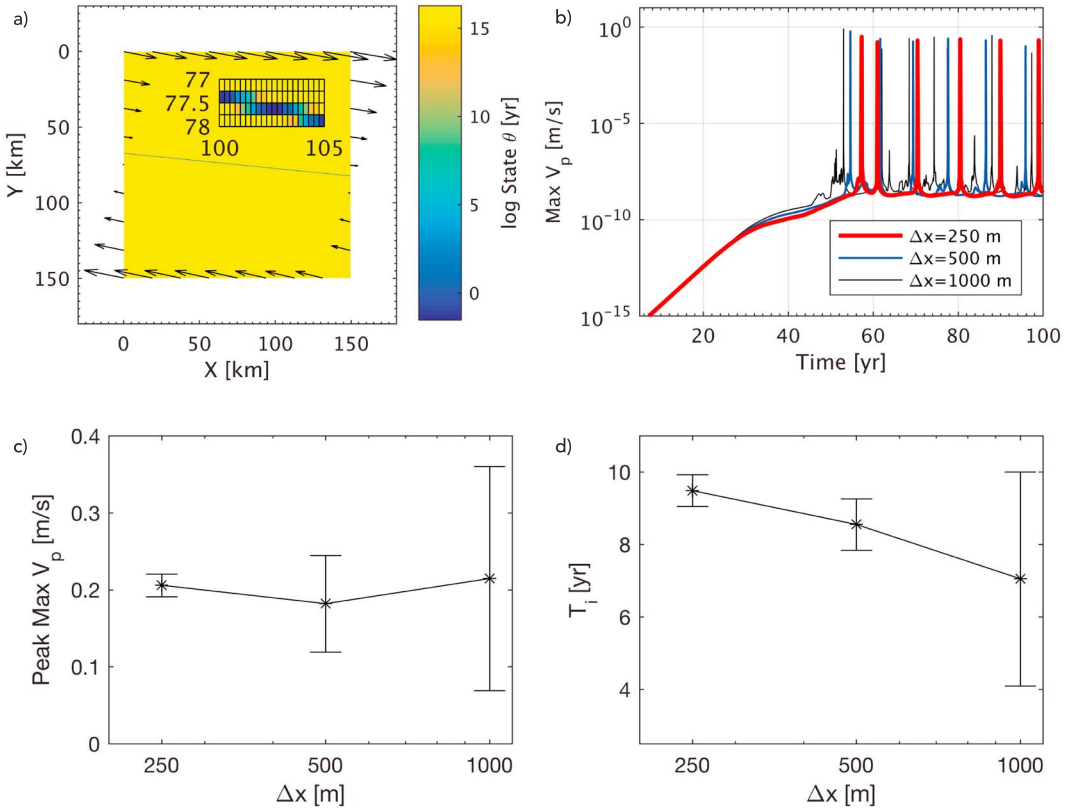

Figure 15. Influence of grid size in a model with a fault, which crosses the numerical grid at an angle. (a) Model setup with velocity boundary conditions as indicated. The fault is defined as a zone with lower initial state than in the surroundings, similarly to the reference model (Figure 3), however, at an angle of 5.71°. Inlet zooms into a small portion of the fault. (b) Time series of maximum slip velocity in the models with Δx = 250, 500, and 1,000 m. Average and standard deviation of peak velocity (c) and interseismic duration (d), calculated from the third to sixth earthquake in each model.

4.2. Sequence of Earthquakes

To evaluate the impact of the numerical parameters on the earthquake cycle results, we compare the first five earthquakes in terms of hypocenter location, average slip and interseismic duration (Table 3). The overall sequence of quasi-periodic earthquakes is not affected by the grid size, time step factor, and internodal interpolation of viscoplastic viscosity (Figures 14a–14f). However, some variations due to changes in these numerical parameters are noticeable starting from the second earthquake, especially in terms of the hypocenter location. These variations can be explained by the modifications of the square root of the second deviatoric stress invariant inside the seismogenic zone from reflected waves after each earthquake. These stress modifications show the highest sensitivity to grid size, time step factor, and internodal interpolation after the first earthquake (Figures S14 and S15). It is expected that the initial stress condition affects the hypocenter location, particularly in our nearly symmetrical model setup.

In addition, the error limit for the Picard iterations, the amount of markers, and the randomness of the initial marker distribution do not play any significant role for the sequence of earthquakes (Figures 14g–14o).

4.3. Grid Convergence Along an Oblique Fault

So far, results are analyzed along a straight, horizontal fault, which is optimally located with respect to the regular numerical grid. In the targeted simulations of our numerical approach, faults may cross the numerical grid at a certain angle. Therefore, we tested the convergence with respect to grid size in a model setup, in which the fault zone with a thickness of one grid size crosses the grid at a small angle (Figure 15 a). Numerical oscillations in slip velocity decrease with increasing spatial resolution (Figure 15b). Furthermore, we calculated the average and standard deviation of the earthquake peak slip velocities and of the interseismic duration (Figures 15c and 15d). The increase of spatial resolution leads to a decrease of the standard deviation, while the average values stabilize, indicating that a convergent solution is approached. The numerical oscillations

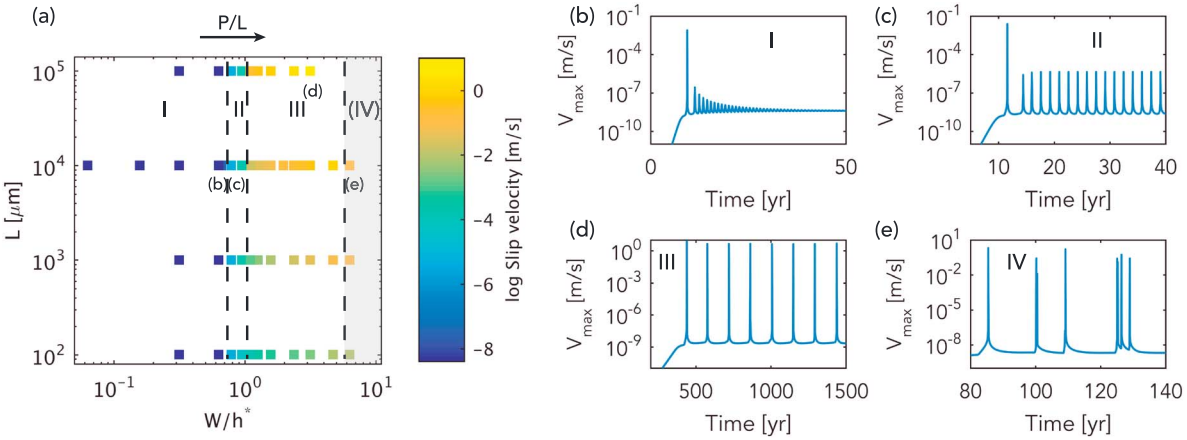

Figure 16. Slip spectrum as a function of nucleation size. (a) Slip spectrum is expressed as slip velocity, as a function of W/h^*, L, and P/L, respectively. Four different regimes are identified: Decaying oscillations toward stable sliding (I), periodic aseismic slip transients (II), periodic (III), and aperiodic seismic slip (IV). Regime IV is characterized by a more complex slip pattern. Since reflected waves from the model boundaries influence this pattern, regime IV is shaded. Panels b–e, which show the maximum slip velocity in the model, provide examples for the different regimes in the slip spectrum: Decaying oscillations (b), period aseismic slip transients (c), period seismic slip (d), and aperiodic seismic slip (e).

at a given spatial resolution are higher than in the respective models with a horizontal fault. This indicates that the numerical solution deteriorates by increasing the fault angle and that, at a given fault angle, a higher spatial resolution is necessary to reach a convergent solution.

5. Slip Response Spectrum

The goal of this section is to test the capability of our STM-RSF approach to resolve different regimes of the slip spectrum from stable sliding, aseismic slip transients to seismic slip. Furthermore, we investigate the role of compressibility and dynamic pressure.

5.1. Role of the Nucleation Size

Based on the reference model, we conduct a parameter study motivated by the work of Liu and Rice (2007). By using classical earthquake cycle simulations, they found that for a given $a - b$ distribution, the transition between regimes in the slip spectrum depends on the ratio between the width W of the rate-weakening zone and the nucleation size h^*. Rubin (2008) found similar transitions but concluded that the ratios, at which the transitions occur, depend on the value of $a - b$. We keep the width of the rate-weakening zone fixed and change h^* by varying L and P_B. For each of four values of L between 10^{-4} and 10^{-1} m, we systematically increase P_B and thereby W/h^* and P_B/L (Figure 16a). Note that we keep all other physical model parameters fixed. Therefore, for a given W/h^* (and hence P_B/L), theoretical estimates of h^* and Λ_0 are constant. Simulations were done with a grid size of 500 m for $W/h^* < 4$ and 250 m for $W/h^* > 4$; other numerical parameters are the same as in the reference setup.

The simulated slip spectrum consists of four regimes (Figure 16a): decaying oscillations toward stable sliding (I), periodic aseismic slip transients (II), periodic (III), and aperiodic seismic slip (IV). In regime I ($W/h^* < 0.65$), the amplitude of slip transients decays with time until the fault steadily slips at the loading slip velocity V_l (Figure 16b). The number of oscillations until stable sliding increases with higher W/h^*. In regimes II and III, events recur in a periodic manner after an initial loading and stabilizing period. The slip velocity of these events in each model increases as a function of P_B/L from aseismic in regime II (Figure 16 c) to seismic rates in regime III (Figure 16d). In regime II ($0.65 < W/h^* < 1.0$), slip velocity increases with P_B and is independent of L. In contrast, in regime III ($1.0 > W/h^* > 6.0$), higher slip velocities are reached for larger L at constant W/h^*. Therefore, slip velocity increases both with P_B and L (Figure S9a). For $W/h^* = 3.16$, profiles of τ_{II} along the fault at the end of the nucleation process of the first event nearly overlap after dividing them by P_B (Figure S9c). Consequently, simulated h^* and Λ are the same in these models as predicted by the corresponding theoretical estimates. A similar overlap is obtained for the snapshots during dynamic rupture propagation and slip weakening of τ_{II} (Figure S9d and S9e). Regime IV is characterized by an aperiodic earthquake pattern with variable recurrence times and maximum slip velocities (Figure 16e). The complexity arises as events, which rupture only parts or the full rate-weakening zone, occur in between full size ruptures. Partial rupture occurs

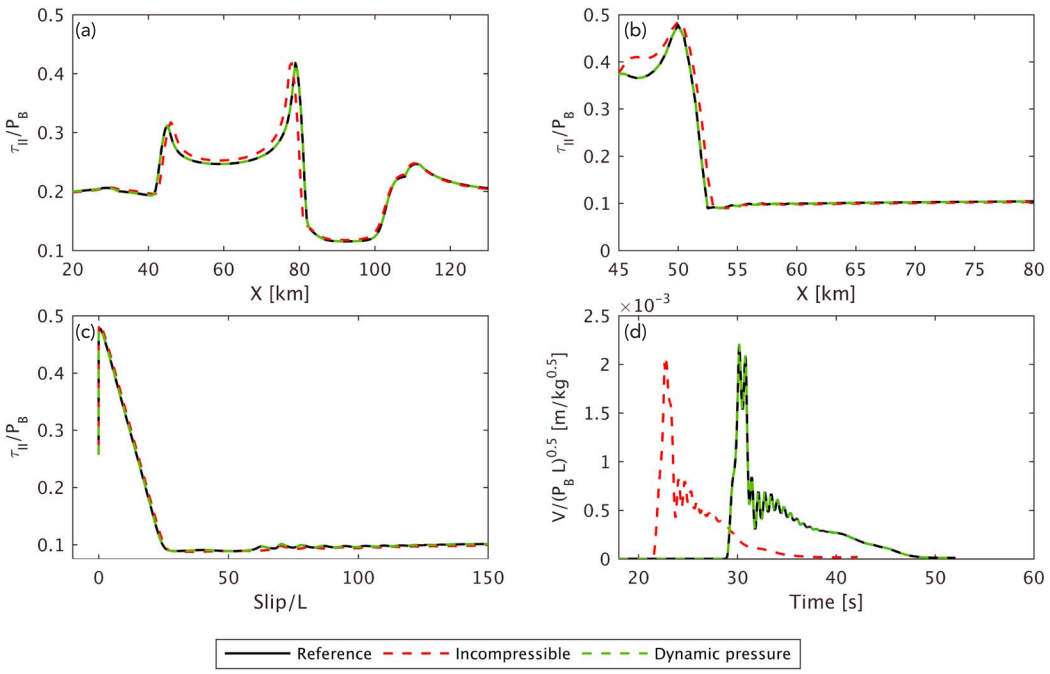

Figure 17. Dynamic pressure and compressibility. Comparison between reference solution ($K = 50$ GPa, constant $P_B = 5$ MPa in rate- and state-dependent friction, RSF, equation), an approximately incompressible solution (red, $K = 500$ TPa, $P_B = 7.5$ MPa constant in RSF equation) and a solution using a dynamically evolving pressure in RSF equations (green, $K = 50$ GPa, $P_B = 5$ MPa, true P is used in RSF equation). (a) Snapshot of the square root of the second deviatoric stress invariant at the end of the nucleation process. (b) Snapshot of the square root of the second deviatoric stress invariant when the rupture front has reached $X = 50$ km. (c) Slip weakening of the square root of the second deviatoric stress invariant in $X = 50$ km during the passage of the rupture front. Effective stress is normalized by the initial background pressure P_B. (d) Evolution of slip velocity in $X = 50$ km, divided by $\sqrt{P_B L}$.

due to the interaction with reflected waves from the boundaries, which act as stopping waves. Therefore, this regime is rather numerical (i.e., influenced by nonabsorbing boundary conditions) in contrast to the other physical regimes.

5.2. Dynamic Pressure in RSF

For simplicity and to compare our results to classical earthquake cycle simulations, we kept thus far the value of P in the invariant RSF formulation (equation (20)) constant at the initial background pressure P_B. We rerun the reference model ($\Delta x = 500$ m) by using the dynamic value P in the RSF equations. We observe only minor changes in rupture properties (<1.0%) with respect to the reference solution for the first event (Figure 17) and the sequence of first events (not shown). P varies outside the fault zone during the rupture passage, but it remains at the initial background pressure P_B within the fault zone. The reason is that large plastic deformation rates along the fault effectively create a weak layer that is surrounded by the strong host rock. The pressure stays constant, because the weak layer cannot accommodate any significant dynamic pressure gradients. Consequently, the obtained solution is almost identical to the case when $P = P_B$ in RSF equations. While this result indicates that it is sufficient to treat P as a constant along mature fault zones, dynamic pressure becomes important for the long-term fault evolution in terms of fault orientation and self-consisting fault strength (e.g., Meyer et al., 2017).

5.3. Incompressible Medium

To understand the difference with respect to incompressible models commonly used in geodynamic simulations (e.g., Gerya, 2010), we run the reference model (with $\Delta x = 500$ m) with a nearly incompressible medium ($K = 500$ TPa, $\nu \approx 0.5$). In this incompressible medium, only shear waves occur because p waves cannot exist. To compare models with the same theoretical nucleation and cohesive zone lengths, we change P_B to 7.5 MPa. Correcting for the different P_B leads to similar profiles of the square root of the second deviatoric stress invariant during nucleation and dynamic rupture propagation and to similar slip weakening of the square root of the second deviatoric stress invariant and time series of slip velocity (Figures 17a–17d). Consequently,

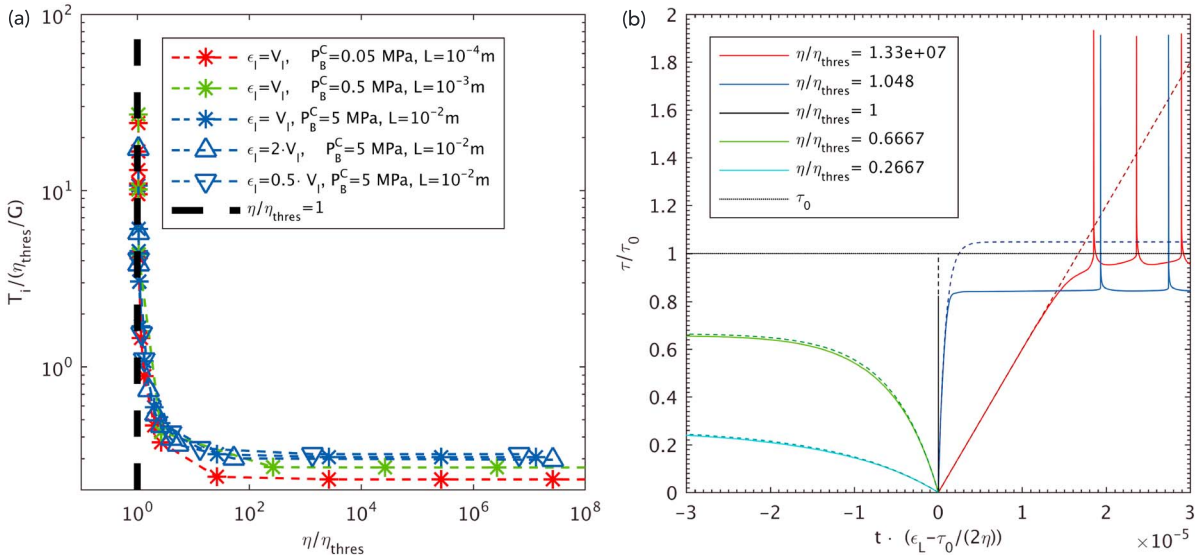

Figure 18. Influence of viscous deformation on the earthquake cycle. (a) Interseismic duration T_i scaled by the threshold relaxation time scale η_{thres}/G versus viscosity scaled by the threshold viscosity $\eta_{\text{thres}} = \frac{\mu_0 P_B}{2\dot{\epsilon}_L}$. Results are shown for simulations conducted with the same W/h^*, but different P_B, L and $\dot{\epsilon}^l$ as indicated in the legend. (b) Impact of viscosity in the reference model. Solid lines are simulated curves, colored dashed lines show for each model the analytical viscoelastic stress evolution for constant loading strain rate $\dot{\epsilon}^l$ (equation (36)). Time axis is scaled by the difference between the asymptotic viscoelastic stress $\tau_d = 2\eta\dot{\epsilon}^l$ and the reference RSF strength τ_0, divided by 2η. Negative or zero values imply that τ_d is smaller than or equal to τ_0, respectively, which corresponds to η being smaller than or equal to the threshold viscosity η_{thres}, respectively.

simulated values for the nucleation length, cohesive zone lengths, and slip weakening distance of the first event are the same as in the compressible case to within one grid cell. It appears that the normalized stress at and ahead of the rupture front is larger than in the compressible case (Figure 17b). This could reflect the rigidity of the incompressible medium, which is unable to redistribute stresses by volumetric elastic deformation. Similarly, the rupture arrives earlier at the reference location in the incompressible case (Figure 17d).

6. Role of Viscous Deformation

One advantage of our approach is the ability to include various viscous deformation mechanisms in the simulation of earthquake cycles. In the models analyzed thus far, viscous deformation is essentially inactive due to the extremely high effective viscosity of $5 \cdot 10^{26}$ Pa s. In nature, especially in deep and warm portions of subduction zones, faults could be surrounded by or penetrate material with distinctly lower viscosity (10^{18} – 10^{19} Pa s), as suggested in simulations (e.g., van Dinther, Gerya, Dalguer, Mai, et al., 2013) and from observations after larger earthquakes (e.g., Freed et al., 2017). To test the influence of viscous deformation on the earthquake cycle, we run simulations ($\Delta x = 500$ m) with lower effective viscosities in the entire model. Note that viscosity is constant and linear during these simulations, which for simplicity ignores any nonlinear effects of temperature and strain rate on the viscous rheology typically adopted in geodynamic models. We furthermore note that in the simulations with a low viscosity, the adaptive time step is controlled by Δt_{vep} to resolve the viscoelastic Maxwell relaxation time scale.

For a wide range of high viscosities, viscous deformation is essentially ineffective (Figures 18a and 18b). This represents the brittle regime. Lowering viscosity reveals the existence of a viscosity threshold η_{thres}, below which only viscous deformation occurs in the ductile regime (Figure 18b). In the transition between brittle and ductile deformation, in which the viscosity decreases the threshold, interseismic duration increases sharply as $\frac{\eta}{\eta - \eta_{\text{thres}}}$ with decreasing viscosity (Figure 18a). Furthermore, the hypocenter location of the earthquake is shifted to the center and the amount of seismic slip decreases. At low viscosities, for which the hypocenter is located at the center of the rate-weakening zone, the nucleation length increases with decreasing viscosities until it reaches the size of the rate-weakening zone so that no earthquakes nucleate.

The role of viscous deformation can be understood by considering its influence on the interseismic loading and subsequent nucleation process. Lowering viscosity reduces the Maxwell relaxation time scale η/G toward the interseismic duration of the earthquake cycle (Figure 18b). Furthermore, it reduces the strain rate

dependent ductile strength ($\tau_d = 2\eta\dot{\varepsilon}_I$) to which stress approaches after several Maxwell times. As we showed in sections 3.2 and 3.3.1, interseismic loading increases the stress toward $\tau_0 = \mu_0 P_B$ first in the rate-strengthening and then in the rate-weakening zone, which induces stable slip. The growing patch of stable slip in the rate-weakening zone becomes unstable when its length reaches the nucleation size. This process is prevented if τ_I is smaller than τ_0 (Figure 18b). Following this notion, a transition is to be expected if $\tau_I = \tau_0$, which leads us to the definition of the viscosity threshold as

$$\eta_{\text{thres}} = \frac{\tau_0}{2\dot{\varepsilon}_I} = \frac{\mu_0 P_B}{2\dot{\varepsilon}_I}. \tag{44}$$

The corresponding threshold Maxwell relaxation time is η_{thres}/G. Equation (44) predicts that the viscosity threshold is a function of P_B, loading strain rate $\dot{\varepsilon}_I$ and μ_0. We test this prediction by investigating the role of viscosity in models with different P_B (at constant W/h^*) and different loading strain rate (Figure 18a). We find that the simulated thresholds agree with the theoretical prediction. Curves of interseismic duration as a function of viscosity overlap if we normalize viscosity by η_{thres} and interseismic duration by the threshold Maxwell relaxation time (Figure 18a). Note that $\dot{\varepsilon}_I$, and, hence, the viscosity threshold depend on the size of the model domain in the present simulations, since the distance of the fault to the top and bottom boundary influences the loading rate.

This shows that the application of RSF in a viscoelastic medium agrees with the existing basic understanding of the brittle-ductile transition zone in that viscous deformation becomes dominant as soon as the strain rate dependent ductile strength is smaller than the pressure-dependent brittle strength. We show that in case of RSF, this brittle strength is defined by the reference strength, because this is the stress level at which earthquakes nucleate. Decreasing the loading strain rate decreases the ductile strength and increases the interseismic duration toward the Maxwell relaxation time scale. This implies that slowly loaded faults are more prone to be affected by viscous deformation as their long interseismic loading phase might interfere with the Maxwell characteristic time of the surrounding medium.

7. Discussion

We have presented the distinctly improved STM-RSF approach for modeling a wide slip spectrum in a continuum. In the following we discuss the invariant reformulation of RSF and the results concerning its numerical accuracy and stability, including arising issues related to wave reflections and the internodal interpolation scheme. Finally, we provide an outlook to the full potential of the STM-RSF approach to address current limitations.

7.1. Invariant RSF

The key new model component of our STM-RSF approach is the implementation of an invariant reformulation of RSF in a continuum-mechanics framework. By replacing physical quantities related to the fault's orientation by invariants of these properties, we obtain a formulation of RSF that is applicable in a continuum.

From a geodynamic modeling point of view it may appear strange that RSF assumes permanent yielding. We show that this is always the case in our simulations, even during the elastic loading phase. Following the viewpoint of Nakatani (2001), the classical geodynamic and RSF friction concepts can be reconciled if we consider the term $\sigma_n \left[\mu_0 + b \ln\left(\frac{\theta V_0}{L}\right)\right]$ as the interface strength, at which noticeable plastic deformation appears. The smooth property of RSF in vicinity of this interface strength due to the "viscosity-like" direct effect facilitates resolution of the onset of noticeable plastic deformation. Resolving this onset is a problem for the commonly applied nonsmooth static yield functions in geodynamic simulations, which assume that no plastic deformation occurs if the square root of the second deviatoric stress invariant is lower than a constant, pressure-dependent material strength. The same problem occurred for rate-dependent friction in the previous STM approach (van Dinther, Gerya, Dalguer, Mai, et al., 2013), such that implementing RSF has lead to a large improvement with respect to that.

From the classical earthquake cycle modeling point of view it may appear strange that we apply RSF in the entire medium as it is usually applied only along planar, infinitely thin fault planes in an otherwise elastic continuum. Here we showed in an extreme case that RSF is applicable in the entire medium, while it only becomes active in regions of low state.

Our approach enables to take the role of the dynamic pressure into account. Most simulations in this study, though, have been conducted with a constant value for pressure in the RSF equations as in classical earthquake cycle simulations. We have found that the dynamic pressure plays only a minor role in the investigated simple setup of a mature straight fault zone. The dynamic pressure field is expected to become more important for the long-term fault evolution in terms of fault orientation and self-consistent fault strength (e.g., Meyer et al., 2017).

7.2. Numerical Accuracy and Stability

In the analysis of one earthquake cycle in the mature strike-slip fault setup, we demonstrated a good agreement of our results with a number of analytical approximations. The initial interseismic stress increase due to the imposed shear motion at the model boundaries agrees with a purely elastic stress built up. Toward the end of the subsequent nucleation process, a cohesive zone starts to be created, whose length agrees with the quasi-static cohesive zone length Λ_0 (Day et al., 2005; Lapusta & Liu, 2009). The length of the slip area, along which the instability evolves, is close to the nucleation length given in Rubin and Ampuero (2005) for the suited a/b range. During rupture propagation, the evolution of the cohesive zone length follows the changes in rupture velocity, as predicted by equation (38). In stage I, during which the rupture front approaches, stress follows closely the direct effect. The stress decrease in stage III agrees with the slip weakening approximation given in Lapusta and Liu (2009). Furthermore, during the postseismic phase, the relaxation process in the rate-strengthening region and the interseismic healing in the interseismic agree with analytical curves. This very good agreement with analytical approximations shows that our new approach resolves all stages of an earthquake cycle from slow interseismic loading, to the nucleation, to dynamic rupture propagation, and postseismic deformation. As no analytical solution for the whole earthquake cycle exists, it may be desirable in the future to conduct a benchmark comparison study between STM-RSF and classical earthquake cycle simulations in a similar way to what has been done in the dynamic rupture modeling community (Harris et al., 2009). However, it would be necessary to extend our method to 3-D in order to compare the models in a vertical strike-slip models using an anti-plane formulation.

We also showed that the simulated slip spectrum is in good agreement with findings of Liu and Rice (2007) and Rubin (2008). Our results confirm their findings that for a fixed $a-b$ distribution, the transitions between decaying oscillations, period aseismic, and period seismic slip depend on the ratio between the length W of the rate-weakening zone and the nucleation length h^*. Differences between our and other studies in this ratio at the transitions and maximum slip velocity within the slip regimes are likely due to the different model setups (e.g., a-b distribution, loading and boundary conditions) and the different definition of the nucleation size. Similar reasons might approach apply for the fact that we did not find a complex periodic behavior for W/h^* close to unity as in Liu and Rice (2007) and Rubin (2008) or periodic doubling of slow and fast ruptures as in Veedu and Barbot (2016).

For the first time with our approach, we are able to simulate seismic slip rates, at which shear waves, and, following the implementation of compressibility, pressure waves are generated. We do not implement any damping through a radiation term typically applied in quasi-dynamic simulations (Rice, 1993), such that we solve for the full dynamics of the rupture propagation process. This is important as the application of the quasi-dynamic approach leads to quantitative differences in rupture speed and slip velocity in comparison to the full dynamic approach in case of the standard RSF (Thomas et al., 2014). However, related to this progress is a limitation, because we do not have yet implemented absorbing boundaries in our approach such that these waves are reflected back to the fault. Interaction of these reflected waves with the fault shows the need to implement absorbing boundary conditions. In the reference model, they slightly modify the stress distribution along the fault at the beginning of the postseismic period such that the hypocenter locations of the subsequent events are affected. In models with smaller nucleation sizes, a numerical regime of slip spectrum arises, because reflected waves stop the rupture such that partial ruptures appear. Implementing absorbing boundary conditions should eliminate these problems.

We have tested the numerical stability by assessing the convergence of our results with respect to several numerical parameters. We showed that results are virtually independent of the number of and initial distribution of markers, which we use to track rock properties to avoid numerical diffusion in large deformation experiments. This is a clear improvement with respect to our previous STM approach, which showed a distinct sensitivity to the initial marker distribution for the rupture time of the first event and the subsequent sequence of events (van Dinther, Gerya, Dalguer, Mai, et al., 2013). The improvements are mainly the results

of introducing the grid-based calculation of the yield strength and the new marker-based advection scheme for stresses, pressure, and velocities. This marker-based advection scheme is optimized for small deformation but potentially requires further investigation in case of large deformation experiments.

Results converge with respect to grid size and the time step factor in case of the first event. For subsequent events, results converge as well, but variations are introduced due to the different resolution of the waves reflected from the boundaries. The amplitude of numerical oscillations in the slip rate function due to insufficient resolution of the grid size decreases with increasing spatial resolution. In addition, a key indicator for numerical precision—the rupture arrival time—converges with decreasing grid size. We showed that the interpolation of the viscoplastic viscosity—either harmonic or arithmetic averaging—introduces some changes in the results, although the general pattern of events remains unaffected. These differences between interpolation schemes become smaller with decreasing grid size. Arithmetic averaging shows a faster convergence with grid size than harmonic averaging. Day et al. (2005) suggested to resolve the cohesive zone during dynamic rupture propagation with at least 3 grid points in boundary integral method and with at least 5 points in finite difference methods that use split nodes to discretize a fault continuity. The smallest cohesive zone length in the reference model 2 km in the reference model (Figure 8), which is 8 times larger than grid size at which a convergent solution is obtained for arithmetic averaging. The difference in convergence rate might be related to the different effective thickness of the fault zone. Whereas the fault zone thickness is one Δx in case of harmonic averaging, it is less than Δx or even a plane in case of arithmetic averaging. The slower convergence for an inelastic zone of one grid size is comparable to the grid convergence for finite fault zones using the stress glut method (Andrews, 1999). This stress glut method showed a slower convergence than using the finite-difference method with split nodes in dynamic rupture simulations (Dalguer & Day, 2006). Despite the better convergence of dynamic rupture properties in case of arithmetic averaging, we suggest to use harmonic averaging. In addition to the convergence analysis in the reference model setup, we analyzed the convergence with respect to grid size in case of a purely rate-strengthening fault in Text S5. It shows that harmonic averaging leads to a convergent solution, whereas a slight divergence is noted for arithmetic averaging. Also, preliminary tests show that arithmetic averaging produces problems in convergence with Picard iterations if we consider a fault that evolves spontaneously. Alternatively, it might be necessary to reconsider the way plastic slip is calculated on the staggered grid nodes to avoid the need to interpolate viscosities.

7.3. Limitations and Outlook

The reference model setup has mainly been designed for testing purposes. This model setup, which resembles the ones used in 2-D classical earthquake cycle simulations, has the following disadvantages. First, the loading of the fault is dependent on the size of the domain. The smaller it is, the faster the loading and the shorter the recurrence time. The model size therefore also influences the loading strain rate dependent viscosity threshold, below which viscous deformation dominates. This model size dependency, typical in such model setups (Hajaroalsvadi & Elbanna, 2017), highlights the need to simulate realistic tectonic loading in a crustal-scale environment. Second, the analysis in this setup is limited to an essentially predefined fault plane without complexity in the off-fault medium. Third, viscosity is linear and independent of temperature and strain rate. These limitations can be overcome if the full potential of the STM-RSF is explored in a large-scale model setup with temperature-dependent viscosity similar to the one adopted in van Dinther, Gerya, Dalguer, Corbi, et al. (2013). One part of this potential—the ability to simulate the spontaneous generation and propagation of faults—will be explored in future. It is then possible to study earthquakes in relation to the long-term history of a fault. In these models, in which faults cross the numerical grid, it is necessary to improve the accuracy of this method. To resolve the fault zone thickness, it will be essential to prevent localization either by diffusion through a physical mechanism or by numerical regularization (e.g., nonlocal plasticity).

The current numerical code requires optimization to increase computational efficiency to simulate earthquake cycles in 2D with more realistic, and hence, smaller nucleation zone lengths in a lithospheric scale setup and to extend this approach to the third dimension to facilitate comparison to nature (Lapusta & Liu, 2009). Potential improvements of our approach to enhance computational speed are (i) Message Passing Interface (MPI) parallelization, (ii) combined explicit-implicit time descritization, as demonstrated in Kaneko et al. (2011) to reduce the number of time steps solved with the expensive implicit method, and (iii) an adaptive staggered grid (Gerya, 2010; Gerya et al., 2013) to resolve only the finest length scale with the highest resolution.

8. Conclusions

We have significantly improved the self-consistent STM numerical modeling approach (van Dinther, Gerya, Dalguer, Corbi, et al., 2013; van Dinther, Gerya, Dalguer, Mai, et al., 2013) to simulate earthquake cycles in a viscoelastoplastic compressible continuum. This paper shows that our continuum mechanics-based methodology, originally designed for lithospheric-scale geodynamic problems (Gerya & Yuen, 2007), resolves all stages of an earthquake cycle with realistic rupture properties. This includes seismic slip rates and rupture speeds at which shear and pressure waves are generated. The key ingredient of our STM-RSF modeling approach is the implementation of an invariant reformulation of the rate- and state-dependent friction equations, which provides the possibility to apply RSF to simulate earthquakes along spontaneously evolving faults. Other methodological improvements with respect to the previous STM approach are grid-based Picard iterations to minimize the error in the yield function and an optimized advection scheme for small displacements. Furthermore, we modified the adaptive time step usually applied in classical earthquake cycle simulation (Lapusta & Liu, 2009; Lapusta et al., 2000). While the original scheme is crucial during earthquake cycles in an essentially elastic medium, we included additional constraints to resolve the initial loading phase, state-independent friction, and the Maxwell time scale in viscoelastic simulations. Furthermore, we do not adopt a minimum time step cutoff, because it reduces the convergence rate during Picard iterations.

To enable the comparison to analytical approximations and theoretical estimates, we have chosen a simple setup of a mature, essentially predefined strike-slip fault zone embedded in predominantly elastic medium. In the analysis of an earthquake cycle, we found a good fit between our results with analytical approximations, including postseismic relaxation in the rate-strengthening zone, interseismic fault healing, and the linear slip weakening process of stress at the rupture front. That, and the good agreement between the simulated and theoretical estimates of nucleation and cohesive zone lengths, demonstrates that our calculations are sufficiently accurate along a fault, which is aligned with the numerical grid. Convergence with respect to important numerical parameters, including grid size and time step, shows that our method is stable. Interpolation of viscoplastic viscosity between staggered grid points leads to variations of the results, but the periodic and characteristic earthquake cycle is not affected and variations become smaller for higher spatial resolution. We suggest to use harmonic interpolation of viscoplastic viscosity, which is based on the stability of the results presented here and applications in more complicated setups. One not yet resolved issue is that the generated seismic waves are reflected back from the model boundaries. Furthermore, resolving the fault zone thickness will be necessary to accurately resolve earthquakes also along faults in an oblique angle to the numerical grid.

Finally, we demonstrated that the approach is applicable to study the role of viscous deformation in the earthquake cycle. We identified and theoretically confirmed the existence of a viscosity threshold below which earthquakes cannot nucleate. The threshold is shown to depend on the reference strength of RSF ($= \mu_0 P$) and the loading strain rate, which is in agreement with previous work on the brittle-ductile transition.

Acknowledgments
This study was supported by Swiss National Science Foundation grants 200021-153524 and 200021-169880. The numerical simulations for this work were conducted on "Monch" at the Swiss National Supercomputing Centre (CSCS Lugano). We are grateful for Pablo Ampuero for an initial discussion about rate- and state-dependent friction during his visit at ETH Zurich. We thank Luca dal Zilio and particularly Casper Pranger for useful discussions. Furthermore, we thank Sylvain Barbot and an anonymous reviewer for constructive comments and suggestions, which helped to improve the manuscript. Simulation data and Matlab scripts to produce the figures presented in the paper are available in the repository cited in the references (Herrendörfer et al., 2018). Additionally, this repository contains parts of the code, which show the implementation of the invariant RSF formulation. Furthermore, it contains an executable, with which the reference model (section 3) can be rerun.

References

Ampuero, J.-P., & Rubin, A. M. (2008). Earthquake nucleation on rate and state faults — Aging and slip laws. *Journal of Geophysical Research*, *113*, B01302. https://doi.org/10.1029/2007JB005082

Andrews, D. J. (1999). Test of two methods for faulting in finite-difference calculations. *Bulletin of the Seismological Society of America*, *89*(4), 931–937.

Barbot, S., & Fialko, Y. (2010). A unified continuum representation of postseismic relaxation mechanisms: Semi-analytic models of afterslip, poroelastic rebound and viscoelastic flow. *Geophysical Journal International*, *182*(3), 1124–1140. https://doi.org/10.1111/j.1365-246X.2010.04678.x

Ben-Zion, Y., & Rice, J. R. (1993). Earthquake failure sequences along a cellular fault zone in a three-dimensional elastic solid containing asperity and nonasperity regions. *Journal of Geophysical Research*, *98*(B8), 14,109–14,131. https://doi.org/10.1029/93JB01096

Ben-Zion, Y., & Rice, J. R. (1997). Dynamic simulations of slip on a smooth fault in an elastic solid. *Journal of Geophysical Research*, *102*(B8), 17,771–17,784. https://doi.org/10.1029/97JB01341

Bhattacharya, P., Rubin, A. M., Bayart, E., Savage, H. M., & Marone, C. (2015). Critical evaluation of state evolution laws in rate and state friction: Fitting large velocity steps in simulated fault gouge with time-, slip-, and stress-dependent constitutive laws. *Journal of Geophysical Research: Solid Earth*, *120*, 6365–6385. https://doi.org/10.1002/2015JB012437

Bizzarri, A., & Cocco, M. (2003). Slip-weakening behavior during the propagation of dynamic ruptures obeying rate- and state-dependent friction laws. *Journal of Geophysical Research*, *108*, 2373. https://doi.org/10.1029/2002JB002198

Chen, J., & Spiers, C. J. (2016). Rate and state frictional and healing behavior of carbonate fault gouge explained using microphysical model. *Journal of Geophysical Research: Solid Earth*, *121*, 8642–8665. https://doi.org/10.1002/2016JB013470

Chester, F. M. (1994). Effects of temperature on friction: Constitutive equations and experiments with quartz gouge. *Journal of Geophysical Research*, *99*(B4), 7247–7261. https://doi.org/10.1029/93JB03110

Cocco, M., & Bizzarri, A. (2002). On the slip-weakening behavior of rate- and state dependent constitutive laws. *Geophysical Research Letters*, *29*(11), 1516. https://doi.org/10.1029/2001GL013999

Cocco, M., Bizzarri, A., & Tinti, E. (2004). Physical interpretation of the breakdown process using a rate- and state-dependent friction law. *Tectonophysics*, *378*(3–4), 241–262. https://doi.org/10.1016/j.tecto.2003.09.015

Corbi, F., Funiciello, F., Moroni, M., van Dinther, Y., Mai, P. M., Dalguer, L. A., & Faccenna, C. (2013). The seismic cycle at subduction thrusts: 1. Insights from laboratory models. *Journal of Geophysical Research: Solid Earth*, *118*, 1483–1501. https://doi.org/10.1029/2012JB009481

Corbi, F., Herrendörfer, R., Funiciello, F., & van Dinther, Y. (2017). Controls of seismogenic zone width and subduction velocity on interplate seismicity: Insights from analog and numerical models. *Geophysical Research Letters*, *44*, 6082–6091. https://doi.org/10.1002/2016GL072415

Dal Zilio, L., van Dinther, Y., Gerya, T. V., & Pranger, C. C. (2018). Seismic behaviour of mountain belts controlled by plate convergence rate. *Earth and Planetary Science Letters*, *482*, 81–92. https://doi.org/10.1016/j.epsl.2017.10.053

Dalguer, L. A., & Day, S. M. (2006). Comparison of fault representation methods in finite difference simulations of dynamic rupture. *Bulletin of the Seismological Society of America*, *96*(5), 1764–1778. https://doi.org/10.1785/0120060024

Day, S. M., Dalguer, L. A., Lapusta, N., & Liu, Y. (2005). Comparison of finite difference and boundary integral solutions to three-dimensional spontaneous rupture. *Journal of Geophysical Research*, *110*, B12307. https://doi.org/10.1029/2005JB003813

de Borst, R., & Sluys, L. J. (1991). Localisation in a Cosserat continuum under static and dynamic loading conditions. *Computer Methods in Applied Mechanics and Engineering*, *90*(1), 805–827. https://doi.org/10.1016/0045-7825(91)90185-9

Deubelbeiss, Y., & Kaus, B. J. P. (2008). Comparison of Eulerian and Lagrangian numerical techniques for the Stokes equations in the presence of strongly varying viscosity. *Physics of the Earth and Planetary Interiors*, *171*(1), 92–111. https://doi.org/10.1016/j.pepi.2008.06.023

Dieterich, J. H. (1972). Time-dependent friction in rocks. *Journal of Geophysical Research*, *77*(20), 3690–3697. https://doi.org/10.1029/JB077i020p03690

Dieterich, J. H. (1978). Time-dependent friction and the mechanics of stick-slip. *Pure and Applied Geophysics*, *116*(4–5), 790–806. https://doi.org/10.1007/BF00876539

Dieterich, J. H. (1979). Modeling of rock friction: 1. Experimental results and constitutive equations. *Journal of Geophysical Research*, *84*(B5), 2161–2168. https://doi.org/10.1029/JB084iB05p02161

Dieterich, J. H. (1981). Constitutive properties of faults with simulated gouge. In N. L. Carter, M. Friedman, J. M. Logan, & D. W. Stearns (Eds.), *Mechanical behavior of crustal rocks: The Handin volume, Geophysical Monograph Series* (Vol. 24, pp. 103–120). Washington, DC: American Geophysical Union.

Dieterich, J. H. (1992). Earthquake nucleation on faults with rate-and state-dependent strength. *Tectonophysics*, *211*(1–4), 115–134. https://doi.org/10.1016/0040-1951(92)90055-B

Dieterich, J. H., & Kilgore, B. D. (1994). Direct observation of frictional contacts: New insights for state-dependent properties. *Pure and Applied Geophysics PAGEOPH*, *143*(1–3), 283–302. https://doi.org/10.1007/BF00874332

Drucker, D. C., & Prager, W. (1952). Soil mechanics and plastic analysis of limit design. *Quarterly of Applied Mathematics*, *10*, 157–165.

Duretz, T., Gerya, T. V., & May, D. A. (2011). Numerical modelling of spontaneous slab breakoff and subsequent topographic response. *Tectonophysics*, *502*(1–2), 244–256. https://doi.org/10.1016/j.tecto.2010.05.024

Erickson, B. A., & Dunham, E. M. (2014). An efficient numerical method for earthquake cycles in heterogeneous media: Alternating subbasin and surface-rupturing events on faults crossing a sedimentary basin. *Journal of Geophysical Research: Solid Earth*, *119*, 3290–3316. https://doi.org/10.1002/2013JB010614

Erickson, B. A., Dunham, E. M., & Khosravifar, A. (2017). A finite difference method for off-fault plasticity throughout the earthquake cycle. *Journal of the Mechanics and Physics of Solids*, *109*, 50–77. https://doi.org/10.1016/j. jmps.2017.08.002

Freed, A. M., Hashima, A., Becker, T. W., Okaya, D. A., Sato, H., & Hatanaka, Y. (2017). Resolving depth-dependent subduction zone viscosity and afterslip from postseismic displacements following the 2011 Tohoku-oki, Japan earthquake. *Earth and Planetary Science Letters*, *459*, 279–290. https://doi.org/10.1016/j.epsl.2016.11.040

Gerya, T. V. (2010). *Introduction to numerical geodynamic modelling*. Cambridge, UK: Cambridge University Press.

Gerya, T. V., May, D. A., & Duretz, T. (2013). An adaptive staggered grid finite difference method for modeling geodynamic Stokes flows with strongly variable viscosity. *Geochemistry, Geophysics, Geosystems*, *14*, 1200–1225. https://doi.org/10.1002/ggge.20078

Gerya, T. V., & Yuen, D. A. (2003). Characteristics-based marker-in-cell method with conservative finite-differences schemes for modeling geological flows with strongly variable transport properties. *Physics of the Earth and Planetary Interiors*, *140*(4), 293–318. https://doi.org/10.1016/j.pepi.2003.09.006

Gerya, T. V., & Yuen, D. A. (2007). Robust characteristics method for modelling multiphase visco-elasto-plastic thermo-mechanical problems. *Physics of the Earth and Planetary Interiors*, *163*, 83–105. https://doi.org/10.1016/j.pepi. 2007.04.015

Hajaroalsvadi, S., & Elbanna, A (2017). A new hybrid numerical scheme for simulating fault ruptures with near-fault bulk heterogeneities. *ArXiv e-prints*.

Harris, R. A., Barall, M., Archuleta, R., Dunham, E., Aagaard, B., Ampuero, J. P., et al. (2009). The SCEC/USGS dynamic earthquake rupture code verification exercise. *Seismological Research Letters*, *80*(1), 119–126. https://doi.org/10.1785/gssrl.80.1.119

Herrendörfer, R., Gerya, T., & van Dinther, Y. (2018). Model data set to "An invariant rate- and state-dependent friction formulation for visco-elasto-plastic earthquake cycle simulations". https://doi.org/10.3929/ethz-b-000263453

Herrendörfer, R., van Dinther, Y., Gerya, T., & Dalguer, L. A. (2015). Earthquake supercycle in subduction zones controlled by the width of the seismogenic zone. *Nature Geoscience*, *8*(6), 471–474. https://doi.org/10.1038/ngeo2427

Heuret, A., Lallemand, S., Funiciello, F., Piromallo, C., & Faccenna, C. (2011). Physical characteristics of subduction interface type seismogenic zones revisited. *Geochemistry, Geophysics, Geosystems*, *12*, Q01004. https://doi.org/10.1029/2010GC003230

Kagan, Y. Y., & Jackson, D. D. (2013). Tohoku earthquake: A surprise. *Bulletin of the Seismological Society of America*, *103*(2B), 1181–1194. https://doi.org/10.1785/0120120110

Kaneko, Y., Ampuero, J.-P., & Lapusta, N. (2011). Spectral-element simulations of long-term fault slip: Effect of low-rigidity layers on earthquake-cycle dynamics. *Journal of Geophysical Research*, *116*, B10313. https://doi.org/10.1029/2011JB008395

Kaneko, Y., Lapusta, N., & Ampuero, J.-P. (2008). Spectral element modeling of spontaneous earthquake rupture on rate and state faults: Effect of velocity-strengthening friction at shallow depths. *Journal of Geophysical Research*, *113*, B09317. https://doi.org/10.1029/2007JB005553

Lambert, V., & Barbot, S. (2016). Contribution of viscoelastic flow in earthquake cycles within the lithosphere-asthenosphere system. *Geophysical Research Letters*, *43*, 10,142–10,154. https://doi.org/10.1002/2016GL070345

Lapusta, N. (2003). Nucleation and early seismic propagation of small and large events in a crustal earthquake model. *Journal of Geophysical Research*, *108*, B42205. https://doi.org/10.1029/2001JB000793

Lapusta, N., & Barbot, S. (2012). Models of earthquakes and aseismic slip based on laboratory-derived rate and state friction laws. In A. Bizzarri & H. S. Bhat (Eds.), *Mech. faulting from lab. to real earthquakes* (Vol. 661, pp. 153–207). Kerala India: Research Signpost.

Lapusta, N., & Liu, Y. (2009). Three-dimensional boundary integral modeling of spontaneous earthquake sequences and aseismic slip. *Journal of Geophysical Research*, *114*, B09303. https://doi.org/10.1029/2008JB005934

Lapusta, N., Rice, J. R., Ben-Zion, Y., & Zheng, G. (2000). Elastodynamic analysis for slow tectonic loading with spontaneous rupture episodes on faults with rate- and state-dependent friction. *Journal of Geophysical Research*, *105*, 23,765–23,789. https://doi.org/10.1029/2000JB900250

Lavier, L. L., Buck, W. R., & Poliakov, A. N. B. (2000). Factors controlling normal fault offset in an ideal brittle layer. *Journal of Geophysical Research*, *105*, 23,431–23,442. https://doi.org/10.1029/2000JB900108

Leeman, J. R., Saffer, D. M., Scuderi, M. M., & Marone, C. (2016). Laboratory observations of slow earthquakes and the spectrum of tectonic fault slip modes. *Nature Communications*, *7*, 11104. https://doi.org/10.1038/ncomms11104

Liu, Y., & Rice, J. R. (2007). Spontaneous and triggered aseismic deformation transients in a subduction fault model. *Journal of Geophysical Research*, *112*, B09404. https://doi.org/10.1029/2007JB004930

Lyakhovsky, V., & Ben-Zion, Y. (2009). Evolving geometrical and material properties of fault zones in a damage rheology model. *Geochemistry, Geophysics, Geosystems*, *10*, Q11011. https://doi.org/10.1029/2009GC002543

Lyakhovsky, V., Ben-Zion, Y., Ilchev, A., & Mendecki, A. (2016). Dynamic rupture in a damage-breakage rheology model. *Geophysical Journal International*, *206*(2), 1126. https://doi.org/10.1093/gji/ggw183

Marone, C. (1998). Laboratory-derived friction laws and their application to seismic faulting. *Annual Review of Earth and Planetary Sciences*, *26*, 643–696.

Marone, C., Hobbs, B. E., & Ord, A. (1992). Coulomb constitutive laws for friction: Contrasts in frictional behavior for distributed and localized shear. *Pure and Applied Geophysics*, *139*(2), 195–214. https://doi.org/10.1007/BF00876327

Meyer, S. E., Kaus, B. J. P., & Passchier, C. (2017). Development of branching brittle and ductile shear zones: A numerical study. *Geochemistry, Geophysics, Geosystems*, *18*, 2054–2075. https://doi.org/10.1002/2016GC006793

Moresi, L., Dufour, F., & Mühlhaus, H.-B. (2003). A Lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials. *Journal of Computational Physics*, *184*(2), 476–497. https://doi.org/10.1016/S0021-9991(02)00031-1

Moresi, L., Zhong, S., & Gurnis, M. (1996). The accuracy of finite element solutions of Stokes's flow with strongly varying viscosity. *Physics of the Earth and Planetary Interiors*, *97*(1), 83–94. https://doi.org/10.1016/0031-9201(96)03163-9

Nagata, K., Nakatani, M., & Yoshida, S. (2012). A revised rate- and state-dependent friction law obtained by constraining constitutive and evolution laws separately with laboratory data. *Journal of Geophysical Research*, *117*, B02314. https://doi.org/10.1029/2011JB008818

Nakatani, M. (2001). Conceptual and physical clarification of rate and state friction: Frictional sliding as a thermally activated rheology. *Journal of Geophysical Research*, *106*(B7), 13,347–13,380. https://doi.org/10.1029/2000JB900453

Noda, H., & Shimamoto, T. (2012). Transient behavior and stability analyses of halite shear zones with an empirical rate-and-state friction to flow law. *Journal of Structural Geology*, *38*, 234–242. https://doi.org/10.1016/j.jsg.2011.08.012

Palmer, A. C., & Rice, J. R. (1973). The growth of slip surfaces in the progressive failure of over-consolidated clay. *Proceedings of the Royal Society of London. Series A, Mathematical, Physical and Engineering Sciences*, *332*(1591), 527–548. https://doi.org/10.1098/rspa.1973.0040

Peng, Z., & Gomberg, J. (2010). An integrated perspective of the continuum between earthquakes and slow-slip phenomena. *Nature Geoscience*, *3*(9), 599–607.

Perfettini, H., & Ampuero, J.-P. (2008). Dynamics of a velocity strengthening fault region: Implications for slow earthquakes and postseismic slip. *Journal of Geophysical Research*, *113*, B09411. https://doi.org/10.1029/2007JB005398

Rice, J. R. (1993). Spatio-temporal complexity of slip on a Fault. *Journal of Geophysical Research*, *98*(B6), 9885–9907.

Rice, J., Lapusta, N., & Ranjith, K. (2001). Rate and state dependent friction and the stability of sliding between elastically deformable solids. *Journal of the Mechanics and Physics of Solids*, *49*, 1865–1898.

Rice, J., & Ruina, A. (1983). Stability of steady frictional slipping. *Journal of Applied Mechanics*, *50*, 343–349.

Rubin, A. M. (2008). Episodic slow slip events and rate-and-state friction. *Journal of Geophysical Research*, *113*, B11414. https://doi.org/10.1029/2008JB005642

Rubin, A. M., & Ampuero, J. (2005). Earthquake nucleation on (aging) rate and state faults. *Journal of Geophysical Research*, *110*, B11312. https://doi.org/10.1029/2005JB003686

Ruina, A. (1983). Slip instability and state variable friction laws. *Journal of Geophysical Research*, *88*, 10,359–10,370.

Schenk, O., & Gärtner, K. (2004). Solving unsymmetric sparse systems of linear equations with PARDISO. *Future Generation Computer Systems*, *20*(3), 475–487. https://doi.org/10.1016/j.future.2003.07.011

Schenk, O., & Gärtner, K. (2006). On fast factorization pivoting methods for sparse symmetric indefinite systems. *Electronic Transactions on Numerical Analysis*, *23*, 158–179.

Sleep, N. H. (1995). Ductile creep, compaction, and rate and state dependent friction within major fault zones. *Journal of Geophysical Research*, *100*(B7), 13,065–13,080. https://doi.org/10.1029/94JB03340

Sleep, N. H. (1997). Application of a unified rate and state friction theory to the mechanics of fault zones with strain localization. *Journal of Geophysical Research*, *102*(B2), 2875–2895. https://doi.org/10.1029/96JB03410

Sobolev, S. V., & Muldashev, I. A. (2017). Modeling seismic cycles of great megathrust earthquakes across the scales with focus at postseismic phase. *Geochemistry, Geophysics, Geosystems*, *18*, 4387–4408. https://doi.org/10.1002/2017GC007230

Stein, S., & Okal, E. A. (2007). Ultralong period seismic study of the December 2004 Indian Ocean earthquake and implications for regional tectonics and the subduction process. *Bulletin of the Seismological Society of America*, *97*(1A), S279–S295. https://doi.org/10.1785/0120050617

Thielmann, M., May, D. A., & Kaus, B. J. P. (2014). Discretization errors in the hybrid finite element particle-in-cell method. *Pure and Applied Geophysics*, *171*(9), 2165–2184. https://doi.org/10.1007/s00024-014-0808-9

Thomas, M. Y., Lapusta, N., Noda, H., & Avouac, J.-P. (2014). Quasi-dynamic versus fully dynamic simulations of earthquakes and aseismic slip with and without enhanced coseismic weakening. *Journal of Geophysical Research: Solid Earth*, *119*, 1986–2004. https://doi.org/10.1002/2013JB010615

van Dinther, Y., Gerya, T. V., Dalguer, L. A., Corbi, F., Funiciello, F., & Mai, P. M. (2013). The seismic cycle at subduction thrusts: 2. Dynamic implications of geodynamic simulations validated with laboratory models. *Journal of Geophysical Research: Solid Earth*, *118*, 1502–1525. https://doi.org/10.1029/2012JB009479

van Dinther, Y., Gerya, T. V., Dalguer, L. A., Mai, P. M., Morra, G., & Giardini, D. (2013). The seismic cycle at subduction thrusts: Insights from seismo-thermo-mechanical models. *Journal of Geophysical Research: Solid Earth*, *118*, 6183–6202. https://doi.org/10.1002/2013JB010380

van Dinther, Y., Mai, P. M., Dalguer, L. A., & Gerya, T. V. (2014). Modeling the seismic cycle in subduction zones: The role and spatiotemporal occurrence of off-megathrust earthquakes. *Geophysical Research Letters*, *41*, 1194–1201. https://doi.org/10.1002/2013GL058886

Veedu, D. M., & Barbot, S. (2016). The Parkfield tremors reveal slow and fast ruptures on the same asperity. *Nature*, *532*, 361–365.

Vermeer, P. A. (1998). Non-associated plasticity for soils, concrete and rock. In H. J. Herrmann, J.-P. Hovi, & S. Luding (Eds.), *Phys. Dry Granul. Media* (pp. 163 – 196). Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-94-017-2653-5_10

Wang, K. (2007). Elastic and viscoelastic models of crustal deformation in subduction earthquake cycles. In T. Dixon & J. Moore (Eds.), *The seismogenic zone subduction thrust faults* (pp. 540 – 575). New York: Columbia University Press.