Sobolev, S. V., Muldashev, I. (2017): Modeling Seismic Cycles of Great Megathrust Earthquakes Across the Scales With Focus at Postseismic Phase. - *Geochemistry Geophysics Geosystems (G3)*, *18*, 12, pp. 4387—4408.

DOI: http://doi.org/10.1002/2017GC007230

AGU PUBLICATIONS

Geochemistry, Geophysics, Geosystems

RESEARCH ARTICLE
10.1002/2017GC007230

Modeling Seismic Cycles of Great Megathrust Earthquakes Across the Scales With Focus at Postseismic Phase

Stephan V. Sobolev[1,2] and Iskander A. Muldashev[1]

[1]GFZ German Research Center for Geosciences, Potsdam, Germany, [2]Institute of Earth and Environmental Sciences, University of Potsdam, Germany

Key Points:
- We model seismic cycles of great subduction earthquakes in time scales from minute to millennia with nonlinear transient rheology
- Viscosity in the mantle wedge drops by more than three orders of magnitude during a great earthquake with a magnitude above nine
- Pronounced viscoelastic relaxation in mantle wedge starts already hours or days after the great earthquake

Supporting Information:
- Supporting Information S1

Correspondence to:
S. Sobolev,
stephan@gfz-potsdam.de

Citation:
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

Received 12 SEP 2017
Accepted 10 NOV 2017
Accepted article online 15 NOV 2017
Published online 8 DEC 2017

Abstract Subduction is substantially multiscale process where the stresses are built by long-term tectonic motions, modified by sudden jerky deformations during earthquakes, and then restored by following multiple relaxation processes. Here we develop a cross-scale thermomechanical model aimed to simulate the subduction process from 1 min to million years' time scale. The model employs elasticity, nonlinear transient viscous rheology, and rate-and-state friction. It generates spontaneous earthquake sequences and by using an adaptive time step algorithm, recreates the deformation process as observed naturally during the seismic cycle and multiple seismic cycles. The model predicts that viscosity in the mantle wedge drops by more than three orders of magnitude during the great earthquake with a magnitude above 9. As a result, the surface velocities just an hour or day after the earthquake are controlled by viscoelastic relaxation in the several hundred km of mantle landward of the trench and not by the afterslip localized at the fault as is currently believed. Our model replicates centuries-long seismic cycles exhibited by the greatest earthquakes and is consistent with the postseismic surface displacements recorded after the Great Tohoku Earthquake. We demonstrate that there is no contradiction between extremely low mechanical coupling at the subduction megathrust in South Chile inferred from long-term geodynamic models and appearance of the largest earthquakes, like the Great Chile 1960 Earthquake.

1. Introduction

Subduction, being a key process of plate tectonics, is also a driving mechanism of the largest and most dangerous earthquakes (Schubert et al., 2001; Wang et al., 2012). Presently, modeling of the long-term (millions of years' time scale) deformation at plate boundaries (e.g., Baes et al., 2016; Burov et al., 2014; Gerya, 2011; Popov et al., 2012; Quinteros & Sobolev, 2013) and modeling of earthquakes and their seismic cycles (e.g., Barbot et al., 2012; Lambert & Barbot, 2016; Lapusta & Barbot, 2012; Liu & Rice, 2007; Rice, 1993; Rice & Ben-Zion, 1996) are independent. A few models (Sobolev & Babeyko, 2004; van Dinther et al., 2013a, 2013b, 2014) attempted to close the gap between these temporal scales, but these studies were limited to a minimum time resolution of several years. In particular, van Dinther et al. (2013b) complemented a long-term geodynamic model of subduction by the rate-weakening friction law at the subduction plate interface, which allowed simulation of seismic cycles. However, due to technical reasons, "rupture propagation" in that model lasted more than a decade. Obviously, in reality, even the greatest subduction zone earthquakes with magnitude above nine last only a few minutes and are followed by postseismic deformation in the ensuing hours to months and years. In this postseismic phase, whose modeling requires much higher temporal resolution than a year, the crust and mantle are adapting to the deformation imposed by the earthquakes, and the stress system is being prepared for the next great earthquake (Scholz, 1998; Wang et al., 2012).

As satellite-based geodetic observations have become widely available over the last decades, knowledge about the processes that operate during postseismic and interseismic periods has advanced rapidly (Bürgmann & Dresen, 2008; Wang et al., 2012). It is widely accepted that two relaxation processes dominate during the postseismic period (Bürgmann & Dresen, 2008; Wang et al., 2012): (1) a decaying slip at the fault following the main event called afterslip and (2) viscoelastic relaxation in the crust and mantle. Afterslip is localized at the fault, usually at its shallowest and deepest parts (Bürgmann & Dresen, 2008; Wang et al., 2012) and acts like a slow continuation of the earthquake. Contrary to the afterslip, viscoelastic relaxation after great earthquakes with magnitudes over nine involves large volumes of the

mantle and lower crust with dimensions of hundreds to thousands of kilometers (Bürgmann & Dresen, 2008; Wang et al., 2012). An important discovery was that the viscosity of the mantle after an earthquake is time-dependent (Freed & Bürgmann, 2004; Pollitz, 2003; see also reviews by Bürgmann & Dresen, 2008 and by Wang et al., 2012, and references therein) and decreases during the earthquake by up to 10–20 times and then increases back during the postseismic period. These viscosity changes affect the timing of the postseismic relaxation processes. It is widely accepted that viscoelastic relaxation begins to control surface velocities years (Scholz, 1998) or months (Hoechner et al., 2011; Wang et al., 2012) after a great earthquake and before that deformation is driven by afterslip at the fault (Marone et al., 1991; Perfettini and Avouac, 2004).

Simulating earthquake processes is highly challenging because of the high complexity of earthquake phenomenology and the very wide range of space and time scales involved (Ben-Zion, 2008). In particular, a longstanding challenge has been to model earthquakes in a self-consistent way as a part of a seismic cycle (Wang et al., 2012). In addition, most of the recent seismic-cycle models with several exceptions (Freed & Bürgmann, 2004; Freed et al., 2006, 2007, 2010, 2012; Lambert & Barbot, 2016; Masuti et al., 2016; van Dinther et al., 2013b) use simplified linear rheological models, such as Burger's rheology (see review Wang et al., 2012, and references therein), that are not based on mineral-physics laboratory data.

In this study, we intend to develop a two-dimensional (2-D) cross-scale self-consistent model of subduction including earthquakes and multiple seismic cycles. The model aims to fulfill three conditions. First, it should employ nonlinear rheology based on experimental data on deformation of rocks and mineral-physics models. Second, it should replicate long-term (million years' time scale) evolution of subduction zone. Finally and most important, the model should generate earthquakes as spontaneous mechanical instabilities and should replicate all following stages of the seismic cycle over multiple cycles at time scales from minutes (earthquake) to thousands of years (multiple seismic cycles). Note, however, that here we do not aim to model the rupture propagation process and generation of seismic waves during the earthquakes. Including these processes would require much higher spatial and temporal resolution than we could currently achieve in our models.

2. Cross-Scale Model

In this section, we describe the main features of our model, while more details of the modeling techniques and its testing are presented in the Appendix A. The model consists of two stages. At the first stage, we simulate long-term subduction process with the time scale of million years. As a result, we obtain a model with the required geometry of the slab and realistic distribution of stresses and temperature in the lithosphere and underlying mantle. The integration time step at this stage is constant and is typically about 10^4 years. At the next stage, we zoom in at time scales of the seismic cycle and multiple seismic cycles using the long-term model as an initial condition. Note that at this second stage, deformation processes are transient rather than steady state and their time scales are strongly variable, from a minute (earthquake) to years and centuries (seismic cycle) and millennia (multiple seismic cycles). Therefore, modeling of these processes requires modification of the rheological models and implementation of an adaptive time stepping procedure.

2.1. Long-Term Model (Time Scale of Million Years)

We use a 2-D version of the thermomechanical finite element numerical code SLIM3D (Popov & Sobolev, 2008), which solves conservation equations of mass, momentum, and energy. The technique employs nonlinear rheology, assuming that the cumulative strain rate is the sum of elastic strain rate, viscous creep strain rate, and plastic strain rate. The steady state creep parameters for diffusion and dislocation creep are taken from published laboratory experiments with major rock types (see Table A1 in the Appendix A) and all models have a true free-slip upper boundary condition. We use typical thermal parameters for all rock types (see section A1.1 in the Appendix A). The model is 2-D, has a horizontal dimension of 900 km and vertical (depth) dimension of 300 km. The finite elements are isometric and their size in most of the models is 3 km, but we have also computed high-resolution (finite element size 1.5 km) and low-resolution versions (finite element size 6 km) of our reference model.

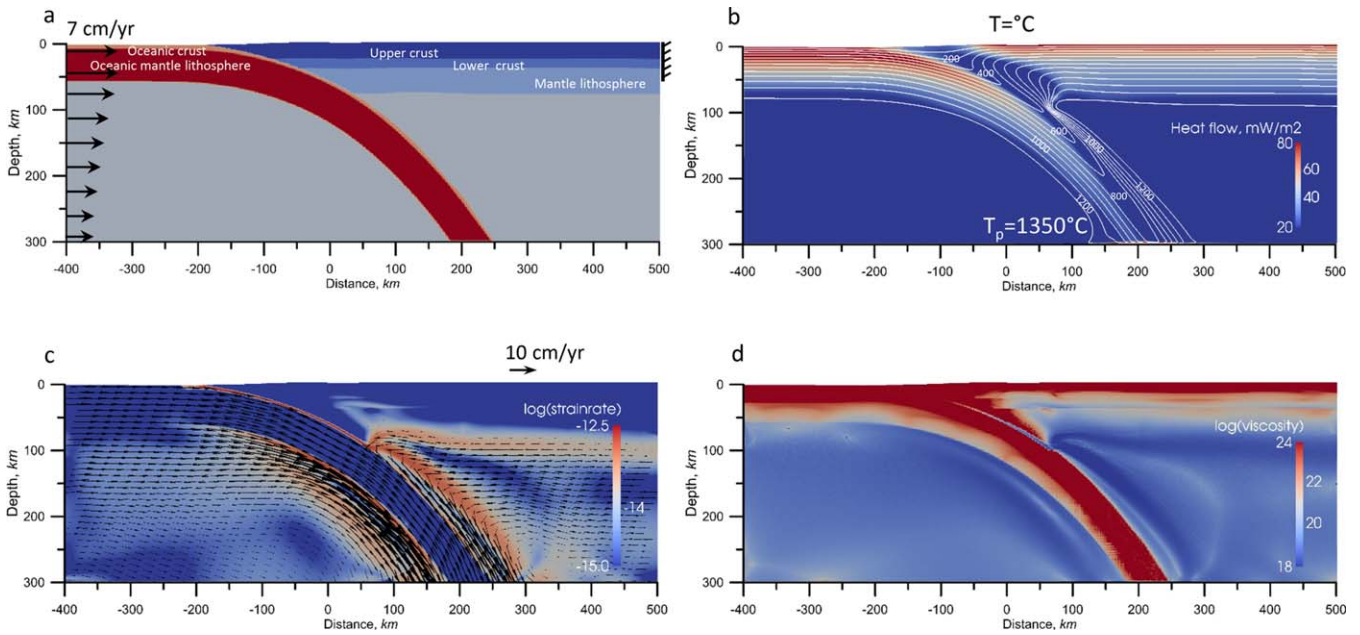

Figure 1. (a)Spatial distribution of materials, (b) magnitude of heat flow and temperatures, (c) strain rates and velocities, and (d) viscosity (d) in a long-term thermomechanical model of subduction in South Chile before implementation of transient viscous rheology and rate-and-state friction.

The model is aimed to reproduce the lithospheric structure of the South Andes where the Great Chile Earthquake occurred in the year 1960. It consists of an oceanic subducting slab with the basaltic crust and harzburgitic lithospheric mantle; overriding plate with two-layer continental crust, and lithospheric mantle and asthenospheric mantle (Figure 1a). For the asthenospheric mantle, we use olivine diffusion and dislocation creep parameters corresponding to the water content of 1,000 ppm H/Si (Hirth & Kohlstedt, 2003). The potential temperature in the asthenosphere is set to 1,350°C. We consider as a "model shear zone" a top layer of the oceanic crust at a depth of less than 100 km and prescribe to it a special type of rheology. This includes a very low effective static friction coefficient of 0.015–0.02 in accord with the long-term evolution models of South Andes (Sobolev & Babeyko, 2005) and the weakest wet quartz dislocation creep rheology (Ranalli, 1997).

We "prepare" the long-term model by running the two-plate model for 10 Myr with the fixed no-slip upper 50 km of the continental lithosphere at right boundary and the rest of the right boundary flow-open. The temperature is 0°C at the surface, and the potential temperature is set to be 1,350°C at the model bottom. At the model sides, the normal component of the heat flow vector is set to be 0. We generate subduction by advancing oceanic lithosphere in the direction of the continental lithosphere with a 17° dipping low-strength fault between oceanic and continental lithospheres. The bottom boundary is flow-open and the left boundary is kinematically prescribed. Horizontal velocities at the upper 50 km of the left boundary (oceanic lithosphere) were fixed at 7 cm/yr. Velocities at the left boundary deeper than 50 km (Figure 1a) are chosen by trial and error to achieve the geometry of the subducted slab consistent with the seismic observations (Dzierma et al., 2012). The resulting long-term model is demonstrated in Figure 1. It has slab geometry, upper plate structure, and temperature consistent with observations and previous thermal models (Currie & Hyndman, 2006). The viscosity is strongly heterogeneous (Figure 1d) with the lowest viscosity slightly lower than 10^{18} Pa s in the asthenospheric mantle wedge.

2.2. Short-Term Model (Time Scale From Minute to n × 1000 Years)

At the second stage, we take the long-term model as an initial setup for the modeling of the seismic cycles of great earthquakes. As side and bottom boundary conditions, we use velocities from the long-term model and the upper boundary is kept free. To model a seismic cycle, we switch to completely different and strongly variable time scales from about a few minutes (duration of the great earthquake) to years and hundreds of years (duration of postseismic and interseismic periods). To accomplish that, we modify the rheology of the crust and mantle to consider transient creep processes following ideas of Karato (1998) and

implement classical aging rate-and-state friction (hereafter RSF) weakening rheological model (Ruina, 1983; Scholz, 1998) in the model shear zone of finite thickness, which is considered to be a continuum, The slip velocity is defined through shear strain rate in the shear zone (see section 2.2.2). We also implement special procedures to model earthquakes and postseismic processes, which include an adaptive time stepping procedure and procedure of identification of mechanical instabilities (earthquakes). We test our modeling technique by comparing modeling results with the previous high-resolution numerical models of seismic cycles (Lapusta et al., 2000).

2.2.1. Rheology Modification: Transient Creep

For long-term tectonic deformation, it is suitable to use steady-creep rheology with parameters provided by laboratory experiments, like those for the olivine-dominated rocks (Hirth & Kohlstedt, 2003), see also references in the supporting information Table. In our long-term geodynamic models, we routinely use both (competing) diffusion creep and dislocation creep steady rheologies but in the active subduction zone setting and at the depth range of less than 300 km, the dominant steady-creep process is dislocation creep. Whether dislocation density will change during the postseismic period depends on the magnitude of the stress increase imposed by the elastic strain during the earthquake. When the stress increase is significant (say a doubling of the square root of the second invariant of the stress deviator) then dislocation multiplication occurs (S. Karato, personal communication, 2016). Sudden changes of the stress field during the earthquakes lead to the transient rheological behavior. We use a model for transient dislocation creep rheology based on the ideas of Karato (1998).

$$\dot{\varepsilon}=\dot{\varepsilon}_{ss}\left(1+(\beta-1)\exp\left(-\frac{\varepsilon_{ss}(t)}{\varepsilon_{el}^{eq}}\right)\right), \tag{1}$$

$$\dot{\varepsilon}_{ss}=B_{dis}(\tau_{II})^{n}\exp\left(-\frac{H_{dis}}{RT}\right), \tag{2}$$

where $\dot{\varepsilon}$ is the square root of the second invariant of the power law creep strain rate tensor; $\dot{\varepsilon}_{ss}$ is the same for steady state power law creep; ε_{el}^{eq} is the square root of the second invariant of the elastic strain tensor induced by an earthquake; $\varepsilon_{ss}(t)$ is an accumulated steady state creep strain at time t after the earthquake (i.e., time integral of $\dot{\varepsilon}_{ss}$ since the earthquake) and it is 0 at $t = 0$, i.e., at a moment of an earthquake; β is the ratio between initial transient and steady state strain rates, which is a constant about 10 for olivine (Bai et al., 1991); τ_{II} is the square root of the second invariant of the stress deviator; n is the power law exponent; H_{dis} is an activation enthalpy of the dislocation creep; R is the gas constant; T is the absolute temperature, and B is a preexponential factor. Parameters B, n, and H_{dis} for olivine are taken from experimental data (Hirth & Kohlstedt, 2003).

Equation (1) has a more general form than the original equation (15) in Karato (1998) as it does not assume linear dependency of the steady state creep strain on time. If this dependency is linear, i.e., $\varepsilon_{ss}(t)=\bar{\dot{\varepsilon}}\cdot t$ (where $\bar{\dot{\varepsilon}}$ is an average steady state strain rate) then equation (1) reduces to the form of equation (15) in Karato (1998) also used in Freed et al. (2012) (equation (4)), where relaxation time $\tau=\varepsilon_{el}^{eq}/\bar{\dot{\varepsilon}}$. The advantage of our more general formulation is that it contains only one parameter (apart from the steady state rheological parameters), which is β and all other quantities are calculated.

Equation (1) is applied to the finite elements, where the square root of the second invariant of the stress deviator is changing by more than 2 times at one time step. Because of this condition, the transient deformation in our models is almost completely restricted to the asthenospheric mantle wedge, where viscosity and long-term stresses are low. Note that the quantities $\varepsilon_{ss}(t)$ and ε_{el}^{eq} are recalculated at every sudden significant change of stress (i.e., earthquake), which allows cycles of transient creep, similar to Masuti et al. (2016). However, differently to Masuti et al. (2016), our transient creep model is defined for a single event. That means that the model can be applied to the next event only if the steady state creep regime has already been achieved after the previous event. The typical time for that is less than a few years (Karato, 1998). In our models of great earthquakes, this time is even much smaller, being less than one month (see section A1.5). Therefore, our transient creep model is fully applicable for the seismic cycles with the time intervals between the events of a few hundred years.

2.2.2. Rheology Modification: Rate-and-State Friction

In the model shear zone, we employ classic aging rate-and-state friction by Ruina (1983), (see also the review by Scholz, 1998),

$$\tau = c + \mu\left(1 - \frac{P_f}{\sigma_n}\right)\sigma_n, \tag{3}$$

$$\mu\left(1 - \frac{P_f}{\sigma_n}\right) = \mu^* = \mu^*_{st} + a^* \cdot \ln\left(\frac{V}{V_{st}}\right) + b^* \cdot \ln\left(\frac{\Theta V_{st}}{L}\right), \tag{4}$$

$$\frac{d\Theta}{dt} = 1 - \frac{\Theta V}{L}, \tag{5}$$

$$\mu^*_{st} = \mu_{st}\left(1 - \frac{P_f}{\sigma_n}\right);\ a^* = a\left(1 - \frac{P_f}{\sigma_n}\right);\ b^* = b\left(1 - \frac{P_f}{\sigma_n}\right), \tag{6}$$

where τ is the shear stress, c is the cohesion, P_f is the fluid pressure, σ_n is the normal stress, μ, μ^*, and μ^*_{st} are friction, effective friction, and effective static friction, respectively. a, b are parameters for rate-and-state friction law and a^*, b^* are effective parameters for rate-and-state friction law, V is the velocity of slip on the fault, V_{st} is quasi-static (or reference) velocity, i.e., the velocity at which $\mu^* = \mu^*_{st}$ and L is the critical slip distance and Θ is the state variable.

Note that in the original version of equation (3) of RSF law cohesion is absent. In our model, we use a very small value of cohesion of 2 MPa in the model shear zone to avoid destabilization of the uppermost elements in the model. Our tests show that changing this parameter by 2 times in any direction does not affect the modeling results.

Assuming that velocities are constant at any integration time step, we can integrate state equation (5) analytically and define an updated value of the state parameter at the end of the time step (Θ_{n+1}) through its value at the end of the previous time step (Θ_n).

$$\Theta_{n+1} = \frac{L}{V_{n+1}} + \left(\Theta_n - \frac{L}{V_{n+1}}\right) \cdot \exp\left(-\frac{V_{n+1}\Delta t}{L}\right), \tag{7}$$

where Θ_{n+1} and Θ_n are state variables for current and previous time steps, respectively, L is the critical slip distance, V_{n+1} is the velocity of slip on the fault, and Δt is the current time step. Note that equation (7) is not a finite difference approximation of state equation (5) but its exact solution assuming constant velocity at a single time step.

We then approximate velocity V as a product $2\Delta_{el}\dot{\gamma}$, where $\dot{\gamma}$ is the plastic strain rate of material in the model shear zone and Δ_{el} is the finite element size, which is the same in both directions in our model. We also approximate normal stress σ_n and shear stress τ by dynamic pressure (–I/3, where I is the first invariant of the stress tensor) and by the square root of the second invariant of the stress deviator, respectively. All these approximations are the same as were used in our earlier models (Sobolev & Babeyko, 2004) and in models by van Dinther et al. (2013b). Further, by substituting (7) in (4) and (3) we obtain the final relation for the effective friction and shear stress in the channel.

2.2.3. Adaptive Time Step Procedure

The maximum time step in our short-term model is 5 years, which is suitable at the late stage of the postseismic and for the interseismic periods, but is much too long for the earthquake itself and following the early stage of the postseismic period. See the Appendix A for a criterion to define maximum time step. To resolve different stages of the postseismic period, we implement the following adaptive time step procedure. At each time step, we calculate an average plastic strain rate in the entire model shear zone. If its value increases by more than a certain critical amount CA, we decrease the time step, otherwise the time step does not change. The CA value is defined to detect the largest earthquakes. Empirically found suitable CA values are 25–150%. If CA is lower, the algorithm starts to detect many smaller earthquakes together with the large ones. If CA is higher than some 150–200%, we start to miss large events. The procedure works as follows. If, at one of the time steps, the average strain rate in the channel increases by more than (1 + CA/100) times, the time step is then decreased by 2 times and the calculation is repeated. If required, the stepwise decreasing continues till the time step reaches its smallest value of 40 s, which happens at the earthquake. The smallest time step (40 s) in our 2-D model corresponds to the rise time for a great earthquake. After the earthquake, the time step begins to increase stepwise (by 2 times) following decreasing strain rate in the model shear zone until it reaches its maximal value of 5 years. In this way, we can follow in detail rapidly changing deformation during the entire seismic cycle. Evolution of the time step is shown in Figures 2a and 2b.

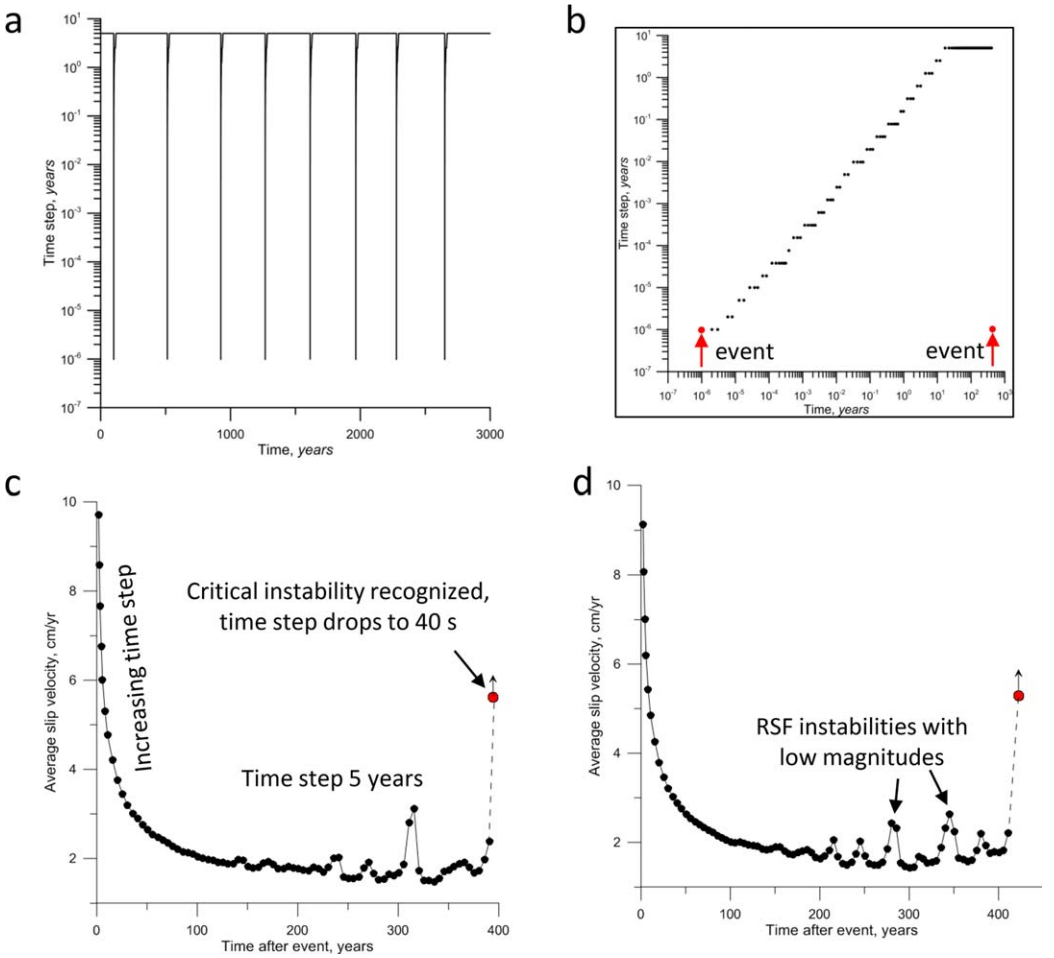

Figure 2. (a) Evolution of modeling time step with time for multiple seismic cycles and (b) single seismic cycle. (c and d) Average slip velocity at the entire fault versus time for two randomly chosen events. Red points show velocities at the moment when critical instabilities are recognized. The model is then recalculated with a smaller time step.

Note that our adaptive time stepping procedure is not suitable to model in detail rupture nucleation and rupture propagation processes. The more appropriate adaptive time stepping algorithm is that by Lapusta et al. (2000). However, this algorithm requires very high spatial and temporal resolution that cannot be achieved currently in our large-scale subduction models. Therefore, we employ an approximate adaptive time stepping algorithm, which nevertheless allows proper modeling of the largest events and their seismic cycles (see below).

2.2.4. Modeling of Earthquakes

In the previous modeling studies (Lapusta & Barbot, 2012; Lapusta et al., 2000), it was demonstrated that to resolve properly processes of nucleation and propagation of rupture, the rupture nucleation length h* must be resolved by at least a few tens of grid points. With our element size of 1.5–3 km and RSF parameter L = 1–10 cm, h*, defined as in Lapusta et al. (2000), is about 2–20 km, so we have only 0.7–14 grid points at nucleation length, which is much less than required. From this consideration, it is clear that it makes no sense to model details of rupture nucleation and propagation with the current model resolution. However, for our purposes, precise modeling of rupture propagation is not required. Instead, what is necessary for modeling the entire seismic cycle is (1) to recognize the onset of the largest instability (= great earthquake) and (2) to reproduce reasonably well the cumulative coseismic slip distribution along the fault. To do so, we develop a modeling procedure that achieves our aims with a resolution of about a few grid points at h* length.

Our modeling procedure is based on the following idea. During an earthquake, the local stress in the channel drops. This is accompanied by a dramatic decrease of slip velocity. During the interseismic phase, stress

builds up but slip velocity in the "locked zone" remains orders of magnitude lower than the convergence rate until critical instability (= earthquake) occurs. While approaching the instability, the average slip velocity at the entire model shear zone rises and begins to fluctuate (Figures 2c and 2d) if the maximum time step is small enough (see analyses in the Appendix A). These fluctuations result from propagation upward and back of the tip of the deep creeping part of the fault. Some of them may be real while others may be numerical artifacts. Anyway, they do indicate nearing critical instability; the closer to the instability the larger are the fluctuations (Figures 2c and 2d). By constantly monitoring strain rate changes at the fault (see section above), we recognize a moment when the slip velocity increases fast enough (according to the CA criterion) and drop our time step, till the smallest time step (here 37 s) is achieved (see section 2.2.3). Our rupture, i.e., rapid shear deformation in the continuum model shear zone, typically takes one or two of the smallest time steps; after that, we gradually increase the time step to simulate postseismic processes. Note that this algorithm is aimed to model the largest events, so we may ignore the small ones.

In a special section in the Appendix A, we present results of the test of our modeling procedure using existing high-resolution solutions for the seismic cycles at transform fault by Lapusta et al. (2000). We demonstrate that our model generates regular sequences of events without the artificial complexity that may arise in models of rupture propagation due to insufficient resolution (Lapusta et al., 2000; Rice, 1993; Rice & Ben-Zion, 1996). We reproduce major features of the solutions by Lapusta et al. (2000), including recurrence times and magnitudes of main events as well as coseismic slip distributions, although we do not reproduce some details of the solution, like small secondary events with seismic moments of a few percent of the main events.

3. Results

3.1. Earthquake Sequences

First, we consider the case with the depth-uniform (till 100 km depth) rate-weakening RSF parameters. We calibrate parameters to reproduce an average slip (about 15–20 m, Moreno et al., 2009) and average stress drop (about 5 MPa, Seno, 2014) of the Great Chile Earthquake of the year 1960 (Figure 3). Interestingly, we simultaneously also reproduce reasonably well recurrence time of the Chile 1960-like events estimated to be about 400–500 years (Cisternas, 2005). At the values of the RSF parameters $(a - b)^* = -4 \times 10^{-4}$, $b/(a - b) = -5$ to 3 and 1 cm $\leq L \leq$ 10 cm, our model generates spontaneous instabilities leading to the stick-slip deformation process with average recurrence times of earthquakes of 350–430 years, average static stress drops of 4–6 MPa and average 2-D seismic moments of $1.5–1.8 \times 10^{17}$ N (Figure 3a). The corresponding 3-D moment magnitudes (Mw = 9.3–9.35 assuming rupture length of 800–900 km) are close to the geodesy-based estimates for the Chile 1960 event (Cisternas, 2005; Moreno et al., 2009). As we discuss in more detail in section 4.3, the $(a - b)^*$ value of -4×10^{-4} gives $(a - b)$ of about -0.01 after correction for the effect of fluid pressure. This value by magnitude is at the upper range of the experimental data and is about 2–3 times larger than expected average values (Blanpied et al., 1995; den Hartog et al., 2012),

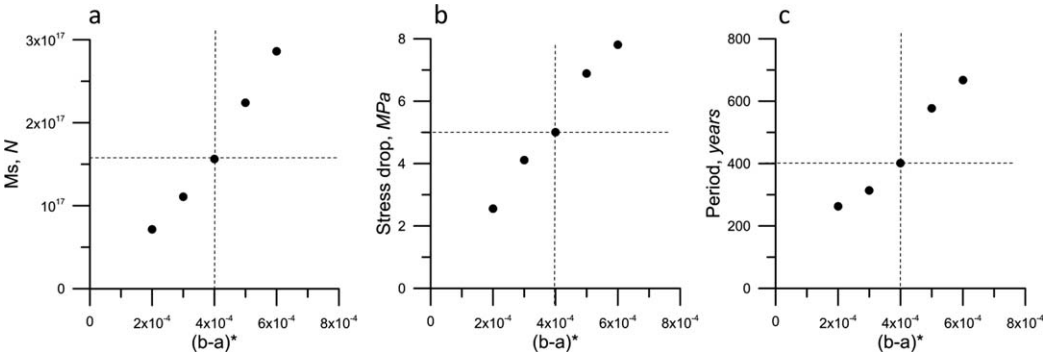

Figure 3. Mean (a) 2-D seismic moment, (b) stress drop, and (c) recurrence time (period) of the earthquakes generated by the reference nonlinear transient model with different RSF parameter $(a - b)^*$. Horizontal dashed lines correspond to the estimated values of stress drop of 5 MPa (Seno, 2014) and recurrence time of 400 years (Cisternas, 2005) for the Great Chile 1960 Earthquake. Vertical lines correspond to appropriate values of RS $(a - b)^*$ parameter.

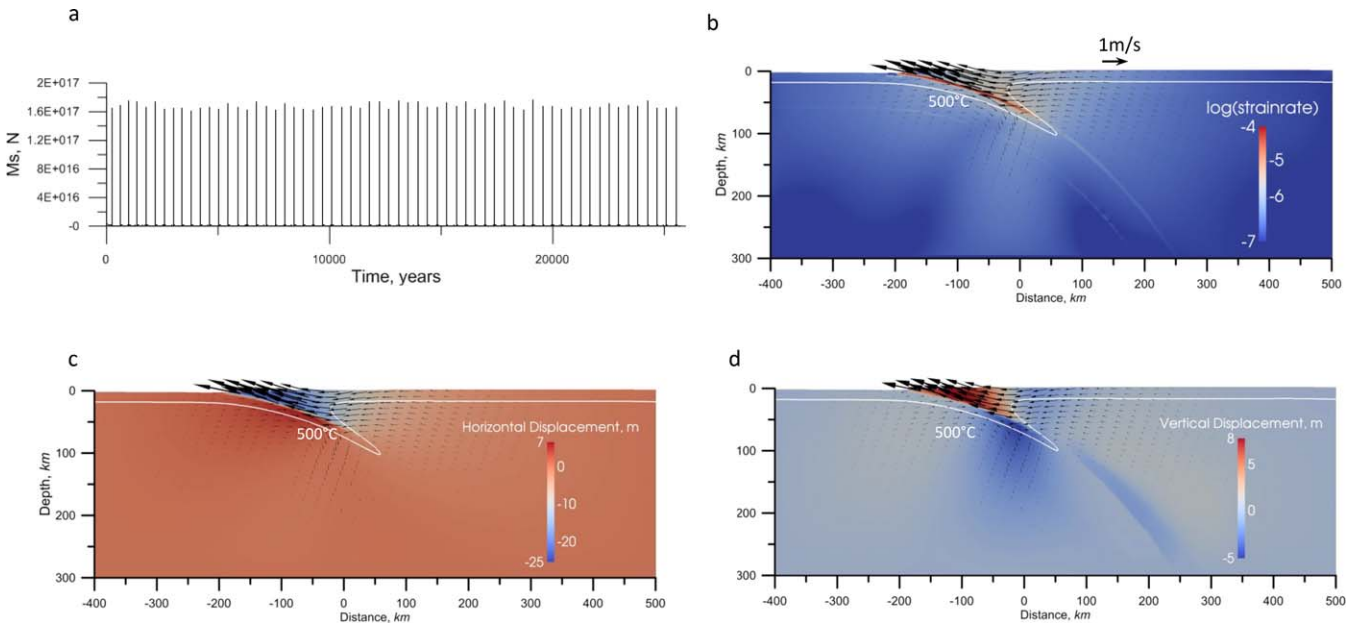

Figure 4. (a) Sequence of the earthquakes generated in the high-resolution version (finite element size of 1.5 km) of reference nonlinear transient model with the RSF L parameter of 1 cm. (b) A snapshot of strain rate spatial distribution (with velocity vectors) for the typical event. The white curve shows 500°C temperature isoline. (c and d) Snapshots of the (c) horizontal and (d) vertical coseismic slip distributions during the same event.

suggesting that other friction-weakening mechanisms are involved during great earthquakes, in accord with previous suggestions (e.g., Lapusta & Barbot, 2012; Rice, 2006) (see section 4.3 for details).

The distribution of velocities, as well as horizontal and vertical displacements during a typical modeled earthquake, are shown in Figures 4b–4d. Interestingly, although in this model we do not set the depth limit for the rate-weakening RSF regime explicitly, the largest ruptures do not penetrate deeper than the depths where the temperature is higher than 450–500°C (Figures 4c and 4d), in agreement with the previous modeling results (van Dinther et al., 2013b). This occurs because of the resistance of the ductile flow in the shear zone below the brittle–ductile transition.

The analyses of the sensitivity of the models to changes in the spatial resolution and technical CA parameter (see Appendix A and supporting information Figures S2 and S3) demonstrate that major features of solutions, like coseismic slip distribution and earthquake recurrence times, are stable and robust.

3.2. Postseismic Deformations

The model shows realistic postseismic deformation patterns (Figure 5) that have been observed at the surface and extend the flow patterns to depth showing patterns that otherwise would be unknown. For example, the model reproduces the landward motion of the near-trench region of the upper plate that begins early after the earthquake (Figure 5). Thanks to the unique ocean bottom geodetic instruments network, this phenomenon was observed and modeled for the first time for the Great Tohoku 2011 Earthquake (Sun et al., 2014; Watanabe et al., 2014). The model also replicates observations that the locking process of the seismogenic zone starts early in the seismic cycle but the trenchward motion of the upper plate continues several decades after the great earthquake (Wang et al., 2012) (Figures 5e and 5f). In particular, our Chile 1960 model demonstrates prolonged postseismic deformations in agreement with the GPS observations (Khazaradze et al., 2002) suggesting trenchward motion of the overriding plate 300–400 km landward from the trench 35 years after the event.

However, perhaps the most striking result of the model is the dramatic drop in the viscosity in the large domain of the mantle wedge just after the onset of the earthquake. While the minimum viscosity in the mantle wedge during the interseismic period in our reference model is approximately 10^{18} Pa s, it drops to about 10^{14} Pa s after the earthquake. This decrease is a result of the stress imposed on the mantle wedge due to the elastic deformation of the earthquake. Stress is increased by more than 10 times in the large

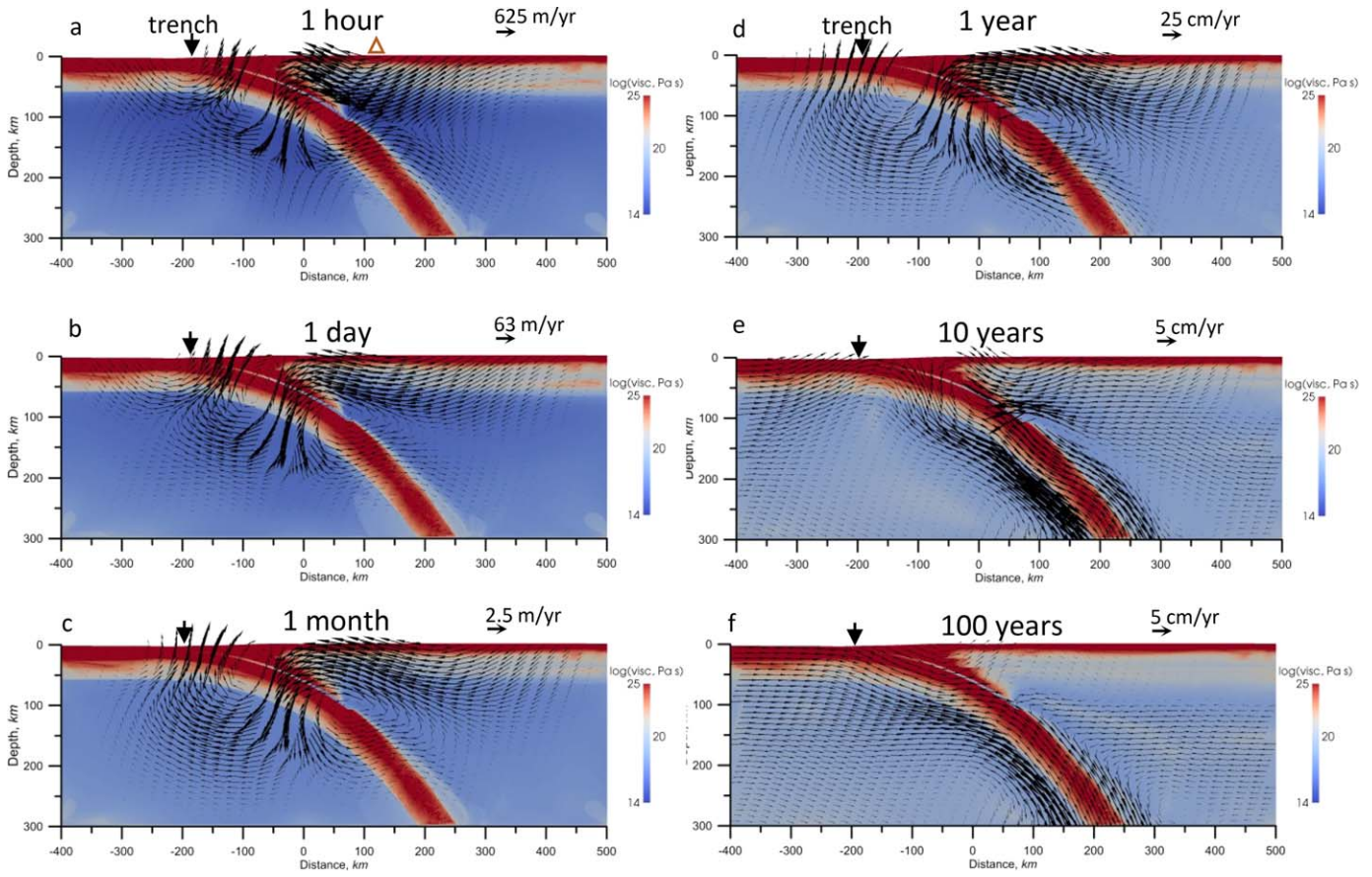

Figure 5. (a)–(f) Snapshots of the spatial distributions of viscosity (background colors) and velocities (vectors) for the different stages of the seismic cycle from 1 h (a) to century (f) of the typical earthquake (Mw 9.3) generated by the high-resolution version of the reference nonlinear transient model. Note different scales of the velocity vectors. The red triangle at the surface indicates the position of the virtual GPS station located about 300 km landward from the trench.

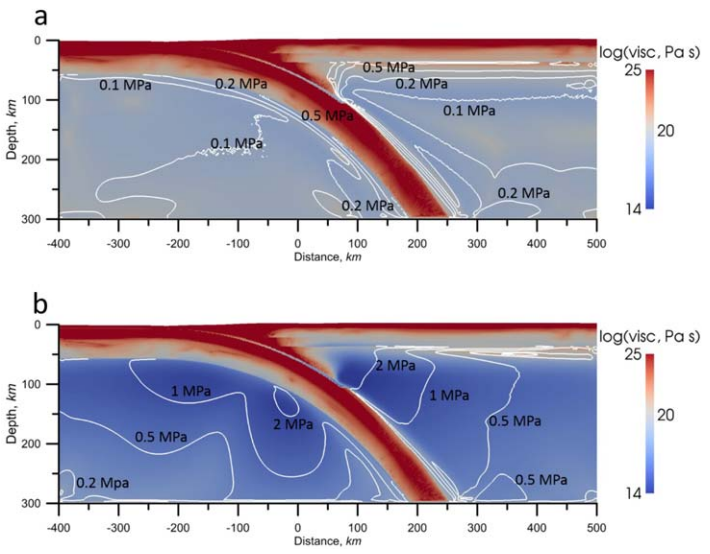

Figure 6. (a and b) Snapshots of the spatial distribution of viscosities (background colors) and stresses (contours) in the high-resolution version of reference nonlinear transient mode of the Mw 9.3 earthquake (a) just before and (b) just after the earthquake.

volume of the mantle wedge (Figures 6a and 6b), which translates to the viscosity drop of about three orders of magnitude because of the power law rheology of the mantle rocks. The additional order of magnitude decrease is due to the transient creep strain rate amplification. Later, during the postseismic phase of the seismic cycle, viscosities gradually increase following the relaxation of stress in the wedge, approaching their steady (interseismic) values about a decade after the earthquake (Figures 5 and 7). For a smaller earthquake (Mw about 8.9), the viscosity drop is less because of the lower stress increase, but it is still as large as 2.5 orders of magnitude (Figure 7). Because the viscosity drop is caused by the change of the stress, the key rheological parameter that controls this behavior for the earthquakes with the same magnitude is the stress exponent n, which is between 3 and 4 for the mantle rocks (Hirth & Kohlstedt, 2003; Karato & Jung, 2003). Viscosity in the mantle wedge changes during the seismic cycle somewhat differently in the models with n = 3 and n = 4 (Figure 7), but in all cases it drops by orders of magnitude during the great earthquake. As we will see from the following analysis, this dramatic drop in viscosity results in a much earlier emergence of the large-scale viscoelastic relaxation process in the mantle wedge than is currently believed.

In conventional models of viscoelastic relaxation, viscosity in the mantle wedge is assumed to be either constant (linear Maxwell viscous

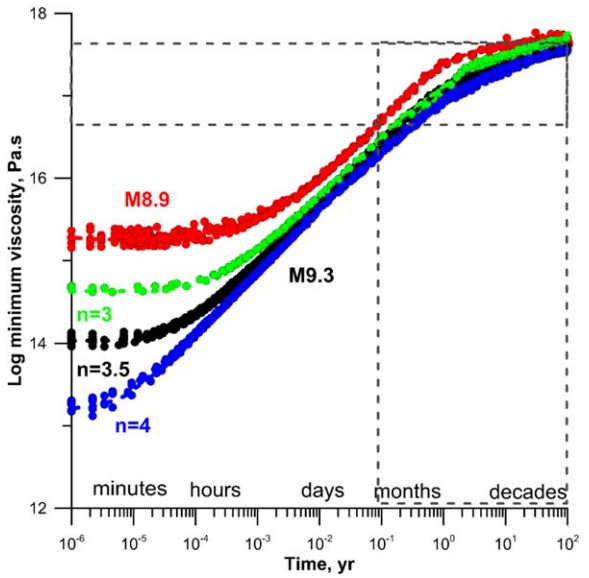

Figure 7. Time evolution of the minimum viscosity in the mantle wedge during the seismic cycles of earthquakes generated by the nonlinear transient models: Mw 9.3 reference model (black symbols); versions of Mw 9.3 reference model with stress exponent n = 3 (green symbols) and n = 4 (blue symbols); Mw 8.9 reference model (red symbols).

model) or changing by up to 10–20 times according to a linear transient viscous (Burger's) model (Wang et al., 2012, and references therein). To compare results of our model with the conventional models, we run models with the accordingly simplified rheology in the mantle but with all other parameters being the same as in our reference model (see Appendix A for details). Similar to our reference model, these models with simpler rheology generate sequences of earthquakes with comparable seismic moments, stress drops, and recurrence times. We compare horizontal velocities for the several models at a virtual GPS station located at a distance of 300 km landward from the trench (Figure 8). Velocities in the reference model (black circles in Figure 8a) are controlled by the continuing rupture propagation, and afterslip in the shallow part of the fault during the first 10 min after the main earthquake shock. However, after just an hour following the earthquake and later, surface velocities are controlled by viscoelastic relaxation in the mantle wedge (see also Figures 5b–5f).

Interestingly, in this model, velocities decay with time as $t^{-0.8} - t^{-1}$ in a large time interval, from an hour until years after an earthquake. Such a decaying behavior is similar to, and could be easily mixed with, the hyperbolic-decay law typical for the fault-controlled afterslip process (Marone et al., 1991; Perfettini & Avouac, 2004).

In the model with linear transient rheology (purple circles in Figure 8a), surface velocities begin to decay due to viscoelastic relaxation in the mantle wedge much later (after a few months) in the postseismic phase than in our reference nonlinear model. That happens even later (a few years after the earthquake) in the model with the constant viscosity in the mantle wedge of 3×10^{18} Pa s (green circles in Figure 8a).

So, the clear difference of our nonlinear model from the conventional linear rheological models is a much earlier emergence of the viscoelastic relaxation process. That is true also for other nonlinear transient models (Figure 8b) with different values of rheological parameters (see section A1.5 of the Appendix A, for a description of the models), as well as for the nonlinear model without transient rheology (blue circles in Figure 8a).

3.3. Models With a Deep Rate-Strengthening Region

All models that we have so-far presented do not contain a rate-strengthening RSF domain, which may exist at megathrust faults deeper than 40–45 km (Scholz, 1998) and which is believed to be responsible for the afterslip at the fault (Marone et al., 1991; Perfettini & Avouac, 2004; Scholz, 1998; Wang et al., 2012). To test its effect, we have added the rate-strengthening RSF domain in the deep part (deeper than 42 km) of the fault in our reference model without changing any other parameters. The resulting model we hereafter call the afterslip reference model. We select values for the rate-strengthening parameters based on estimates by Barbot et al. (2009) that $(a-b)\left(1-\frac{P_f}{\sigma_n}\right)\sigma_n$ should be about 0.7 MPa. Because σ_n in the deep section of the model shear zone is around 1.5–1.7 GPa, this gives us a value of the $(a-b)^* = (a-b)\left(1-\frac{P_f}{\sigma_n}\right)$ parameter of 4×10^{-4} (see Table A2 in the Appendix A). The a^* parameter was taken to be 1.2×10^{-3} and 8×10^{-4}, but we found no significant dependence of our results from the a^* parameter if the $(a-b)^*$ parameter is fixed.

In Figure 8c, we compare postseismic horizontal velocities at a virtual GPS station on the surface (300 km landward from the trench) in the reference model (blue circles) with velocities in the afterslip reference model (red circles). For this comparison, we have chosen events with identical seismic moments. The models are clearly different during the first few tens of minutes following the earthquake. However, after an hour and later, models with and without classical afterslip are becoming almost identical. This demonstrates that even if the classical afterslip process is acting at the fault, the viscoelastic relaxation in the mantle wedge with transient power law rheology becomes a dominant deformation process already an hour or so after an Mw 9.2 earthquake. The same analysis for a smaller earthquake (Mw 8.9) shows that in this case,

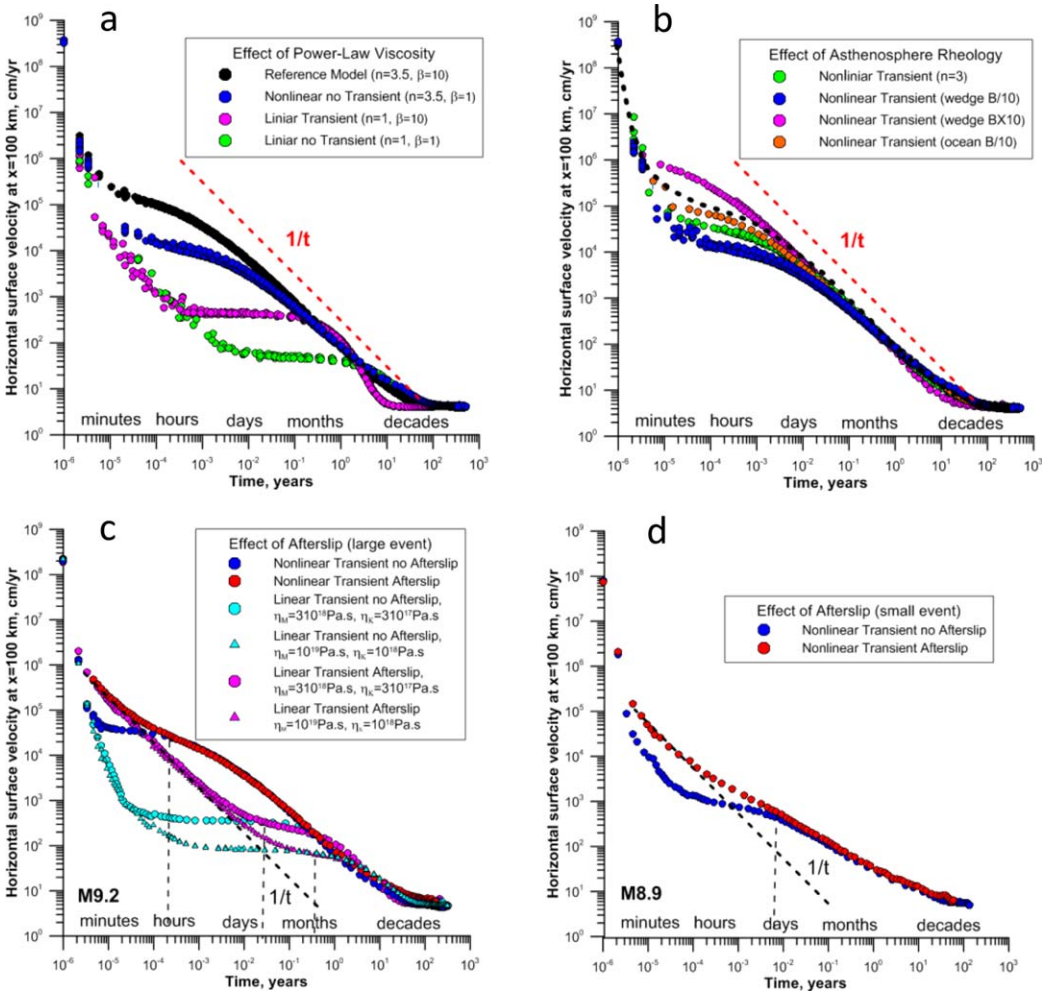

Figure 8. (a) Time evolution of the horizontal surface velocities at the virtual GPS station located 300 km landward from the trench during seismic cycles of the earthquakes generated by different models. Black solid circles correspond to the reference nonlinear transient model, blue circles correspond to nonlinear "nontransient" model, magenta solid circles correspond to the linear transient model and green solid circles to the linear "nontransient" model. Note the log-log scale of the axes. Red dashed line shows 1/t function trend. (b) The same as (a) for the additional nonlinear transient models: model with stress exponent n = 3 (green solid circles), model with 10 times reduced B parameter in the entire asthenosphere (blue solid circles), model with 10 times increased B parameter in the entire asthenosphere (magenta solid circles), model with 10 times reduced B parameter in the oceanic asthenosphere only (orange solid circles), (c) the same as (a) for the model's pairs consisting of model including and model not including rate-strengthening part of the fault with $(a - b)^* = 4 \times 10^{-4}$ below the depth of 42 km. All models have the same parameters (as in reference model) in the rate-weakening part of the fault above the depth of 42 km with parameters. (d) The same as (c) but for the smaller magnitude Mw 8.9 model.

viscoelastic relaxation in the mantle wedge becomes a dominant deformation process later, but still as early as a few days after the earthquake (Figure 8d). In Figure 8c, we also show the effect of afterslip for the "conventional" models with different transient linear viscosities in the mantle wedge (see the legend in the figure or section A1.6 for details). In the linear models, viscoelastic relaxation in the wedge becomes a dominant deformation process only some weeks or some months after the great event.

Surface velocities in the models with and without rate-strengthening part of the fault are very similar at a distance of 300 km (Figure 8c) and 200 km (Figure 9a) from the trench. Closer to the trench (100 km, Figure 9b), surface velocities change sign (upper plate begins to move landward faster than oceanic plate) some time after the event. This happens when deformation due to the viscoelastic relaxation in the mantle wedge overcomes deformation due to the afterslip (Sun et al., 2014). In the "conventional" linear afterslip model, that occurs a week after the Mw 9.2 event (magenta triangles in Figure 9b), while in the nonlinear afterslip model only an hour after the event (red circles in Figure 9b).

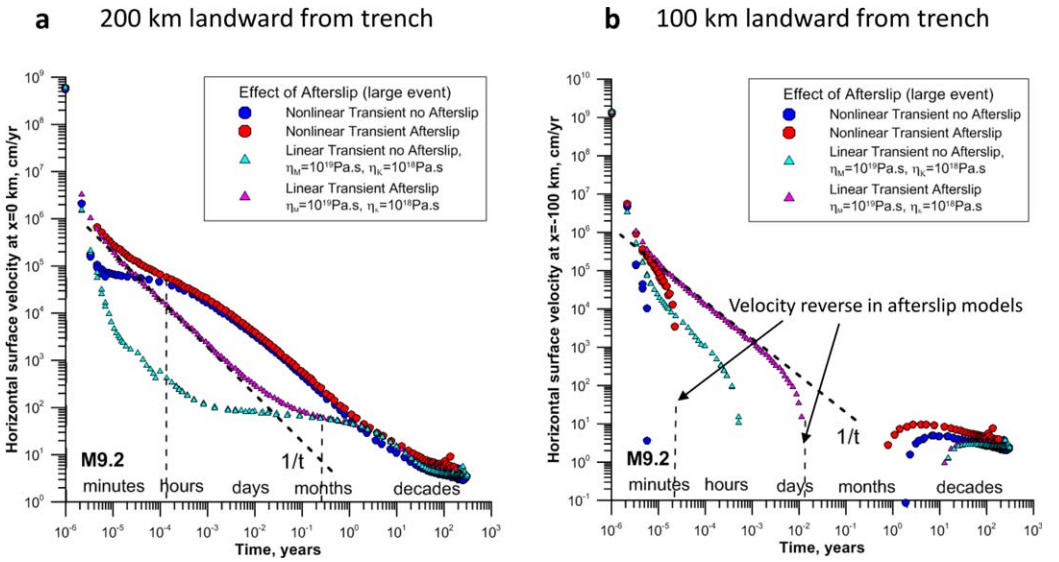

Figure 9. (a) Horizontal surface velocities relative to the ocean (positive oceanward) at the location of 200 km landward from the trench for the models with and without rate-strengthening part of the fault with $(a - b)^* = 4 \times 10^{-4}$ below the depth of 42 km. (b) The same as (a) at the location of 100 km landward from the trench. Note that points in (b) are not shown when the upper plate is moving landward faster than the oceanic plate.

4. Discussion

4.1. Comparison With Conventional Models

Our results suggesting extensive viscoelastic relaxation deformation very early after a great earthquake contradict the conventional view that viscoelastic relaxation begins to control surface velocities only months or even years after an earthquake (see reviews by Bürgmann and Dresen, 2008, Scholz, 1998, and Wang et al., 2012). If our model is correct, why then can conventional models successfully fit the observations? It is clear from Figure 8 that surface velocities in the region above the central and deep parts of a fault during the first few months or even a year in our reference model can be easily misinterpreted as resulting from afterslip because of the similar decay behavior for both processes. Note that in our models an afterslip is a dominant deformation mechanism only during 10 min to an hour in the Mw 9.2 model at any distance from the trench (Figures 8c and 9) and only during the first few days in the Mw 8.9 model (Figure 8d). Note also that from a month or so after the great earthquake till the end of the seismic cycle, the viscosity in the mantle wedge in our models changes by about 10–20 times (Figure 7), which is remarkably close to the viscosity change required to fit observations in the conventional models of viscoelastic relaxation. We emphasize that this 10–20 times change of viscosity during the postseismic period, which is frequently used to fit GPS data (see reviews by Bürgmann and Dresen, 2008 and Wang et al., 2012), is not assumed but is predicted by our model.

4.2. Model Test for Tohoku 2011 Event

As a further test for our model, we compare its predictions with the data for the Great Tohoku earthquake. Note that slab geometry at Tohoku is very similar to South Chile (see Schellart & Rawlinson, 2013). Therefore, we apply the same slab geometry to model the Tohoku event. As our model is 2-D, we select the data from the stations located around the line perpendicular to the trench and crossing the patch of high slip of the Tohoku event (Figure 10a) similar to Muto et al. (2016). For each station shown by a small blue point in Figure 10a, we plot daily solutions for horizontal trench-perpendicular displacements relative to its location at the first day after the Tohoku event. We use all daily solutions for the time interval of 4 years after the event (The Geospatial Information Authority of Japan, http://mekira.gsi.go.jp/project/f3_10_5/en/index.html).

The size of the main asperity in the Tohoku event (around 200 km) is unusually small for its magnitude. Therefore, there is a significant 3-D geometric effect in coseismic deformation at the land GPS stations that are all located more than 200 km away from the trench. To minimize 3-D geometrical effects, we normalize

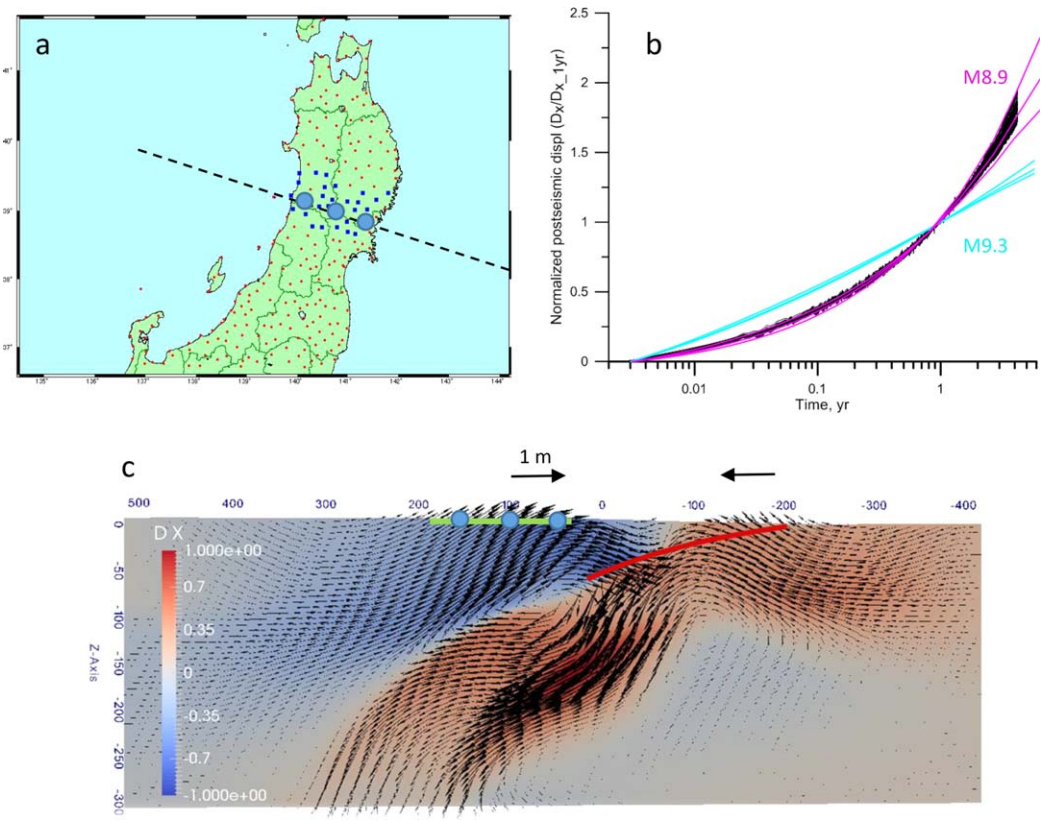

Figure 10. (a) Location of the GPS stations (small blue points) used for the comparison of the displacements after the Tohoku 2011 earthquake with the models. The dashed line shows the surface trace of the model 2-D profile and large blue circles show virtual GPS stations on the model profile. (b) Time evolution of the normalized trench-perpendicular observed postseismic horizontal displacements (black curves) versus calculated horizontal displacements at three virtual GPS stations for the Mw 9.3 reference model (blue curves) and Mw 8.9 model (magenta curves). All displacements are trench-perpendicular, and are calculated relative to the position of the stations at the next day after the Tohoku 2011 event and are normalized by the displacements at one year after the event. *Note logarithmic scale of the time axis*. (c) Spatial distribution of the horizontal displacements (background colors) and displacement vectors (arrows) 2.5 years after the event in the Mw 8.9 model. The red line shows a subduction fault.

the postseismic displacements for each station by its cumulative postseismic displacement during one year after the event. Normalized displacement curves for all stations plot in a rather narrow band and are shown by black curves in Figure 10b. Note the logarithmic time scale in Figure 10b that highlights displacements at small times. All displacement curves show significantly higher velocities shortly after the event: more than one-third of the yearly displacement occurs during the first month after the event. Blue curves show the displacements at three virtual GPS stations (large blue circles in Figure 10a) predicted by our reference model for the event with Mw 9.3. We see that this model suggests that about half of the yearly displacement would occur during the first month after the event, instead of the one-third observed. From that, we conclude that the Mw 9.3 model predicting four orders of magnitude drop of viscosity is clearly falsified for the case of the Tohoku event. However, the predicted displacements for the model of the event with the Mw 8.9 (purple curves), more representative of the Tohoku event, fit the observations remarkably well for the entire time range, including the first days after the event.

In Figure 10c, we show our model for the Mw 8.9 event at the time of 2.5 years after the earthquake. The colors represent horizontal displacements and arrows show displacement vectors. Note about 0.5 m landward displacement of the upper plate domain close to the trench and about double as large a displacement of the virtual land GPS stations in the opposite direction. Note also downward displacement of the surface of the upper plate some 100 km landward from the trench. All these modeling results agree well with the observations (Sun et al., 2014; Watanabe et al., 2014).

4.3. How Realistic Are Our Estimated RSF Parameters?

The suitable effective (a − b)* parameter to generate the largest earthquakes in our model is -4×10^{-4} (Table A2 in the Appendix A, third row from the top). To compare this value with the laboratory-derived values, we need first to make corrections for the high fluid pressure effect. The fluid pressure factor can be estimated using equation (2), the values for effective static friction in the channel of 0.015 and friction coefficient for dry rocks in the model shear zone of about 0.5. This gives a fluid pressure factor of 0.03. By dividing -4×10^{-4} by this factor we obtain an estimate of the "real," not affected by fluid pressure values of (a − b) of −0.013 in rate-weakening RSF depth range. If the effective static friction in the channel is 0.03, a fluid pressure factor is 0.06 and "real" value of (a − b) is −0.007. These absolute values (0.007–0.013) are at the upper range of the experimental values and are about 2–4 times larger than expected average values in the seismogenic zone (Blanpied et al., 1995; den Hartog et al., 2012). Note, however, that during the large earthquakes additional friction-weakening processes like flash heating and fluid pressurization are expected (Rice, 2006) and modeled (Noda & Lapusta, 2013). We may expect that fluid pressurization is particularly extensive at subduction faults with initially high fluid pressure like in South Chile. Therefore, we find it to be a reasonable suggestion from our model that for the largest earthquakes, which we consider here, about a half or more of the friction weakening is related to the processes additional to RSF, i.e., processes like flash heating and fluid pressurization.

As half or more of the friction weakening may be in fact related to other than RSF type of friction-weakening mechanisms, we have also run models with (a − b) parameters of 1/3 of the (a − b) parameter in the reference model and additional linear slip-weakening friction law (Appendix A). With the appropriate values of the slip-weakening parameter leading to similar stress drops during earthquakes, the resulting models are similar to our reference model (see supporting information Figure S5). Therefore, all conclusions of this paper also apply to the case of multiple friction-weakening mechanisms. However, we emphasize that in these models more than 60% of the seismic moments of earthquakes is due to non-RSF friction-weakening processes.

In the rate-strengthening part of the fault, we use an (a − b)* value of 4×10^{-4}. After correction for fluid pressure effect, this leads to (a − b) values of 0.01–0.013, which are consistent with experimental data for granite and phyllosilicates at a temperature higher than 300–400°C (Blanpied et al., 1995; den Hartog et al., 2012).

4.4. Model Limitations

The largest model limitation is its two dimensionality, which does not allow comparing model predictions with important details of observations. For example, surface coseismic displacements of the Tohoku 2011 event clearly show 3-D effects. Postseismic deformations are more evenly distributed in space but also have some 3-D features. Therefore, we are forced to use only normalized displacements to study the time evolution of mantle rheology with our 2-D model.

Another important limitation of our current models is their relatively low resolution, which does not allow us to resolve properly the processes of rupture nucleation and propagation (Lapusta et al., 2000). In the future, much higher resolution within the model shear zone (about 0.1 km of grid point spacing) should be implemented.

5. Concluding Remarks

We have developed a cross-scale model of subduction and subduction earthquakes' seismic cycles which cover the time scale range from 1 min (earthquake) to long-term (multiple seismic cycles and tectonic evolution, 10^3–10^7 years). The model employs nonlinear transient rheology in the crust and mantle and RSF rheology in the model shear zone. It generates spontaneous earthquake sequences and using an adaptive time step algorithm, recreates the deformation process as observed naturally during the seismic cycle and multiple seismic cycles.

The key feature of our model is nonlinear transient viscous rheology, where nonlinear stress dependency of viscosity plays a most important role. In this respect we follow, confirm, and further develop the study by Freed & Bürgmann (2004) and its extensions (Freed et al., 2006, 2007, 2010, 2012), which suggested about 10–20 times decrease of the viscosity in the mantle just after the earthquake. However, we show that viscosity drop in the mantle during the earthquake can be as high as three to four orders of magnitude, i.e., much larger than previously suggested. Our model is supported by the recent discovery that the

viscoelastic relaxation in the mantle has controlled postseismic deformation at least since one month after the Great Tohoku earthquake and that the afterslip downdip of the main rupture zone might have been substantially overestimated (Freed et al., 2017; Sun et al., 2014; Sun & Wang, 2015; Hu et al., 2016). However, we infer that even at the time range from one hour to a month after the great earthquake, viscoelastic deformation may be very active and therefore the deforming domain may be much larger than previously thought and imaged with the conventional linear models. This finding may have implications for delayed triggering of distant earthquakes (Freed, 2005) and volcanic eruptions. Furthermore, an increase of the stress accompanied by the increase of dislocation density in the rocks of the mantle wedge during the earthquake may increase seismic wave attenuation (Farla et al., 2012), thus diminishing the magnitude of ground shaking during the great earthquakes.

Our models predict about 10–20 times change of viscosity in the mantle wedge from a month after the great earthquake till the end of the seismic cycle, which is remarkably close to the viscosity change required to fit surface displacements in the conventional models of viscoelastic relaxation.

We also note that our results demonstrate that there is no contradiction between extremely low mechanical coupling (effective friction coefficient of 0.015) at the subduction megathrust in South Chile inferred from the long-term geodynamic model (Sobolev & Babeyko, 2005) and the appearance of a sequence of the largest earthquakes, like the Great Chile Earthquake of 1960.

From a comparison of our calibrated (a − b) RSF parameter with typical experimental values, we infer that more than half of the observed friction drop during great earthquakes may result from other than RSF friction-weakening processes, like flash heating and fluid pressurization. All conclusions of this study also apply to the models with multiple friction-weakening mechanisms, like a combination of RSF and slip-weakening mechanisms, which generate earthquakes with the similar average stress drops.

Finally, we note that although this particular study is focused at seismic-cycle time scale predictions, the model is also ready to study the still poorly understood effects of the multiple seismic cycles on the long-term deformation of the upper plate (Rosenau & Oncken, 2009; Wang, 1995) as well as for investigation of effects of various subduction zone parameters on the maximal magnitude of earthquakes (Heuret et al., 2011; Schellart & Rawlinson, 2013).

Giving credit to the model, we also admit that it certainly does not pretend fully to solve the extremely complex problem of linking long-term and earthquake time scale processes in subduction zones. Our model has several limitations due to its relatively low resolution, which does not allow it properly to model rupture propagation and to model small earthquakes and slow-slip events occurring between main events. Moreover, the model is 2-D, which significantly limits its application to natural events. All these issues are a matter for future work.

Appendix A: Details of the Modeling Technique

A1.1. Long-Term Model and Transition to Short-Term Model

A1.1.1. Long-Term Model

We used a 2-D version of SLIM3D (Popov & Sobolev, 2008), which is a thermomechanical finite element code based on the Arbitrary Lagrangian-Eulerian method, complemented by particle-in-cell technique. It solves the coupled thermomechanical conservation equations of momentum, energy, and mass, and employs nonlinear elastoviscoplastic rheology and true free-surface upper boundary condition. The governing equations in our models are:

$$\alpha\frac{DT}{Dt}-\frac{1}{K}\frac{DP}{Dt}=\frac{\partial v_i}{\partial x_i}, \tag{A1}$$

$$\frac{\partial\sigma_{ij}}{\partial x_j}+\rho g_i=\rho\frac{Dv_i}{Dt},$$

$$\rho C_p\frac{DT_p}{Dt}=k\nabla^2T+\left(\dot{\varepsilon}_{ij}^{vs}+\dot{\varepsilon}_{ij}^{pl}\right)\tau_{ij}+A,$$

where σ_{ij} is the Cauchy stress tensor, ρ is density, g_i is gravity acceleration vector, C_p is heat capacity, T is temperature, T_p is potential temperature, P is dynamic pressure, t is time, k is thermal conductivity, $\dot{\varepsilon}_{ij}^{vs}$ is

ductile creep strain rate tensor, $\dot{\varepsilon}_{ij}^{pl}$ is plastic strain rate tensor, τ_{ij} is the Cauchy stress deviator, A is the volumetric heat source, α is the coefficient of thermal expansion, k is thermal conductivity, K is the bulk modulus, v_i is velocity vector, and D/Dt is the substantial time derivative. Law of summation on repeating indexes applies.

The inertial term $\rho\frac{Dv_i}{Dt}$ in (A1) is ignored in the long-term model, but is considered during modeling of earthquakes.

The evolution of elastic, viscous, and plastic deformation is based on additive decomposition of the total deviatoric strain rate ($\dot{\varepsilon}$) as:

$$\dot{\varepsilon}_{ij}=\dot{\varepsilon}_{ij}^{el}+\dot{\varepsilon}_{ij}^{vs}+\dot{\varepsilon}_{ij}^{pl}=\frac{1}{2G}\hat{\tau}_{ij}+\frac{1}{2\eta_{eff}}\tau_{ij}+\dot{\gamma}\ \frac{\partial Q}{\partial\tau_{ij}}, \tag{A2}$$

where G is the elastic shear modulus, $\hat{\tau}_{ij}$ is the objective stress rate, η_{eff} is the effective viscosity, $\dot{\gamma}$ is the plastic multiplier or plastic strain rate, and Q is the plastic potential function.

The ductile temperature-dependent and stress-dependent viscosity in our models simultaneously accounts for the two creep mechanisms of diffusion and dislocation as follows:

$$\eta_{eff}=\frac{1}{2}\tau_{II}\left(\dot{\varepsilon}_{diff}+\dot{\varepsilon}_{dis}\right)^{-1}, \tag{A3}$$

where τ_{II} is the square root of the second invariant of the stress deviator and $\dot{\varepsilon}_{diff}$ and $\dot{\varepsilon}_{dis}$ are the square root of the second invariants of the strain rate tensors due to the diffusion and dislocation creeps, respectively, which are defined as:

$$\dot{\varepsilon}_{diff}=B_{diff}\tau_{II}\exp\left(-\frac{E_{diff}+PV_{diff}}{RT}\right), \tag{A4}$$

$$\dot{\varepsilon}_{dis}=B_{dis}\left(\tau_{II}\right)^{n}\exp\left(-\frac{E_{dis}+PV_{dis}}{RT}\right), \tag{A5}$$

in which B_{diff} is the preexponential constant for diffusion creep, E_{diff} is the activation energy for diffusion creep, P is pressure, V_{diff} is the activation volume for diffusion creep, R is the universal gas constant, T is temperature, B_{dis} is the preexponential constant for dislocation creep, E_{dis} is the activation energy for dislocation creep, V_{dis} is the activation volume for dislocation creep, and n is the power law exponent for dislocation creep.

Brittle (plastic) deformation is incorporated using the Drucker-Prager failure model with the yield surface (F) and plastic potential (Q) defined as:

$$F=\tau_{II}-c-\mu P;\ Q=\tau_{II}, \tag{A6}$$

where μ is the coefficient of friction and c is the cohesion.

The steady state creep parameters for diffusion and dislocation creep are determined by laboratory experiments with major rock types. The table with rheological parameters used in this study is presented in the Table A1 (Appendix A). We use typical thermal parameters for all rock types, C_p (heat capacity) of 1. kJ/(kg K) for all materials, k (thermal conductivity) of 2.5 W/(m K) for crust, and 3.3 W/(m K) for mantle and A (radiogenic heat production) of 1.5 nW/kg in continental upper crust, 0.2 nW/kg in continental lower crust, 0.02 nW/kg in oceanic crust, and 0 in mantle.

The model is 2-D, has horizontal dimension of 900 km and vertical (depth) dimension of 300 km. The finite element size in most of the models is 3 km, but we have also computed high-resolution (finite element size 1.5 km) and low-resolution versions (finite element size 6 km) of our reference model.

A1.1.2. Transition to Short-Term Model

The typical integration time step in the long-term model is 10 thousand years, while the maximal time step required for our short-term model is only a few years, so a time step change of more than 1,000 times is required. While in the model with viscoplastic rheology change of time step does not lead to a change of the stress field, in the model with elastoviscoplastic rheology the stress is changing. To take that into account, we performed the following "time zoom-in" strategy. First, we fix kinematic boundary conditions

Table A1
Material Parameters for the Reference Model

Material	Diffusion creep Log(B) (1/(Pa s))	E (kJ/mol)	V (cm^3)	Dislocation creep Log(B) (1/(Pa s))	E (kJ/mol)	V (cm^3)	n	Coefficient of friction μ_0*	Shear modulus G (GPa)
Model shear zone[#]	n.a.	n.a.	n.a.	17.3/−17.48	154	0	2.3	**0.015**	40
Oceanic crust[xx]	n.a.	n.a.	n.a.	−14.75	356	0	3.0	0.5	40
Oceanic lithosphere[xxx]	n.a.	n.a	n.a.	−15.2	530	17	3.5	0.5	74
Upper continental crust[x]	n.a.	n.a.	n.a.	−28.0	223	0	4.0	0.5	36
Lower continental crust[xx]	n.a.	n.a.	n.a.	−14.75	356	0	3.0	0.5	40
Continental lithosphere[xxxx]	−10.59	300	9	−14.7	480	12	3.5	0.5	74
Asthenosphere[xxxx]	−10.59	300	9	−14.7	480	12	3.5	0.5	74

Note. [x]wet quartzite, Gleason and Tullis (1995); [xx]wet anorthite, Rybacki and Dresen (2000); [xxx]dry olivine, 10 mm grain radius, Hirth and Kohlstedt (2003); [xxxx]wet olivine with water content of 1000 ppm H/Si, 10 mm grain radius, Hirth and Kohlstedt (2003); [#]wet quartz Ranalli (1997). Cohesion is 10 MPa in all materials accept model shear zone, where it is 2 MPa.

at all boundaries except the surface (remaining to be a stress-free boundary) using the velocities from the last time step of the long-term model. Then, we decrease the time step to five years and run the model till the velocity field becomes identical to the velocity field of the long-term model at its last time step. Empirically, we have found that 300–500 5 years' time steps are enough to get reasonable consistency of the velocity fields of long-term and short-term models.

In our models, we try to keep the model shear zone as narrow as possible. Therefore, we delete channel-phase markers in the parts of the channel (mostly in its uppermost part) where the channel in long-term model becomes thicker than 3–5 elements. After that, the long-term model is run a few hundred thousand years to reequilibrate the stress field.

A1.2. Modeling of Earthquakes

A1.2.1. Criterion for the Maximum Time Step

The maximum time step is defined to be small enough that RSF instability can develop. Assume for simplicity a steady state RSF regime. In this case, change of slip velocity from slow velocity V_{sl} to fast V_{ft} at time step Δt will cause stress drop $\Delta\sigma$

$$\Delta\sigma = \sigma_n \cdot (b-a)^* \cdot \ln\left(\frac{V_{ft}}{V_{sl}}\right), \tag{A7}$$

which in turn can be related to the slip at the fault

$$\Delta\sigma = C \cdot G\frac{V_{ft} \cdot \Delta t}{W}, \tag{A8}$$

where C is constant of the order of unity, G is the shear modulus and W is the characteristic rupture distance (width). Combining (A7) and (A8) we obtain the following simple equation for $Y = V_{ft}/V_{sl}$

Table A2
Reference Frictional Parameters in Model Shear Zone

Parameter	All models $0 < Z < 9$ km	Reference model $9 \text{ km} < Z < 100$ km	Ref. afterslip model $9 \text{ km} < Z < 43$ km	Ref. afterslip model $43 \text{ km} < Z < 100$ km	Combined RSF-Slip weakening models (1/2)
μ_{st}*,	0.015	0.015	0.015	0.015	0.015
μ_{st}	0.5	0.5	0.5	0.5	0.5
(a − b)*	0	-410^{-4}	-410^{-4}	-410^{-4}	−0.013
(a − b)	0	−0.013	−0.013	−0.013	−0.004
a*	0	1.610^{-3}	1.610^{-3}	810^{-4}	510^{-4}
A	0	0.053	0.053	0.027	0.017
$\Delta\mu_0$*	0	0	0	0	0.005/0.007

Note. (L=1 cm, $\left(1-\frac{p_f}{\sigma_n}\right)=0.03$, $V_0 = 7$ cm/yr).

$$\gamma Y = \ln Y, \text{ where } \gamma = C \cdot G \frac{V_{sl} \cdot \Delta t}{W \cdot \sigma_n \cdot (b-a)^*} \quad \text{(A9)}$$

Equation (A9) has a solution only if value of γ fulfills the following condition

$$\gamma = C \cdot G \frac{V_{sl} \cdot \Delta t}{W \cdot \sigma_n \cdot (b-a)^*} \le \frac{1}{e} = \frac{1}{2.72}. \quad \text{(A10)}$$

That defines a maximum value of the time step at which RSF instability can develop

$$\Delta t < \frac{W \cdot \sigma_n \cdot (b-a)^*}{2.72C \cdot G \cdot V_{sl}}. \quad \text{(A11)}$$

A1.2.2. Definition of the Minimum Time Step

As we do not model rupture propagation, we must assume some value for the minimum time step, Δt_{min}. For the 2-D model, a reasonable guess for Δt_{min} would be earthquake rise time scaled with the rupture width. For the great megathrust earthquakes that should be about several tens of seconds and much smaller for the strike-slip earthquakes with much smaller rupture widths. It is important that the modeling results do not depend much on the choice of Δt_{min}, as follows from the analysis below.

If in (A3) $\gamma \ll 1/e$, then (A9) has two solutions

$$Y_1 \cong 1 (\text{slowly creeping fault}) \text{ and } Y_2 \cong \frac{1}{\gamma} \ln \left(\frac{1}{\gamma} \ln \frac{1}{\gamma} \right). \quad \text{(A12)}$$

The second solution in (A12) represents fast slip during an earthquake. According to (A12), in the first approximation, the fast slip velocity is inversely proportional to the value of γ and hence also to Δt_{min}. It means that stress drop during the earthquake (see equation (A7)) depends on the logarithm of Δt_{min} and it changes only by about 5% if Δt_{min} changes by e times from 40 to 110 s, $\approx \sigma_n \cdot (b-a)^* \approx \Delta\sigma / \ln \left(\frac{V_{fl}}{V_{sl}} \right) \approx \Delta\sigma/18$.

At minimum time steps, we take into account the inertial term in (A1) in an explicit way during iterations. The explicit procedure efficiently converges because even at our smallest time steps (7.5 s) in tests of the modeling procedure (see below) the inertial term does not exceed a few percent of other terms in the equation of momentum conservation in (A1).

A1.2.3. Test of Modeling Procedure by Comparison With a Known High-Resolution Solution

We test our modeling procedure by simulating a seismic cycle at a strike-slip fault in a setup similar to the study by Lapusta et al. (2000) (supporting information Figure S1a). We aim to reproduce both solutions by Lapusta et al. (2000) for the case of relatively large L = 8 mm, h* = 0.94 km, and relatively small L = 2 mm, h* = 0.235 km. In the case of L = 8 mm, h* = 0.94 km Lapusta et al. (2000) obtained regular large events with the recurrence time of about 100 years, while in the case of smaller L = 2 mm, h* = 0.235 km, they obtained large events with a recurrence time of 70–80 years accompanied by much smaller events (with seismic moments less than 5% of the large event's moments) occurring some 20 years after the large events. We assume depth distribution of RSF parameters a, b, and L as in the study by Lapusta et al. (2000). Model finite element size is 0.1–0.3 km, which makes only about 0.7 grids at the smallest h* length, similar to our models for the megathrust earthquakes. The maximum time step is defined as 1 year, which fulfills condition (A11). We assume the smallest time step to be 7.5 s, about half of the rupture duration in Lapusta et al. (2000). To recognize the appearance of instability, we monitor strain rates at the entire 24 km of fault and run models with a value of CA parameter of 100%.

Our modeling procedure generates regular periodic sequences of large events with periods of about 100 years for h* = 0.94 km (supporting information Figure S1b), and about 70–80 years for h* = 0.235 km (supporting information Figure S1c), very similar to the much higher-resolution models by Lapusta et al. (2000). However, as expected, our algorithm is missing small events modeled by Lapusta et al. (2000) for h* = 0.235 km. Coseismic slip distributions for large events in our models (supporting information Figure S1d) are also similar to the solutions by Lapusta et al. (2000). Note that our solution for h* = 0.94 km lies between quasi-dynamic (dashed green curve in supporting information Figure S1d) and fully dynamic solutions (dashed blue curve in supporting information Figure S1d) by Lapusta et al. (2000). We conclude that our modeling procedure successfully passes the test by comparison with the high-resolution solution (Lapusta et al., 2000). Although we cannot reproduce small details of the solution, like small events, we do

reproduce major features of the solution, including recurrence times and magnitudes of main events as well as coseismic slip distributions.

A1.3. Robustness of the Models

A1.3.1. Model Convergence

Numerical tests show that changing the minimum time step by 2 times in both directions changes the average period and average moment of events by less than 4% (larger period and moment for the shorter time step) in agreement with the theoretical analysis presented in the previous section. Two times increase of grid size (to 6 km, low resolution), with all other parameters fixed, decreases the average 2-D seismic moment of the earthquakes from 1.57×10^{17} N (reference resolution) to 1.27×10^{17} N (20% difference). Two times decrease of grid size (to 1.5 km, high resolution) increases the average 2-D moment of the earthquakes from 1.57×10^{17} N to 1.66×10^{17} N that makes less than 6% difference (supporting information Figure S2a). Note that this small difference is entirely due to the difference in the stress distribution in the deep ductile-flow portion of the model shear zone. The width of the zone of the localized ductile flow is 1.5 times larger in the model with a grid size of 3 km (supporting information Figure S2c) than in the model with a grid size of 1.5 km (supporting information Figure S2b). This results in a slightly shallower point of brittle-ductile transition in the model with lower resolution; the consequence is slightly smaller seismic events. This effect is easy to compensate by decreasing by 1.5 times the preexponential coefficient (B) in the rheological model for the dislocation creep (equation (A5)) in the material of the model shear zone. Results for this corrected model with the grid size of 3 km (blue symbols in supporting information Figure S2a) are identical to the results of the high-resolution model.

A1.3.2. Effect of the Instability Criterion (CA)

Modeled recurrence time of earthquakes, seismic moment, stress drop and distribution of slip at the fault do not depend on the value of CA criterion unless CA becomes less than some 30% or higher than some 180% for megathrust events (supporting information Figure S3). This demonstrates the robustness of our procedure for identifying the appearance of instability and estimation of cumulative slip distribution.

A1.3.3. Stable Solutions at an Abrupt Decrease of Time Step at Earthquake

Despite the abrupt change of the integration time step in the modeling of the earthquake, we obtain stable solutions. Supporting information Figure S4 shows strain rate fields and velocity vectors during several successive time steps just before, during, and after the earthquake. One time step of 5 years just before the earthquake, we obtain deformation pattern typical for the interseismic period (supporting information Figure S4a). At the next 5 years' step, the critical instability is recognized (supporting information Figure S4b). In our modeling procedure, this is a proxy for rupture nucleation. This step is recalculated with the small time step of 37 s. This and next time step solutions (supporting information Figures S4c and S4d) are proxies for rupture propagation and arrest. Solutions at next steps show decaying afterslip in the deep (rate strengthening) portion of the fault. All solutions are stable and reproducible. Note the changing scales for strain rate and velocity vectors in Fig 4.

A1.4. Nonlinear Transient Cross-Scale (Reference) Models

Our *reference model* has the following parameters: Transient rheology parameter $\beta = 10$, RSF parameters: L = 1 cm, b/(a − b) = −3, (a − b)* = 0 at the depth range of 0–9 km, and (a − b)* = -4×10^{-4} at depth >9 km without depth limit (Table A2). The *reference afterslip model* has the same parameters as the reference model but (a − b)* = 4×10^{-4} at depth >42 km. All these models generate earthquakes with Mw about 9.2–9.3 and are called "Mw 9.2" or "Mw 9.3" models.

By changing (a − b) to -1×10^{-4} with all other parameters as in the reference models, we obtain an earthquake with about four times lower seismic moment. The corresponding models are called "Mw 8.9" models.

Table A3
Mantle Rheological Parameters for Additional Nonlinear Transient Models

Parameter	Reference model	Model n3	Model n4	Model Bx10	Model B/10	Model B/10 ocean
N	3.5	3	4	3.5	3.5	3.5
Log(B_{dis}) wedge	−14.7	−12.2	−17.7	−13.7	−15.7	−14.7
Log(B_{dis}) ocean	−14.7	−12.2	−17.7	−13.7	−15.7	−15.7

A1.5. Other Nonlinear Models

First, we explore the effect of transient rheology parameter β. Changing β from 10 (reference model) to 1 (no transient rheology) affects deformation during the first few weeks after the Mw 9.3 event, but after that time the models are identical (Figure 8a). This means that the time required to achieve steady creep in our models is less than 1 month, i.e., is much smaller than the recurrence time of great earthquakes.

We also compute nonlinear transient models with different rheological parameters (see Table A3). First, we vary the preexponential

coefficient (B) in the rheological model for the dislocation creep (equation (A5)) by 10 times. The corresponding model with B coefficient increased by 10 times we mark as the "Bx10" model and model with B coefficient decreased by 10 times as the "B/10" model. Decreasing (or increasing) this coefficient by 10 times relative to the reference model is equivalent to decreasing (or increasing) the water content by about 8 times or decreasing (or increasing) the mantle temperature by about 90–100°C.

In the next set of models, we explore the effect of changing the stress exponent parameter n. In these models, we scale the preexponential coefficient B to keep viscosity constant at typical long-term (interseismic) stresses in the asthenospheric wedge of $\tau_0 = 10^5$ Pa. The scaled B (n) is then defined by

$$B(n) = B(3.5) * \left(\tau_0 = 10^5 Pa\right)^{3.5-n}, \tag{A13}$$

where B(3.5) is the dislocation creep preexponential coefficient for the reference model with n = 3.5 ($B(3.5) = 10^{-14.7}$, see Table A1). Thus defined coefficients for the models with n = 3 and n = 4 are $B(3) = 10^{-12.2}$ and $B(4) = 10^{-17.2}$, respectively.

We also computed the model with 10 times lower B coefficient in the oceanic part of the asthenosphere than in the asthenospheric wedge, having in mind possibly lower water content in the oceanic mantle than in the asthenospheric wedge (Masuti et al., 2016). The model is marked as the "ocean B/10" model. Rheological parameters for the above models are given in Table A2.

A1.6. Linear Rheology Models

The linear transient model has a constant viscosity of 3×10^{18} Pa s in the asthenosphere, transient rheology parameter $\beta = 10$ and RSF parameters as in the reference model. The linear transient afterslip model (a proxy for the conventional model) has a constant viscosity of 3×10^{18} Pa s in the asthenosphere (depth >80 km), transient rheology parameter $\beta = 10$ and RSF parameters as in the reference afterslip model. Linear steady models have the same parameters as linear transient models, but $\beta = 1$, which leads to no transient viscosity changes (see equation (1)). To examine the effect of afterslip, we also computed a linear transient model with wedge viscosity of 10^{19} Pa s and $\beta = 10$ (see Figure 8c).

A1.7. Models With Both RSF and Slip-Weakening Friction

As half or more of the friction weakening may be in fact related to other than RSF type of friction-weakening mechanisms (see Discussion in main text), we have also run models with (a − b) parameters of 1/3 of (a − b)* parameter in the reference model and additional linear slip-weakening friction law.

In these models, we replace equation (4) by the following equation

$$\mu^* = \mu^*_{st} + a^* \cdot \ln\left(\frac{V}{V_{st}}\right) + b^* \cdot \ln\left(\frac{\Theta V_{st}}{L}\right) - \Delta\mu^*(D);\ \Delta\mu^*(D) = \Delta\mu^*_0 \cdot \min\left(\frac{D}{D_0}, 1\right), \tag{A14}$$

where D is the slip at any point of the fault during the first time step of the earthquake, $\Delta\mu^*_0$ is the maximum non-RSF friction drop and D_0 is a characteristic displacement. If the value of D_0 is large enough (about a meter or so), the additional friction drop takes place only during a large earthquake. We have run two models similar to our reference model but with reduced $(a - b)^* = -1.3 \times 10^{-4}$, and additional slip-weakening friction with parameters $\Delta\mu^*_0 = 0.005$, $D_0 = 1$ m (model 1), and $\Delta\mu^*_0 = 0.007$, $D_0 = 1$ m (model 2) (see supporting information Figure S5). These models generate seismic cycles with events with average 2-D moments of 1.25×10^{17} N (model 1, equivalent to Mw = 9.27 at rupture length of 800 km) and of 1.8×10^{17} N (model 1, equivalent to Mw = 9.37 at rupture length of 800 km). The recurrence times are about 250 years for model 1 and 400–450 years for model 2 (supporting information Figure S5). Stress drop varies between 2.5 and 3.5 MPa for the events of model 1 and between 6 and 7 MPa for the events of model 2. In model 1, about 60% and in model 2, about 70% of the seismic moment is due to non-RSF friction-weakening processes. However, all major features of these models are the same as in the reference model and hence all conclusions of the paper also apply.

Acknowledgments
The research by IM was funded by GEOSIM initiative of the GEO-X program. We are particularly grateful to Andrey Babeyko for the fruitful discussions and help during editorial process. Discussions with Shun Karato, Kelin Wang, Jean Philippe Avouac, Rolland Bürgmann, and Pablo Ampuero were very helpful. Anton Popov provided useful consulting concerning modification of SLIM3D code and Jonathan Bedford helped during editorial process. We also thank Sylvain Barbot, Ylona van Dinther, and anonymous reviewers for helpful comments. All data related to this paper can be obtained from S. Sobolev (stephan@gfz-potsdam.de).

References

Baes, M., Gerya, T., & Sobolev, S. V. (2016). 3-D thermo-mechanical modeling of plume-induced subduction initiation. *Earth and Planetary Science Letters*, *453*, 193–203. https://doi/org/10.1038/nature18319

Bai, Q., Mackwell, S. J., & Kohlstedt, D. L. (1991). High-temperature creep of olivine single cyrstals. 1: Mechanical results for buffered samples. *Journal of Geophysical Research*, *96*, 2441–2463. https://doi/org/10.1029/90JB01723

Barbot, S., Fialko, Y., & Bock, Y. (2009). Postseismic deformation due to the Mw 6.0 2004 Parkfield earthquake: Stress-driven creep on a fault with spatially variable rate-and-state friction parameters. *Journal of Geophysical Research*, *114*, B07405. https://doi/org/10.1029/2008JB005748

Barbot, S., Lapusta, N., Avouac, J. P., Leprince, S., Lucas, A., & Mattson, S. (2012). Under the hood of the earthquake machine: Toward predictive modeling of the seismic cycle. *Science*, *336*, 707–710.

Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults: Continuum-discrete transitions, evolutionary changes and corresponding dynamic regimes. *Reviews of Geophysics*, 46, RG4006. https://doi/org/10.1029/2008RG000260

Blanpied, M. L., Lockner, D. A., & Byerlee, J. D. (1995). Frictional slip of granite at hydrothermal conditions. *Journal of Geophysical Research*, *100*, 13045–13064.

Bürgmann, R., & Dresen, G. (2008). Rheology of the lower crust and upper mantle: Evidence from rock mechanics, geodesy and field observations. *Annual Review of Earth and Planetary Sciences*, *36*, 531–567.

Burov, E., Francois, T., Agard, P., Le Pourhiet, L., Meyer, B., Tirel, C., . . . Brun, J.-P. (2014). Rheological and geodynamic controls on the mechanisms of subduction and HP/UHP exhumation of crustal rocks during continental collision: Insights from numerical models. *Tectonophysics*, *631*, 212–250. https://doi/org/10.1016/j.tecto.2014.04.033

Cisternas, M. (2005). Predecessors of the giant 1960 Chile earthquake. *Nature*, *437*, 404–407.

Currie, C. A., & Hyndman, R. D. (2006). The thermal structure of subduction zone back arcs. *Journal of Geophysical Research*, *111*, B08404. https://doi/org/10.1029/2005JB004024

den Hartog, S. A. M., Niemeijer, A. R., & Spiers, C. J. (2012). New constraints on megathrust slip stability under subduction zone *P* – *T* conditions. *Earth and Planetary Science Letters*, *353–354*, 240–252. https://doi/org/10.1016/j.epsl.2012.08.022.

Dzierma, Y., Thorwart, M., Rabbel, W., Siegmund, C., Comte, D., Bataille, K., . . . Prezzi, C. (2012). Seismicity near the slip maximum of the 1960 Mw 9.5 Valdivia earthquake (Chile): Plate interface lock and reactivation of the subducted Valdivia Fracture Zone. *Journal of Geophysical Research*, *117*, B06312. https://doi/org/10.1029/2011JB008914

Farla, R. J. M., Jackson, I., Gerald, J. D. F., Faul, U. H., & Zimmerman, M. E. (2012). Dislocation damping and anisotropic seismic wave attenuation in Earth's upper mantle. *Science*, *336*, 332–335.

Freed, A. M. (2005). Earthquake triggering by static, dynamic, and postseismic stress transfer. *Annual Review of Earth and Planetary Sciences*, *33*, 335–367.

Freed, A. M., & Bürgmann, R. (2004). Evidence of power-law flow in the Mojave Desert mantle. *Nature*, *430*, 548–551.

Freed, A. M., Bürgmann, R., Calais, E., & Freymueller, J. T. (2006). Stress-dependent power-law flow in the upper mantle following the 2002 Denali, Alaska, earthquake. *Earth and Planetary Science Letters*, *252*, 481–489.

Freed, A. M., Bürgmann, R., & Herring, T. A. (2007). Far-reaching transient motions after Mojave earthquakes require broad mantle flow beneath a strong crust. *Geophysical Research Letters*, *34*, L19302. https://doi/org/10.1029/2007GL030959

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.

Freed, A. M., Herring, T., & BüRgmann, R. (2010). Steady state laboratory flow laws alone fail to explain postseismic observations. *Earth and Planetary Science Letters*, *300*, 1–10. https://doi/org/10.1016/j.epsl.2010.10.005

Freed, A. M., Hirth, G., & Behn, M. D. (2012). Using short-term postseismic displacements to infer the ambient deformation conditions for the upper mantle. *Journal of Geophysical Research*, *117*, B01409. https://doi/org/10.1029/2011JB008562

Gerya, T. (2011). Future directions in subduction modelling. *Journal of Geodynamics*, *52*, 344–378. https://doi/org/10.1016/j.jog.2011.06.005

Gleason, G. C., & Tullis, J. (1995). A flow law for dislocation creep of quartz aggregates determined with the molten salt cell. *Tectonophysics*, *247*, 1–23.

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

Hirth, G., & Kohlstedt, D. L. (2003). Rheology of the upper mantle and the mantle wedge: A view from the experimentalists. In J. Eiler (Ed.), *Inside the subduction factory, Geophysical Monograph* (Vol. 138, pp. 83–105). Washington, DC: American Geophysical Union.

Hoechner, A., Sobolev, S. V., Einarsson, I., & Wang, R. (2011). Investigation on afterslip and steady state and transient rheology based on postseismic deformation and geoid change caused by the Sumatra 2004 earthquake. *Geochemistry Geophysics Geosystems*, *12*, Q07010. https://doi/org/10.1029/2010GC003450

Hu, Y., Bürgmann, R., Uchida, N., Banerjee, P., & Freymueller, J. T. (2016). Stress-driven relaxation of heterogeneous upper mantle and time-dependent afterslip following the 2011 Tohoku earthquake. *Journal of Geophysical Research: Solid Earth*, *121*, 385–411. https://doi/org/10.1002/2015JB012508

Karato, S. I. (1998). Micro-physics of post glacial rebound. *GeoResearch Forum*, *3–4*, 351–364.

Karato, S. I., & Jung, H. (2003). Effects of pressure on high-temperature dislocation creep in olivine. *Philosophical Magazine*, *83*, 401–414.

Khazaradze, G., Wang, K., Klotz, J., Hu, Y., & He, J. (2002). Prolonged post-seismic deformation of the 1960 great Chile earthquake and implications for mantle rheology. *Geophysical Research Letters*, *29*(22), 2050. https://doi/org/10.1029/2002GL015986

Lambert, V., & Barbot, S. (2016). Contribution of viscoelastic flow in earthquake cycles within the lithosphere-asthenosphere system. *Geophysical Research Letters*, *43*, 10142–10154. https:/doi.org/10.1002/2016GL070345

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.), *The mechanics of faulting: From laboratory to real earthquakes* (pp. 153–207). Kerala, India: Research Signpost.

Lapusta, N., Rice, J. R., Ben-Zion, Y., & Zheng, G. (2000). Elastodynamic analysis for slow testonic loading with spontaneous rupture episodes on faults with rate- and state-dependent friction. *Journal of Geophysical Research*, *105*, 23765–23789. https://doi/org/10.1029/2000JB900250

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

Marone, C. J., Scholz, C. H., & Bilham, R. (1991). On the mechanics of earthquake afterslip. *Journal of Geophysical Research*, *96*, 8441–8452.

Masuti, S., Barbot, S. D., Karato, S.-I., Feng, L., & Banerjee, P. (2016). Upper-mantle water stratification inferred from observations of the 2012 Indian Ocean earthquake. *Nature*, *538*, 373–377.

Moreno, M. S., Bolte, J., Klotz, J., & Melnick, D. (2009). Impact of megathrust geometry on inversion of coseismic slip from geodetic data: Application to the 1960 Chile earthquake. *Geophysical Research Letters*, *36*, L16310. https://doi/org/10.1029/2009GL039276

Muto, J., Shibazaki, B., Iinuma, T., Ito, Ohta, Y. Y., Miura, S., & Nakai, Y. (2016). Heterogeneous rheology controlled postseismic deformation of the 2011 Tohoku-Oki earthquake. *Geophysical Research Letters*, *43*, 4971–4978. https://doi/org/10.1002/2016GL068113

Noda, H., & Lapusta, N. (2013). Stable creeping fault segments can become destructive as a result of dynamic weakening. *Nature*, *493*, 518–521.

Perfettini, H., & Avouac, J.-P. (2004). Postseismic relaxation driven by brittle creep: A possible mechanism to reconcile geodetic measurements and the decay rate of aftershocks, application to the Chi-Chi earthquake, Taiwan. *Journal of Geophysical Research*, *109*, B02304. https://doi/org/10.1029/2003JB002488

Pollitz, F. F. (2003). Transient rheology of the uppermost mantle beneath the Mojave Desert, California. *Earth and Planetary Science Letters*, *215*, 89–104.

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

Popov, A. A., Sobolev, S. V., & Zoback, M. D. (2012). Modeling evolution of the San Andreas Fault System in northern and central California. *Geochemistry, Geophysics, Geosystems*, *13*, Q08016. https://doi/org/10.1029/2012GC004086

Quinteros, J., & Sobolev, S. V. (2013). Why has the Nazca plate slowed since the Neogene? *Geology*, *41*, 31–34. https://doi/org/10.1130/G33497.1

Ranalli, G. (1997). Rheology of the lithosphere in space and time. Burg, J.-P., Ford, M. (Eds.), Orogeny through Time. *Geological Society of London, Special Publication*, *121*, 19–37.

Rosenau, M., & Oncken, O. (2009). Fore-arc deformation controls frequency-size distribution of megathrust earthquakes in subduction zones.. *Journal of Geophysical Research*, *114*, B10311. https://doi/org/10.1029/2009JB006359

Rice, J. R. (1993). Spatio-temporal complexity of slip on a fault. *Journal of Geophysical Research*, *98*, 9885–9907. 7

Rice, J. R. (2006). Heating and weakening of faults during earthquake slip. *Journal of Geophysical Research*, *111*, B05311. https://doi/org/10.1029/2005JB004006

Rice, J. R., & Ben-Zion, Y. (1996). Slip complexity in earthquake fault models. *Proceedings of the National Academy Sciences of United States of America*, *93*, 3811–3818.

Ruina, A. L. (1983). Slip instability and state variable friction laws. *Journal of Geophysical Research*, *88*, 10359–10370.

Rybacki, E., & Dresen, G. (2000). Dislocation and diffusion creep of synthetic anorthite aggregates. *Journal of Geophysical Research*, *105*, 26017–26036.

Schellart, W. P., & Rawlinson, N. (2013). Global correlations between maximum magnitudes of subduction zone interface thrust earthquakes and physical parameters of subduction zones. *Physics of the Earth and Planetary Interiors*, *225*, 41–67.

Scholz, C. H. (1998). Earthquakes and friction laws. *Nature*, *391*, 37–42.

Schubert, G., Turcotte, D. L., & Olsen, P. (2001). *Mantle convection in the earth and planets*. Cambridge, UK: Cambridge University Press.

Seno, T. (2014). Stress drop as a criterion to differentiate subduction zones where Mw 9 earthquakes can occur. *Tectonophysics*, *621*, 198–210. https://doi/org/10.1016/j.tecto.2014.02.016

Sobolev, S. V., & Babeyko, A. A. (2004). A thermomechanical model of low-angle subduction in Central Andes combining geological and recent (GPS) deformations. 2004 Joint Assembly, Montreal, Canada, abstract G23A-04. Retrieved from http://abstractsearch.agu.org/meetings/2004/SM/G23A-04.html

Sobolev, S. V., & Babeyko, A. Y. (2005). What drives orogeny in the Andes? *Geology*, *33*, 617–620.

Sun, T., & Wang, K. (2015). Viscoelastic relaxation following subduction earthquakes and its effects on afterslip determination. *Journal of Geophysical Research: Solid Earth*, *120*, 1329–1344. https://doi/org/10.1002/2014JB011707

Sun, T., Wang, K., Iinuma, T., Hino, R., He, J., Fujimoto, H., . . . Hu, Y. (2014). Prevalence of viscoelastic relaxation after the 2011 Tohoku-Oki earthquake. *Nature*, *514*, 84–87.

van Dinther, Y., Gerya, T., Dalguer, L., Corbi, F., Funiciello, F., & Mai, P. (2013a). 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. (2013b). 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 of off-megathrust earthquakes. *Geophysical Research Letters*, *41*, 1194–1201. https://doi/org/10.1002/2013GL058886

Wang, K. (1995). Coupling of tectonic loading and earthquake fault slips at subduction zones. *Pure Applied Geophysics*, *145*, 537–559.

Wang, K., Hu, Y., & He, J. (2012). Deformation cycles of subduction earthquakes in a viscoelastic Earth. *Nature*, *484*, 327–332.

Watanabe, S., Sato, M., Fujita, M., Ishikawa, T., Yokota, Y., Ujihara, N., & Asada A. (2014). Evidence of viscoelastic deformation following the 2011 Tohoku-Oki earthquake revealed from seafloor geodetic observation. *Geophysical Research Letters*, *41*, 5789–5796. https://doi/org/10.1002/2014GL06113