Numerical modelling of post-seismic rupture propagation after the Sumatra 26.12.2004 earthquake constrained by GRACE gravity data

V. Mikhailov, V. Lyakhovsky, I. Panet, Y. van Dinther, M. Diamant, T. Gerya, O. deViron and E. Timoshkina

1Univ Paris Diderot, Sorbonne Paris Cité, Institut de Physique du Globe de Paris, UMR 7154 CNRS, F-75013 Paris, France. E-mail: mikh@ifz.ru
2Schmidt Institute of Physics of the Earth, Russian Academy of Science, B. Gruzinskaya 10, Moscow 123810, Russia
3Geological Survey of Israel, Jerusalem 95501, Israel
4Institut National de l'Information Géographique et Foresti`ere, Laboratoire LAREG, Univ. Paris Diderot; GRGS, Bat. Lamarck A, Case 7071, 35 rue Hélène Brion, 75205 Paris Cedex 13, France
5Swiss Seismological Service, ETH Zürich, Sonneggstrasse 5, CH-8092 Zurich, Switzerland
6Institute of Geophysics, ETH - Zürich, Sonneggstrasse 5, CH-8092 Zürich, Switzerland

Accepted 2013 April 9. Received 2013 March 1; in original form 2012 October 06

SUMMARY
In the last decades, the development of the surface and satellite geodetic and geophysical observations brought a new insights into the seismic cycle, documenting new features of inter-, co-, and post-seismic processes. In particular since 2002 satellite mission GRACE provides monthly models of the global gravity field with unprecedented accuracy showing temporal variations of the Earth’s gravity field, including those caused by mass redistribution associated with earthquake processes. When combined with GPS measurements, these new data have allowed to assess the relative importance of afterslip and viscoelastic relaxation after the Sumatra 26.12.2004 earthquake. Indeed the observed post-seismic crustal displacements were fitted well by a viscoelastic relaxation model assuming Burgers body rheology for the asthenosphere (60–220 km deep) with a transient viscosity as low as $4 \times 10^{17}$ Pas and constant $\sim 10^{19}$ Pas steady state viscosity in the 60–660-km depth range. However, even the low-viscosity asthenosphere provides the amplitude of strain which gravity effect does not exceed 50 per cent of the GRACE gravity variations, thus additional localized slip of about 1 m was suggested at downdip extension of the coseismic rupture. Post-seismic slip at coseismic rupture or its downdip extension has been suggested by several authors but the mechanism of the post-seismic fault propagation has never been investigated numerically. Depth and size of localized slip area as well as rate and time decay during the post-seismic stage were either assigned a priori or estimated by fitting real geodesy or gravity data. In this paper we investigate post-seismic rupture propagation by modelling two consequent stages. First, we run a long-term, geodynamic simulation to self-consistently produce the initial stress and temperature distribution. At the second stage, we simulate a seismic cycle using results of the first step as initial conditions. The second short-term simulation involves three substeps, including additional stress accumulation after part of the subduction channel was locked; spontaneous coseismic slip; formation and development of damage zones producing afterslip. During the last substep post-seismic stress leads to gradual $\sim 1$ m slip localized at three faults around $\sim 100$-km downdip extension of the coseismic rupture. We used the displacement field caused by the slip to calculate pressure and density variations and to simulate gravity field variations. Wavelength of calculated gravity anomaly fits well to that of the real data and its amplitude provides about 60 per cent of the observed GRACE anomaly. Importantly, the surface displacements caused by the estimated afterslip are much smaller than those registered...
1 INTRODUCTION

During the last decades, new geodetic and geophysical data became available to monitor seismogenic areas providing new insights into dynamics of plate boundaries and rheology of the Earth’s upper layers. Seismological and continuous GPS networks revealed silent earthquakes, episodic tremor and slip and other fundamentally new processes. More recently time varying satellite gravity provided new data for crucial yet poorly monitored areas of underwater subduction zones where recent giant earthquakes occurred (Chen et al. 2007; Ogawa & Heki 2007; Panet et al. 2007; Han et al. 2008 and references herein). Revealing new features of the large-scale mass redistribution involved in the recent megathrust events and post-seismic upper mantle response this data calls for a more complete study of the physical processes governing the entire seismic cycle. In particular, Panet et al. (2007) and Chen et al. (2007) reported positive post-seismic gravity changes by analysing GRACE monthly solutions in the area of the Sumatra 26.12.2004 earthquake. This positive anomaly compensates a coseismic negative anomaly leading to a strikingly fast relaxation of the coseismic gravity variations.

Three main mechanisms are generally invoked to explain post-seismic deformations and gravity field variations: afterslip, poroelastic effects and the viscoelastic response of the crust and mantle to the coseismic stress changes. Ogawa & Heki (2007) also hypothesized that the post-seismic gravity variations might be due to upper mantle water diffusion following the earthquake. Poroelasticity alone is not likely to explain satellite gravity observations, since it results in localized effects (Masterlark et al. 2001). Thus, afterslip and viscoelastic relaxation are usually considered for the modelling observed post-seismic signals, although there is no consensus on their relative importance. Hashimoto et al. (2006) showed that the GPS post-seismic horizontal velocities in the area of the Sumatra 2004 earthquake can entirely be explained by afterslip in the deepest portions of the coseismic rupture area. On the contrary, Pollitz et al. (2006) successfully attributed all GPS displacements to viscoelastic relaxation using Burgers rheology and assuming a low-transient viscosity in the asthenosphere. Therefore, GPS data alone does not allow discriminating between these two processes.

Interpreting both the fast growth of the positive post-seismic gravity anomaly above the trench area after the Sumatra 26.12.2004 event and GPS displacements, Panet et al. (2010) demonstrated that the observation can only be explained by a combination of viscoelastic relaxation and afterslip. Indeed the amplitude and shape of the predicted gravity signal resulting from viscoelastic relaxation differ from the observed one: assuming a low-transient viscosity in the Burgers rheology model, viscoelastic relaxation explains about half of the observed gravity signal. Panet et al. (2010) thus suggested that stress redistribution during the main event provoked slip on the downdip extension of the coseismic rupture. They showed that 75 cm of slip during 33 months on the 100 km interval extending downdip from the base of the coseismic rupture provides the necessary additional gravity signal. Afterslip at depth, following the Sumatra 2004 earthquake, is corroborated by a large number of registered aftershocks (Subarya et al. 2006; Banerjee et al. 2007).

To explain post-seismic gravity changes Han et al. (2008) suggested a low-transient viscosity as low as $5 \times 10^{17}$ Pas in a biviscous Burgers rheology, and also mentioned the importance of the afterslip. Paul et al. (2012) combined viscoelastic relaxation with afterslip assigned only downdip of the coseismic rupture and estimated an average slip velocity up to 50 cm yr$^{-1}$. Hu & Wang (2012) combined viscoelastic relaxation and slip on the coseismic rupture and its immediate deep extension. Hence, the relative role of afterslip and viscoelastic relaxation is still a matter of debate.

The Sumatra subduction zone is not the only region where both viscoelastic relaxation and afterslip are essential to explain post-seismic observations. For example, this is also the case for the 27.03.1964, $M_0$ 9.2 Alaska earthquake. Interpretation of triangulation, tide gauge and levelling measurements collected after the event suggested that both viscoelastic and post-seismic slip are required to explain the observed transient deformations there (Brown et al. 1977; Suito & Freymueller 2009).

The hypothesis that afterslip occurs on a deep extension of the main rupture for giant earthquakes implies that the slow rupture can propagate into the upper mantle. This is supported by seismological data. For instance, Déverchère et al. (1991) or Lindenfeld & Rümpker (2011) detected several earthquakes located well within the mantle beneath the Baikal and the East African Rift, respectively. Heuret et al. (2011) argued that about 70 per cent of the seismogenic zones extend deeper than 10 km below the overriding plate Moho. If the sinking plate transports water along the plate interface, then serpentinization of the mantle wedge could occur, favouring stable aseismic sliding (Hyndman & Peacock 2003 and references therein). However, this is probably not the case for the Sumatra area. Combining data on the pattern and rate of uplift deduced from coral growth with GPS horizontal displacements before the Sumatra earthquake, Simoes et al. (2004) showed that the locked fault zone extended below the forearc Moho. This is confirmed by the existence of deep slip (up to 55 km depth) suggested for the coseismic rupture (Chlieh et al. 2007). Hence, according to these authors, in the particular Sumatra island arc settings, the locked fault zone most probably extends into the mantle. This suggests that either the mantle is not serpentinized, or that the presence of serpentine does not necessarily imply stable sliding.

Propagation of the brittle fracture zone into the upper mantle after a giant earthquake appears compatible with rock mechanics and allows to explain the observations. However, the mechanism...
of the post-seismic fault propagation in the upper mantle still calls for
investigation. Here, we address the mechanism of post-seismic fault
propagation by developing a 2-D numerical model governed by
continuum damage rheology. We simulate the coseismic and post-
seismic state of stress and also investigate whether the post-seismic
stress distribution favours downip propagation of the coseismic
rupture. Then, we estimate the size of damage areas, amplitude
and evolution of the displacement rate. Finally, we compare our
model predictions with the GRACE-derived gravity anomalies for
the Sumatra case.

The numerical modelling is done in two stages. First, we use
a plane-strain coupled petrological and thermomechanical code
I2ELVIS (Gerya & Yuen 2007; Gerya 2011; van Dinther et al. 2013)
to reproduce long-term plate motion associated with subduction.
Then, the obtained model structure, including stress distribution,
temperature and material properties, serves as initial conditions for
a short-term simulation using the Fast Lagrangian Analysis of Con-
tinua (FLAC) algorithm (Cundall & Board 1988; Cundall 1989;
Poliakov et al. 1993; Ilchev & Lyakhovsky 2001). In this second
step, the lithosphere is governed by a viscoelastic damage rheology
accounting for large strain associated with permanent brittle de-
formation (e.g. Lyakhovsky et al. 2011). This stage includes stress
accumulation prior to the main seismic event and the formation of
a damage zone that follows the abrupt coseismic stress drop. Prior
to present our modelling approach we first shortly introduce the
satellite gravity data over the Sumatra area.

2 GRACE GRAVITY DATA

We use the earthquake signal extraction from the total GRACE
geoid time variations as published by de Linage et al. (2009)
and Panet et al. (2010). Here, we briefly describe the GRACE
geoid data set and the key post-processing steps. We refer the
interested reader to the above cited papers for more details. The
GRACE geoid models are the Release 1 decadal gravity models of
the CNES Groupe de Recherche en Géodésie Spatiale (GRGS)
(Bruinsma et al. 2010) for the period between 2002 August
and 2007 September, computed from the GRACE raw satellites
measurements. A complete description of the processing strategy
and models used can be found in Lemoine et al. (2007). These
geoids are provided on the website of the Bureau Gravimetrique
International (http://grgs.obs-mip.fr/grace/variable-models-grace-
legeos/grace-solutions-release-01#tab2) in the form of spherical
harmonics coefficients up to degree 50, which corresponds to 400-
km half wavelength. These data are already corrected for gravity
variations at periods shorter than 1 month using high-frequency gen-
eral circulation models for the atmosphere, the ocean and an ocean
tide model. The remaining signal contains effects of continental
water mass balance, dynamics of the polar ice caps and oceans and
different solid Earth signals. It also contains noise, mostly asso-
ciated with the aliasing of the unmodelled high-frequency signals
remaining in the GRACE measurements, producing longitudinal
stripes in the geoids. In the GRGS Release 1 models, the harmonic
coefficients higher than 30 are gradually constrained to the coeffi-
cients of the static field EIGEN-GL04S (Bruinsma et al. 2010),
thus inconvenient North–South stripes are significantly attenuated.

Temporal variations of the gravity field associated with earth-
quake signals are small in comparison to other contributions, and
aliasing errors. This is the reason why only earthquakes with magni-
tude about 8.8 and higher can be studied from GRACE (Mikhailov
et al. 2004; de Viron et al. 2008), and strategies for earthquake
signal isolation in the GRACE data are needed. To enhance the
earthquake signal in the GRACE data, Panet et al. (2010) thus com-
bined a multiscale filtering of the geoids in the space domain with
an averaging process in the time domain, over various timespans
(Fig. 1). By locally isolating the gravity variations at the earth-
quake spatial scales, this approach allows extracting the earthquake
signal without making any assumption on its temporal variability.
A different signal estimation approach is presented by de Linage
et al. (2009), who assume an exponential time variability of the
post-seismic signal. They estimated associated characteristic time
constants in different parts of the Sumatra area and found an av-
average value around 8.4 months. Whereas the earthquakes that oc-
curred in the Sumatra region since 2004 December, including the
8.3 Nias earthquake of 2005 March, contaminate estimates of the
time-decay constant, the main features of both studies appeared
similar.

To compare the time decays in de Linage et al. (2009) and Panet
et al. (2010) post-seismic gravity relaxation models, we approxi-
mated the temporal variations of the wavelet filtered geoids of Panet
et al. (2010) at different spatial scales by exponential decay func-
tions \( g(t) = A (1 - \exp(-t/\tau)) \), and adjusted the characteristic time
\( \tau \) to fit the wavelet filtered geoid variations at the earthquakes spatial
scales. The estimated values were respectively 8.5 and 8.0 months
for the 1400 and 1000 km scales. This result is consistent with the
8.4 months found by de Linage et al. (2009).

In both papers, the main spatiotemporal features of the post-
seismic temporal variations of the GRGS GRACE models is a pos-
aitive anomaly growth in a broad region including the trench and the
area above the coseismic rupture. The amplitude of the post-seismic
anomaly estimated by Panet et al. (2010) is about 50 per cent of the
amplitude of coseismic minimum in the Andaman Sea. The ampli-
tude of the large-scale positive anomaly centred above the trench
after 26 months, according to de Linage et al. (2009) approaches,
the amplitude of coseismic gravity decrease in absolute value, though
a negative anomaly still remains to the East in result of the different
localizations of both anomalies.

Hence, the three main features of the post-seismic gravity vari-
atations in the region that need to be explained are: (1) growth of
the broad positive anomaly centred above the coseismic rupture
area; (2) exponential time decay with a time constant of about
8 months; (3) the amplitude of the post-seismic anomaly 2.5 yr
after the main event, which according to different estimates comes
up to 50–100 per cent of the coseismic negative anomaly, that is, to
8–15 \( \mu \)Gal.

3 NUMERICAL MODELLING

We adopt 2-D numerical model along a vertical cross-section rep-
resenting the epicentral part of the Sumatra subduction zone. The
size of the rupture surface along the strike is at least 10 times
above its size along the dip and the estimated shear component of
displacement during the Sumatra 25.12.2004 earthquake was rel-
atively small, especially close to the epicentre (e.g. Subarya et al.
2006; Chlieh et al. 2007). Therefore, the 2-D model is appropriate
for the first-order modelling of the downward rupture propagation
and its gravity effect. An important advantage of the 2-D modelling
is its ability to produce significantly better numerical resolution than
large-scale 3-D models. The drawback is that 2-D modelling does
not capture all the details of the Sumatra subduction zone structure
and its evolution, including oblique component of the plate con-
vergence, which mostly accommodates at strike-slip faults running
along Sumatra Island and across the Andaman Sea.
Our modelling approach is subdivided into following stages; (a) long-term geodynamic simulation to produce a self-consistent initial stress and temperature distribution (Section 3.1); (b) short-term simulation that generates damage zones and produces displacements and corresponding dilatation (Section 3.2); (c) final calculation of stress-dependent density distribution to simulate temporal gravity variations (Section 3.3). The second short-term simulation involves three substeps, including (a) additional stress accumulation after part of subduction channel was locked (Section 3.2.1), (b) spontaneous coseismic slip (Section 3.2.2) and (c) post-seismic damage-related aseismic stress relaxation and aftershocks producing afterslip (Section 3.2.3).

The main goal of this study is to model the evolution of the state of stress around rupture area in order to investigate rock damage and fault propagation. Our 2-D model has a depth extent of 200 km. This area is much smaller than the region where post-seismic viscoelastic relaxation occurs after giant earthquakes. Besides, boundary conditions corresponding to ongoing plate motion were assigned at the bottom of the model. As a result our model does not describe post-seismic viscoelastic relaxation at the regional scale, but only damage-related processes localized in the vicinity of the coseismic rupture. Indeed, full-scale modelling of the viscoelastic relaxation for giant earthquakes calls for more global consideration (e.g. Pollitz et al. 2008; Panet et al. 2010). Because of very different spatial scales for viscoelastic relaxation and post-seismic slip, they are often considered separately (e.g. Panet et al. 2010; Hu & Wang 2012), neglecting interactions between both processes.

### 3.1 Long-term simulation (stage 1)

The first modelling stage uses a plane-strain coupled petrological and thermomechanical code, I2ELVIS (Gerya & Yuen 2007), to define the geometry, rock properties and initial stress and temperature distributions as a results of 8 Myr of ongoing subduction. This code has been used to study the earthquake cycle (van Dinther et al. 2013) and extensively analyse long-term deformation phenomena in a subduction settings (e.g. Gerya 2011). It is based on an implicit conservative finite-difference scheme combined with a characteristic-based marker-in-cell technique and solves for the conservation of mass, momentum, and heat using a viscoelastic-plastic Maxwell rheology. Plasticity is implemented using a Drucker–Prager formulation and a viscosity-like parameter, such that the second invariant of the deviatoric stress tensor is corrected towards the pressure-dependent yield stress (Gerya 2010). Hydration (fluid velocity 0.1 cm yr$^{-1}$), sedimentation, erosion and partial melting are included as well (Gerya & Meilick 2011). The numerical grid has dimensions of $1500 \times 200$ km$^2$ and a resolution of 500 m in the zone of interest near the thrust interface ($450 \times 100$ km$^2$) and 2 km outside of it.
The current model setup adapts a more generic active margin subduction setting, also used by Gorczyk et al. (2007) and Gerya & Meilick (2011), by including important regional characteristics of the Sumatran subduction zone (Fig. 2) based on data provided by Klingelhofer et al. (2010). Their seismic refraction and wide-angle reflection surveys provide constraints on the geometry and structure, while a slab age of 60 My and a subduction velocity of 3 cm yr⁻¹ are set after Müller et al. (1997) and Vigny et al. (2005), respectively.

The 2-D numerical model simulates the process of forced subduction of an oceanic plate beneath a continental plate. The size of the modelled area is 200 km in depth and 1500 km in length of which the subducting plate takes up about 900 km and the overriding plate 600 km. In the numerical model, the oceanic crust is represented by 2 km of hydrothermally altered basalt, underlain by 5 km of gabbroic rocks. The lower (15 km) and upper (15 km) layers of the 30-km thick continental crust are felsic in composition. The lithospheric and asthenospheric mantle consists of anhydrous peridotite, which becomes partly hydrated atop the slab during the subduction (Gerya & Meilick 2011). Comprehensive rock properties for these different rock types (see background colour in Fig. 2), including density (P–T dependent), thermal conductivity (P–T dependent) and viscosity (P–T and stress dependent), heat capacity, heat production, internal friction angle, cohesion and shear modulus, are set according to Gerya & Meilick (2011) and Gerya (2010). The resulting temperature and viscosity distributions (contours in Fig. 2) after 8 Myr of long-term numerical model development are exported together with the stress distribution and material properties for each node to the short-term simulation in the next modelling step.

3.2 Short-term simulation (stage 2)

The structure and dimensions of the simulated area as well as the size of the numerical grid elements are almost the same as for the long-term numerical model, while the numerical approach and rock rheology were modified to account for fundamental irreversible aspects of brittle rock deformation including a brittle instability (synthetic seismic events). Most standard numerical approaches fail to simulate a physical instability without numerical instabilities. This is one of the most important advantages of the Fast Lagrangian Analysis of Continua (FLAC) algorithm (Cundall & Board 1988; Cundall 1989; Poliakov et al. 1993; Ilchev & Lyakhovsky 2001) adopted in this study for the short-term simulations. This fully explicit numerical method relies on a large-strain explicit Lagrangian formulation originally developed by Cundall (1989) for an elastoplastic rheology and implemented in the well-known FLAC software produced by ITASCA. A modified version of this code incorporating heat transport is known as PAROVOZ (locomotive in Russian) and is widely used by many researchers. Poliakov et al. (1993) developed an adaptive time stepping scheme and applied the FLAC algorithm for a viscoelastoplastic rheology. Their adaptive scheme does not require iterations, which makes the numerical model stable even for a highly non-linear damage rheology (Lyakhovsky et al. 1993).

The governing equations of the damage rheology model, their thermodynamic basis and experimental verification are presented in Lyakhovsky et al. (1997) and Hamiel et al. (2004). Discussion of the model and numerical results demonstrating evolving fault zone structures have been recently presented by Ben-Zion & Lyakhovsky (2006), Lyakhovsky et al. (2012), Lyakhovsky & Ben-Zion (2009), Finzi et al. (2009), and references therein. The model provides a quantitative treatment of the macroscopic effects of evolving, distributed cracking with local density represented by an intensive damage state variable. It varies from zero for intact rock to one at total failure. The damage model accounts for three general aspects of brittle rock deformation including (1) mechanical response of a solid with an existing crack density, (2) kinetic changes associated with the evolution of the crack density and (3) macroscopic brittle instability (seismic event). This physical framework allows for the simultaneous evolution of damage and its localization into narrow highly damaged zones (faults), earthquakes and associated deformation fields. Synthetic earthquake catalogues generated during the model runs enable analysis of coupled evolution of faults and seismicity pattern.

In a Maxwell element the strains are additive, so the total strain tensor \( \varepsilon_{ij} \) is a sum of strain components associated with different deformation mechanisms

\[
\varepsilon_{ij} = \varepsilon_{ij}^e + \varepsilon_{ij}^d + \varepsilon_{ij}^i,
\]

where \( \varepsilon_{ij}^e \) is elastic strain related to the stress tensor through the damage-related evolving elastic modules (e.g. Lyakhovsky et al. 1997); \( \varepsilon_{ij}^d \) represents ductile strain calculated according the power-law viscosity for the material properties and temperature distribution obtained during the long-term simulation and kept constant during the short-term simulation; \( \varepsilon_{ij}^i \) denotes the damage-related inelastic strain. Detailed analysis of laboratory data (Hamiel et al. 2004) showed that with the onset of acoustic emissions and positive damage evolution, a gradual irreversible (inelastic) strain accumulates before the final macroscopic failure. They suggested that the
rate of the damage-related irreversible strain accumulation is proportional to the rate of positive damage evolution and the characteristic timescale of this process is defined by the rate of the damage accumulation. Ben-Zion & Lyakhovsky (2006) connected the rate of irreversible strain accumulation with partitioning between seismic and aseismic deformation during seismic coupling. They use a non-dimensional ratio, $R$, between the timescale of damage accumulation and the timescale of the damage-related irreversible strain accumulation under given loading conditions and showed that the fraction of elastic strain released during a seismic cycle, that is, the seismic coupling, $\chi$, can be estimated as

$$\chi = \frac{1}{1 + R}. \quad (2)$$

They also demonstrated that the timescale ratio, $R$, or the seismic coupling, $\chi$, is the major factor controlling the aftershock productivity and the rate of aftershock decay. Long aftershock sequences fitted well by Omori law are expected in regions with relatively small $R$ values, or large seismic coupling, $\chi > 50$ per cent. This theoretical prediction is supported by previous estimates of the $R$ value for Westerly Granite (Hamiel et al. 2004) and estimates for Mount Scott granite (Hamiel et al. 2006). For all granite samples the $R$ value falls between 0.3 and 0.6 corresponding to seismic coupling $60 < \chi < 80$ per cent. However, there are no clear constraints for these parameters under relatively high pressures and temperatures typical for the depths of aftershock sequences within the subduction zones. Hence, we present here modelling results using three different values $\chi = 75, 50$ and 25 per cent.

3.2.1 Stress-accumulation stage

The short-term simulation starts with stress, temperature and material property distributions from the above long-term simulation (Fig. 2) and consists of several internal stages. The first stage, explained in this subsection, is the modelling of significant stress accumulation. The hydrated, low-viscosity subduction channel formed during the long-term simulation exhibits interface stresses that are very low, well below the stress level corresponding to the conditions for a large earthquake, such as the $M_{0.5}$ Sumatra one. To accumulate stresses up to an appropriate level, we increased viscosity in the lower part of the low-viscosity subduction channel, thereby locking it and inhibiting ongoing plate motion. In the short-term numerical model we kept damage-free intact rock conditions, that is, damage accumulation was not allowed. We tested three different durations of the stress accumulation stage, running calculations until $\sim 0.75 \times 10^5$ (0.75 per cent), $\sim 10^5$ (1 per cent) and $\sim 1.25 \times 10^5$ (1.25 per cent) strain was accumulated. This is close to or even slightly above the conditions for the onset of failure in the intact rocks under $\sim 100$ MPa confining pressure typical for the brittle upper part of the model. The obtained shear stress and Coulomb Failure Stress (CFS) distribution for 1 per cent of strain are presented in Figs 3 and 4.

Because in our synthetic model we cannot unequivocally locate the main fault corresponding to the Sumatra event, we run two extreme cases locating the fault at the right and the left edge of the shaded rectangle shown in Fig. 3. The difference between the two runs is small. At relatively shallow depths, up to $\sim 50$ km, position of failed elements (interpreted as aftershock location) differs by a few tens of kilometres and is even less at a greater depths. Since we are looking at a large-scale (up to 400 km wavelength) gravity signal and process the data starting half a year after the Sumatra event, we cannot distinguish between runs with different fault locations. Difference in the amplitude of slip is also negligible (see Fig. 4).

3.2.2 Stress release: main seismic event

To model the stress release, we implement an oblique fault zone about 100 km long located within the shaded rectangle in Fig. 3 where damage accumulation was allowed. This leads to a large seismic event with stress drop along the upper part of the bending oceanic plate. The dynamic stress drop during the brittle instability produces a rapid conversion of elastic strain to permanent plastic strain. Lyakhovsky & Ben-Zion (2008) developed a mathematical procedure for the local stress drop that utilizes the Drucker–Prager model, which generalizes the classical Coulomb yield condition for cohesionless material. They use scaling relations between the rupture area and seismic potency values established in seismology and typical range of the stress drop ($\sim 110$ MPa) during earthquakes (e.g. Kanamori & Anderson 1975) to calibrate parameters of their local stress drop procedure. They show that most of the simulated events using their procedure fall within the area bounded by theoretical lines with slope of these lines equal to 2/3 in log–log scale corresponding to theoretical scaling relations derived by Kanamori & Anderson (1975).

To compare the seismic moment obtained in our 2-D model with values estimated from GPS and seismology data we integrated the amplitude of slip along the fault dip and acquired $2.3 \times 10^6$ m$^2$ which gives the seismic moment released in a 2-D cross-section when multiplied to shear modulus. In comparison, the same value for the southern part of the rupture zone varies from 1.4 to $2.9 \times 10^6$ m$^2$ (Banerjee et al. 2007). Therefore, in spite of the several significant simplifications, the obtained seismic moment for $\sim 1$ per cent accumulated strain is close to that reported for the Sumatra earthquake. In the case of 0.75 and 1.25 per cent of the strain the seismic moment of the simulated event is slightly underestimated or overestimated, respectively.

The obtained coseismic stress distribution serves now as an initial condition for the third stage simulating the aftershock sequence for 3 yr.

3.2.3 Post-seismic evolution: aftershock sequence

The last modelling step starts from the stress distribution obtained at the previous stage. Temperature and material properties controlling the ductile (power-law) strain component are almost the same as they were obtained after the first stage of the long-term simulation. The boundary conditions are the same as at the stage of stress accumulation and correspond to ongoing plate motion. For the simulated relatively short period (3 yr), they could be viewed as fixed, except for the free top boundary of the model. Damage accumulation leads to nucleation of synthetic seismic events and their propagation along newly created fault zones. The stress is released by three different mechanisms, including stress drop during seismic events, damage-related aseismic stress relaxation according to the seismic efficiency value and power-law ductile flow. The latest mechanism is, however, only efficient at relatively high temperatures, where ductile flow dominates and prohibits any significant damage accumulation.

Initial stress distribution in all analysed cases (0.75, 1 and 1.25 per cent of accumulated strain) is high enough for the onset of the material failure and gradual damage accumulation, which leads to the formation of several narrow damage zones
Figure 3. (a) Distribution of shear stress along the simulated cross-section at the moment of rupture initiation. Shaded rectangle shows limits for the future coseismic rupture surface. High stresses at the base of the subducting plate result from compression during bending. Note a low-stress wedge between the two plates formed due to transportation of sediments and water along the subduction channel. (b) The Coulomb failure stress. It is similar, but not identical, to those of shear stress.

Therefore, only this case is shown in Fig. 5 demonstrating snapshots after 6, 12 and 36 months with accumulated displacements during the post-seismic period and location of the narrow damage zones. We conclude here that this set of the model parameters seems more realistic. The damage zones marked as 1 and 2 in Fig. 5 were formed in the vicinity of the coseismic rupture of the main event (coseismic rupture marked by red line on the bottom snapshot). Zone 2 is situated in the subducting plate along the downdip extension of the main rupture zone, while the zone 1 is located in the overriding continental plate. These zones are separated by a subduction channel where viscosity was originally lowered by sediment and water transport during the long-term simulation. The maximum displacement occurs in the area of the main rupture in zone 1, at the depth between 30 and 60 km, and in zone 2 it also extends to significant depths below 30 km. In general, these zones are formed at the downdip extension of the main rupture. The less pronounced damage zone 3 is also located in the vicinity of the main rupture and represents slip on an outer-rise normal fault located within the area of...
the first 6 months shear zones 1 and 2 are separated by a block that experiences only very minor deformation. Later on, zone 1 grows and extends into the area of zone 2. After 3 yr (36 months), shear movement in zone 2 occurs in the background of the total downward displacement caused by left-hand side of the zone 1. Finally, all originally localized zones form a strongly deformed area up to 300 km wide around coseismic rupture. The location of this broad and intensively deformed area is compatible with a wide cloud of aftershock activity extending from depths of about 10 km to more than 60 km in Sumatra during 2004–2005 (e.g. Engdahl et al. 2007).

3.3 Comparison with GRACE gravity data

The results of our simulations suggest that formation of damage zones and propagation of coseismic rupture to depths of 30–60 km produces small surface displacements comparing to observed post-seismic GPS displacements in the Sumatra region. The maximum simulated surface displacement in the trench area at the overriding plate relative to its remote eastern edge after 10 months are below 2 cm, while the observed GPS displacements from 2005 January to November excluding effect of Nias earthquake are more than 60 cm (table 1 from Pollitz et al. 2008). As it was mentioned above, we consider fault propagation as additional process to viscoelastic relaxation, which successfully explained amplitude and time decay of the registered surface displacements (Pollitz et al. 2008; Panet et al. 2010) even its gravity effect is noticeably smaller that the gravity anomaly registered by GRACE. Let us consider the gravity effect from modelled damage zones.

The main long wavelength gravity effect (spherical harmonics up to 50) comes from extension and compaction of rocks in the lithosphere and mantle (e.g. Panet et al. 2007). The spatial

Figure 4. Amplitude of post-seismic slip at all rupture surfaces formed in the area of subduction channel for different level of accumulated strain and seismic efficiency $\chi$. For all curves the main fault is located at the left edge of rectangle shown on Fig. 3, except the dotted curve which shows results for a fault located at the right edge. Dark green: 0.75 per cent strain, $\chi = 25\%$; light green: 0.75 per cent strain, $\chi = 50\%$; brown: 1 per cent strain, $\chi = 25\%$; red: 1 per cent strain, $\alpha = 50\%$; dark brown line and dots: 1 per cent strain, $\chi = 75\%$; dark blue: 1.25 per cent strain, $\chi = 25\%$; light blue: 1.25 per cent strain, $\chi = 50\%$.

Figure 5. Displacements in metres at 6 (M6), 12 (M12) and 36 (M36) months after the main coseismic rupture along the thrust interface (red line at the lower plot). Direction is shown by arrows scaled to the maximum value of the each plot. Four damage zones formed after the main event are marked as 1–4. The gravity anomaly ($\mu$Gal) after 36 months at the resolution of the GRACE gravity models is shown at the top.
distribution of volumetric rock deformation is directly extracted from the output of the numerical modelling for every snapshot. The accumulated volumetric rock deformation up to 3 yr after the main event is shown in Fig. 6. Almost the entire area below depths of 40–50 km is under compaction with exception of the low-amplitude region in extension around depths of 100–200 km. Hence in the main part of the region rock density increased and the long-wavelength gravity anomaly corresponding to resolution of the GRACE gravity models is positive (Fig. 5 at the top). Gravity effect from small extensional areas is not seen at the GRACE resolution. Similar to the real data, the positive anomaly covers the area of several 100 km above the coseismic rupture and its amplitude after 36 months exceeds 12 μGal (Fig. 7). The gravity effect at month 26 after the main event is about 40 percent smaller than the real anomaly (red curve in Fig. 7) hence the viscoelastic relaxation is still necessary.

4 DISCUSSION AND CONCLUSION

We present numerical modelling results of post-seismic damage faulting in the area of strong earthquakes to explain fast post-seismic evolution of the gravity anomaly registered by satellites GRACE. Our modelling includes two stages: (a) the long-term tectonic evolution of a subduction zone, which provided the initial distribution of stress, material properties and temperature; (b) a stage of stress accumulation before the main seismic event and its coseismic stress release followed by the post-seismic evolution including the development of fault zones due to rock damage. Evolution of the state of stress at the last stage was used to calculate density variations thus making it possible to calculate temporal variations of the gravity field at the resolution of the GRACE gravity models.

The main earthquake never releases all stresses in vicinity of a locked subduction zone. In particular in the Sumatra area, an active post-seismic process is manifested by high-aftershock activity and continuing displacements recorded by GPS networks, which were reported in many papers (e.g. Paul et al. 2012). The observed post-seismic crustal displacements were fitted well by a viscoelastic relaxation model assuming Burgers body rheology for the asthenosphere (60–220 km deep) with a transient viscosity as low as $4 \times 10^{17}$ Pas and constant $\sim 10^{19}$ Pas steady state viscosity in the 60–660 km depth range. However, this process is not able to explain the registered temporal variations of the gravity field (Panet et al. 2010). To explain the gravity field variations, additional post-seismic slip at the downdip extension of the rupture surface was hypothesized.

In this study we analysed this hypothesis by numerical modelling of three substages of seismic process: stress accumulation, stress release and post-seismic stress evolution leading to downward fault zone propagation. Numerical results demonstrated that the stress

---

**Figure 6.** Accumulated dilatation 36 months after the main coseismic rupture. Density variation in the general part of the area below 50 km is positive, hence the resulting gravity anomaly at GRACE resolution is also positive.

**Figure 7.** Built-up of the positive gravity anomaly (in μGal) above the trench area. Months after the main seismic event are shown by numbers at the right. Red curve—gravity anomaly build up during 26 months (de Linage et al. 2009).
redistribution during the main event leads to the formation of several rupture zones at depth, including the downdip continuation of the main rupture. The damage rheology model with initial conditions close to Sumatra structure and pre-seismic state of stress and with high-seismic efficiency ($\chi = 75$ per cent) successfully describes GPS and seismology based estimates of amplitude of coseismic displacement and seismic moment and suggested amplitude (about 1 m) of post-seismic displacements at the downdip extension of the coseismic rupture. The calculated gravity variations at the post-seismic stage have wavelength of several 100 km, and amplitude up to 12 $\mu$Gal at 36 months after the main event. As a result, rupture propagation explains about 60 per cent of the registered gravity anomaly. It is important to emphasize that rupture propagation at depth predicted by viscoelastic damage rheology model causes only small effects at the surface because the source is distributed between three major fault zones around coseismic rupture, and because of efficient stress relaxation in brittle rock mass due to accumulation of distributed damage and damage-related viscosity. Calculated surface displacements are much smaller than post-seismic displacements registered by GPS network. Hence, to fit both the GPS and GRACE data, this process is suggested to operate together with suggested viscoelastic relaxation, which produces about 50 per cent of the registered gravity anomaly (Panesi et al. 2010).

The main feature of post-seismic evolution is the development of several fault zones in the overriding and subducting plates. Being isolated at the beginning, these zones grow and join together forming finally a broad damage zone covering the area around the main coseismic rupture and extending downwards in the upper mantle. The same modelling procedure can be used to analyse other giant earthquakes, including the 02.27.2010 Maule, Chile, or 11.03.2011 Tohoku-Oki, Japan, earthquakes.

ACKNOWLEDGEMENTS
We thank two anonymous reviewers and Editor Dr. Saskia Goes for thoughtful and helpful comments and suggestions. This study was partly supported by joint CNRS-Russian Foundation for Basic Research grant 12-05-91051. VM and ET acknowledge support of the RFBR by grant 12-05-00276. VL acknowledges support by the US–Israel Binational Science Foundation (Grant 2008248). YD acknowledges support from the Swiss National Science Foundation (SNSF, Grant 200021–125274).

REFERENCES


