Mantle flow distribution beneath the California margin

Sylvain Barbot

Department of Earth Sciences, University of Southern California,
Los Angeles, CA 90089-0740, USA. Email: sbarbot@usc.edu

Abstract

Although the surface deformation of tectonic plate boundaries is well determined by geological and geodetic measurements, the pattern of flow below the lithosphere remains poorly constrained. We use the crustal velocity field of the Plate Boundary Observatory to illuminate the distribution of horizontal flow beneath the California margin. At lower-crustal and upper-mantle depths, the boundary between the Pacific and North American plates is off-centered from the San Andreas fault, concentrated in a region that encompasses the trace of nearby active faults. A major step is associated with return flow below the Eastern California Shear Zone, leading to the extrusion of the Mojave block and a re-distribution of fault activity since the Pleistocene. Major earthquakes in California have occurred above the regions of current plastic strain accumulation. Deformation is mechanically coupled from the crust to the asthenosphere, with mantle flow overlaid by a kinematically consistent network of faults in the brittle crust.

Introduction

Plate tectonics provide a paradigm to understand the shape and dynamics of Earth's surface. Most plate boundaries involve either divergence, like at oceanic spreading centers and continental rifts, or convergence, such as subduction and collision zones, but a small fraction involves transform boundaries that facilitate plate kinematics on the global sphere. These boundaries accommodate plate-parallel relative plate displacement by strike-slip motion on vertical or steeply dipping faults. The San Andreas fault system in California consists of major strike-slip faults and subsidiary thrusts faults that transfer extension at the East Pacific Rise in the Gulf of California spreading center to the south and the relative motion between the Gorda and Pacific plate along the Mendocino fracture zone to the north. Contemporary surface deformation at the California margin is captured by a dense continuous geodetic observatory [1, 2] and the long-term slip accumulation on numerous faults is well established [3], providing a unique opportunity to study the mechanics of strain accumulation at a continental transform. Seismological observations, in particular shear-wave splitting [4, 5] and seismic anisotropy [6], provide unique constraints on the orientation of mantle flow, but a cohesive picture of flow distribution below the margin consistent with surface observations is still missing.

The relative motion between plates is accommodated at intra-continental transforms by various mechanisms of deformation. In the brittle layer, fault slip is the dominant mode of strain accumulation and release, mediated by the complex evolution of frictional resistance that leads to stick-slip and the earthquake and slow-slip phenomena [7, 8]. The bottom of the seismogenic layer is thought to represent an isotherm for the frictional stability of quartz-rich fault gouge [9]. In California, the maximum depth of seismicity varies from 10 to 20 km depending on location [10], inversely correlated with the surface heat flow [11]. Faults are steadily creeping below a locking depth that varies depending on the fault segment [12, 13, 14, 15] and often matches the bottom depth of seismicity [16]. Variations of locking depth across the fault system may be attributed to differences in fluid content, geothermal gradients, and stability of fault slip in the seismogenic zone, and may be non-stationary [17]. Immediately below the seismogenic layer, deformation occurs in localized shear zones that creep steadily down to the brittle-ductile transition, below which the deformation becomes plastic and more distributed, controlled by diffusion creep, grain-boundary sliding, and dislocation creep [18, 19].

The depth of the brittle-ductile transition depends on the geothermal gradient, grain-size, frictional strength, and lithology, and is poorly known. Constraints from mineral physics and numerical modeling indicate the presence of a kilometers-wide shear zone below major strike-slip faults [20, 21, 22, 23, 24], as they provide fluid and heat pathways and stress concentrations, all major weakening constituents. Transient deformation following large California earthquakes indicates distributed lower-crustal flow below fault roots, for example, after the 1992 Landers, 1999 Hector Mine, 2004 Parkfield, and 2010 El Mayor-Cucapah earthquakes [25, 26, 27, 28, 29]. Geophysical data also provide strong evidence for the mobility of the asthenosphere below the California margin, including far-reaching postseismic deformation [27, 30], the development of seismic anisotropy in the upper mantle [31], and strong lateral variations of seismic velocity in the lower-crust and upper-mantle below major strike-slip faults in southern California [32].

The rate of seismicity is controlled by strain accumulation, which is now assimilated to help characterize seismic hazards in California [33, 34]. The long-term kinematics of the California margin is constrained by dating of offset markers along plate boundary faults and the relative motion of tectonic plates at geological time scales. Contemporaneous slip-rates are established with geodetic data using a network of buried faults [1, 35] or block modeling [36, 37, 38, 39, 40]. These methods are of great practical interest because estimation of fault slip-rate is a direct model outcome [41]. However, they make a number of simplifying assumptions that conflict with our understanding of lithosphere mechanics. For example, block modeling gives rise to unrealistic fault-perpendicular offsets due to block rotation, and fault-based models only consider brittle deformation. These models also have difficulty resolving off-fault deformation, which may account for about 50% of the cumulative strain in the Eastern California Shear Zone [42]. Several alternative approaches relax these assumptions, including deformation of a thin plate with variations of elastic thickness [43, 44] and viscoelastic modeling [45, 46], where viscoelastic flow is limited to the postseismic response of large earthquakes. None of these studies reveal how plastic strain is accumulated in the ductile substrate on an on-going basis.

We seek to image the deep deformation that accrues along the California margin compatible with our understanding of lithosphere rheology. We present a method to image state-wide horizontal mantle flow from geodetic data with minimal assumptions and tradeoffs. At long wavelengths, the tectonic contribution of the velocity field at Earth's surface during the interseismic period is caused by fault creep in the brittle layer and distributed flow in the viscoelastic substrate. In a viscoelastic medium,

any parcel that deforms plastically entrains the surrounding rocks by elastic coupling [47, 48, 49, 50]. We exploit this effect to reconstruct the subsurface strain-rate distribution by inversion of the surface velocity field (see Methods). We find that the San Andreas fault system represents the extension of a 50 km-wide trans-lithospheric shear zone in the brittle crust. Major earthquakes have occurred in regions with rapid plastic strain accumulation and the orientation of the principal strain axes is compatible with the earthquake source mechanisms. The numerous parallel dextral faults in Southern California are associated with a return flow below the Mojave section of the Eastern California Shear Zone.

Results

Modeling assumptions

We obtain the velocity field from the Plate Boundary Observatory through the UNAVCO (`unavco.org`) repository (Figure 1). The velocity field is representative of the deformation that accrues for at least 2.5 years up to September 15, 2018, and may include a contribution from the continuing postseismic relaxation of the 1992 Landers, 1999 Hector Mine, and 2010 El Mayor-Cucapah earthquakes. We assume that the deformation is mainly caused by distributed plastic flow in the lower crust and upper mantle. We devise a mesh of $20{\times}20$ km semi-infinite volume elements that extend down from 20 km depth along a coordinate system aligned with the plate boundary axis, with x_1 and x_2 the parallel and perpendicular axes, respectively. We assume that the viscous flow is horizontal and incompressible. This implies that the vertical plastic strain-rate tensor components are zero and that $\dot{\epsilon}_{11} + \dot{\epsilon}_{22} = 0$. The model ignores vertical shear, which accommodates the absolute motion of the plate boundary in a plume-fixed reference system, as this displacement is undistinguishable from a rigid-body translation at the regional scale. The maximum shear strain-rate is presumably aligned with the plate boundary axis, but small rotations are allowed by considering both $\dot{\epsilon}_{12}$ and $\dot{\epsilon}_{11}$ free parameters, strictly enforcing $\dot{\epsilon}_{22} = -\dot{\epsilon}_{11}$ and $\dot{\epsilon}_{12} \leq 0$ to ensure overall right-lateral shear parallel to the plate boundary and left-lateral shear perpendicular to the plate boundary, and incompressibility. The mesh extends from Baja California, Mexico to north of the Mendocino triple junction, encompassing $74{\times}33{=}2442$ volume elements.

To enforce kinematic continuity from the locking depth in the crust to the asthenosphere, we allow strike slip at the deep extension of major faults in the brittle layer, from 15 to 20 km. These embedded surface elements include the entire length of the Garlock, San Jacinto, Elsinore faults, a large section of the San Andreas fault from Gilroy in the Bay Area down to the Salton trough, and the southern section of the Calaveras fault. In addition, we allow shallow fault slip in the creeping section of the San Andreas fault, north of Parkfield. Variation of slip velocity on these deep segments reflect changes in locking depth along a fault and the possibility of a gradual transition from locked to creeping, and do not always represent the long-term fault slip-rate. Large deformation occurs around the Long Valley caldera due to volcanic unrest [51]. Instead of removing the surrounding GPS stations, we model this deformation explicitly with shallow volume elements representing shear and volume change associated with a network of dyke, magma chambers, and connected faults.

Inversion method and resolution

Imaging of localized and distributed deformation can be cast as an inverse problem (see Methods) where plastic strain-rate in volume elements and slip velocity on surface elements are related to surface GPS velocity vectors by Green's functions [52, 53, 54, 55, 56]. The formulation enforces that all the strain tensor components derive from a single-valued velocity field with conservation of angular and linear momentum in a half space [49, 50]. We also simultaneously fit for a rotation and a translation such that the results are independent of the reference frame of the velocity field. The inversion is under-determined with 3 components for each of the 937 GPS stations, summing up to 2811 data points and 5578 model parameters, among which 588 are dedicated to resolving deformation below Long Valley caldera. Stable inversion results can be obtained with first-order Tikhonov regularization with or without positivity constraints, but all results shown derive from a non-linear inversion that imposes a negative shear strain-rate component $\dot{\epsilon}_{12}$.

The imaging power provided by the Plate Boundary Observatory is shown in a checkerboard test in Figure 2, indicating that regions of strain accumulation in the lower-crust and upper mantle roughly 50 km wide can be resolved appropriately. Loss of resolution in the Sierra Nevada centered on the Long Valley caldera is due to the sparsity of the Plate Boundary Observatory geodetic network in the region and the challenging trade-offs associated with resolving volcanic unrest and deep mantle flow. Inspection

of the resolution matrix reveals a spatial pattern similar to the checkerboard test. As the data offer limited resolution at depth, we do not image how the flow changes across the Moho or how the shear zone may broaden in the asthenosphere. The model therefore represents an average strain accumulation, perhaps most representative of the asthenosphere due to its greater depths and dimensions leaving a broad footprint in the velocity field.

Flow distribution

The inversion of the Plate Boundary Observatory velocity field reveals the pattern of horizontal flow in the lower-crust and upper-mantle (Figure 3). Plastic strain accumulation during the interseismic period is heterogeneous, with large strain-rates of the order of 1×10^{-14}/s localized below major faults, but considerably smaller far from the plate-boundary faults. Deep shear zones are concentrated underneath major faults, spreading across a $50\,\mathrm{km}$-wide shear zone at the minimum, implying substantial variations of ambient stress or rheological properties across the margin. Strain may be even more localized than currently resolved, but filtering by the elastic crust eventually limits the recovery of finer details. We surmise that the shear zone is narrower in the lower-crust, broadening to a wider region in the asthenosphere, following theoretical predictions [20, 21]. The deep regions of strain concentration may reflect strain softening due to grain-size evolution, the development of plastic anisotropy, or thermo-mechanical feedbacks with the overlaying brittle section.

Mantle flow is more localized throughout northern and central California, down to the Garlock fault, and in the Salton trough, from the San Bernardino mountains to the East Pacific Rise, with a $120\,\mathrm{km}$ wide restraining step-over separating these two regions. The intervening space corresponds to a broad, spatially coherent shear zone, widely distributed from west to east from the Channel Islands offshore to the Eastern California Shear Zone in the Mojave block, extending as far north as the White Wolf fault and the Little Lake fault zone, across the Garlock fault. Across the central section of the San Andreas fault, the deformation resembles Couette flow on a horizontal plane.

The northern termination of the California margin exhibits a strong horizontal return flow around the Gorda slab, which may overprint an unresolved vertical corner flow due to plunging of the slab below the North American plate. Another return flow in southern California is more widely distributed due to the large extent of the restraining bend in the plate-parallel and plate-perpendicular directions. Flow is more uniform away from the San Andreas fault system. Smaller patches of strain accumulation can be found along Owens Valley, Panamint Valley, and Death Valley, all the way to the Walker Lane in Nevada, but low resolution around Long Valley caldera affects the imaging result in this region. A patchy, low-amplitude, background plastic strain-rate persists throughout the California margin from the Pacific plate to the Basin and Range, presumably reflecting the ambient tectonic stress or small-scale mantle convection driven by density anomalies that are not isostatically compensated [57].

Throughout California, the plate boundary at depth is not exactly aligned with the surface trace of the San Andreas fault, except in the Central Valley, as previously established from the distribution of surficial geodetic strain-rate [14]. To the north, the shear zone is firmly east of the San Andreas fault, spreading across the Maacama and Bartlett Springs fault systems in Mendocino County and centered on the Hayward, Calaveras, and Greenville fault system underneath the Bay Area. In the Salton trough, a zone of high strain concentration spreads across the Coachella segment of the San Andreas fault and the San Jacinto fault, contiguous to a broad secondary zone of more distributed strain accumulation.

Residual velocity

The modeling assumptions are compatible with observations at the California margin, with a sum of squares of residuals of $(0.3\,\mathrm{mm/yr})^2$ on average for each station and an overall variance reduction of 95.5%. The vertical component of velocity is typically smaller and has less influence on the result (Supplementary Figure 1).

The residuals with the Plate Boundary Observatory velocity field in the horizontal direction are shown in Figure 4. Systematic and spatially coherent residuals in the horizontal components can be found near the Mendocino triple junction, near the Parkfield segment of the San Andreas fault, in the Transverse Range and the Los Angeles basin, and in the near-field of recent large earthquakes in Southern California, shown in more details in Figure 5. Strong subsidence in the Central Valley associated with hydrological processes also produce outliers (Supplementary Figure 1). These residuals occur because of shallow processes not accounted for in the model that cannot be explained by horizontal plastic flow in the lower-crust and mantle.

In northern California, the tectonic regime changes from strike-slip faulting on the San Andreas fault system to subduction of the Gorda plate underneath north America at the Cascadia trench. The strain associated with the Gorda down-going slab creates trench-perpendicular residuals around the Klamath mountains in Humboldt County. Other residuals in the Northern Coast Range may be due to hydrological activity [58]. Misfits near Parkfield correspond to small scale variations of creep between the Cholame and creeping segments modulated by postseismic relaxation of the 2004 Parkfield earthquake [59, 60]. The shortening centered on the Oak Ridge fault system and the San Gabriel fault in the Los Angeles Basin is due to slip partitioning around the big bend of the San Andreas fault [61, 62] involving shallow thrusts splaying off a deeper sub-horizontal décollement [63].

Deformation in the Transverse Range therefore represents an example of thin tectonics, i.e., faults confined to the crust without mantle roots, instead of deformation driven by mantle down-welling [57]. Other large residuals in the Los Angeles Basin are due to extraction and injection of groundwater and oil [64]. Finally, the forward model leaves out residuals around recent large earthquakes in southern California due to afterslip and other postseismic processes that were still ongoing in 2018. The postseismic deformation of large strike-slip earthquakes includes vertical shear and short-wavelength features that are not captured in the model. These are most evident in the near-field of the 1999 Hector Mine earthquake and in the Salton trough, north of the El Mayor-Cucapah earthquake rupture, where slip on many shallow faults in the Imperial Valley was accelerated after the mainshock [29].

Relation with faulting and seismicity

The pattern of strain accumulation provides information about seismo-tectonic processes (Figures 3 and 6), including the background stress, as the principal directions of strain-rate are identical to those of the driving deviatoric stress in the ductile region, and the direction of mantle flow at asthenosphere depths. In Mendocino county, the San Andreas fault is poorly oriented with the principal strain direction, contrarily to other faults inland. The strain principal directions are compatible with right-lateral strike-slip faulting on the San Andreas fault from Marin County, north of the Bay Area, to the triple junction at the Garlock fault and, further south, along the Coachella segment near the Salton Sea. Along the big bend of the San Andreas fault, the plastic strain-rate is smaller and less favorably oriented for strike-slip faulting, potentially compensated for by a weaker frictional resistance. The plastic strain-rate below the Garlock fault is optimally oriented for left-lateral strike-slip faulting from the junction with the San Andreas fault to Ridgecrest and Searles Valley, after which a restraining bend of the Garlock fault makes it less favorably oriented, compatible with the slip-rate of the Garlock decreasing to the east [65]. In the Mojave desert, plastic strain-rate is optimally oriented for right-lateral slip for all of the Helendale, Lenwood, Emerson, Camp Rock, Blackwater, Calico, Pisgaw and Ludlow faults that together give rise to bookshelf faulting.

The background stress is also compatible with the source mechanisms of recent large Southern California earthquakes, including left-lateral and right-lateral slip for the 2019 Ridgecrest Mw 6.4 foreshock and Mw 7.1 mainshock, respectively [66], left-lateral and right-lateral slip for the 1992 Big Bear and Landers earthquakes, respectively, and right-lateral slip for the 1999 Hector Mine earthquake. The compressive axis of crustal seismicity [67] is perpendicular to major thrust faults in the Transverse Range and oblique to strike-slip faults in the Mojave block, but generally aligned with the compressive axis of plastic strain-rate. An even better alignment is found with the surface geodetic strain [67], indicating a small rotation of the stress tensor towards the surface, presumably due to the short-range influence of faulting in the brittle layer. These results indicate that deep deformation dictates the large-scale stress patterns in the brittle layer that can be relaxed by different types of faults of various orientations [61], each producing earthquakes with a specific focal mechanism.

Discussion

Geodetic constraints on lower-crustal and mantle strain accumulation illuminate how shallow brittle crust and the deeper plastic layer accommodate regional deformation consistently with significant interaction. Plastic strain accumulation in northern California spreads from west to east across the San Andreas, Maacama, and Bartlett Springs fault systems in the Coastal Range and across the San Andreas, Hayward, and Calaveras fault systems in the Bay Area. In the central segment, the San Andreas fault sits above a similarly wide shear zone that stretches from the Oceanic and Rinconada fault zones to the west to the Great Valley thrust system to the east. In southern California, the deep shear zone is distributed over a wide region that spans across the Newport-Inglewood-Rose Canyon, Elsinore, and San Jacinto

fault systems and the Coachella section of the San Andreas fault, possibly extending farther offshore. These results are compatible with the view that a fault network coalesces to a broad zone of distributed flow in the ductile layer [68], with a separation distance for faults commensurate with the depth of the brittle layer. Theoretical considerations indicate that the thickness of parallel fault strands depends on the viscosity contrast between the lower and upper crusts [69]. However, the partitioning of fault slip to accommodate the underlying mantle flow can be accomplished by a set of parallel strike-slip faults, like in the northern and southern ends of the San Andreas fault system, or by a combination of strike-slip and en échelon folds and reverse faults, like in the central section. These results contrast with the simplistic view that major faults exist in isolation and extend from the surface to great sub-crustal depths with a similar degree of localization.

The broad deformation zone associated with a major restraining bend at asthenosphere depths impacts California tectonics from the Transverse Range to the Eastern California Shear Zone. A wide return flow surrounds the northern termination of the Salton trough shear zone, extending to north of the Garlock fault (Figures 3 and 6b), where the flow direction rotates from SE-NW on the Pacific side, to W-E below the Mojave desert, to NW-SE on the North American side. The flow pattern may shed light on the orientation of shear-wave splitting in southern California. The development of seismic anisotropy in the upper mantle is thought to represent lattice preferred orientation of olivine minerals along the direction of maximum shear [70], which can be estimated assuming a stationary flow by integrating the plastic strain-rate along the streamlines [71]. Considering a stable configuration of the San Andreas fault system in the last 4-6 Myr with particle velocity of 25 mm/yr and an average strain-rate of 5×10^{-15}/s corresponds to 90-120 km of particle motion along the current streamlines building up about 95% plastic strain, enough to eliminate any previous texture. Computing synthetic seismic anisotropy is outside the scope of this work, but if the plastic strain-rate is sufficiently uniform upstream of the streamline, the direction of maximum shear of the current plastic strain-rate may represent a good proxy for the fast axis of shear wave splitting. From the California Borderland and Peninsular Range to the Mojave desert, where these conditions are met, there is good agreement with seismic observations [31]. However, shear-wave splitting measurements have low depth resolution and may be sensitive to deformation below the asthenosphere, like a counter flow driven by the sinking of the Farallon slab [71].

The broad return flow below the most major step-over of the San Andreas fault also drives the eastward extrusion of the Mojave block. The advection of the eastern blocks of the San Andreas-Garlock unstable triple junction explains the formation of a new left-lateral shear zone along the White Wolf fault zone, which hosted the Mw=7.3 Kern County earthquake in 1952 [72], and may eventually supplant the role of the Garlock fault for the regional block kinematics (Figure 7). The extrusion of the Mojave block perpendicular to plate motion at a rate of ~17 mm/yr is responsible for the peculiar long-term evolution of fault activity in the Mojave desert with increasing cumulative slip from west to east, as faults move past a stable sub-crustal shear zone [73]. The easternmost fault, the Ludlow fault, which simultaneously holds the largest cumulative offset and the smallest Holocene slip-rate, is no longer aligned with the Salton trough shear zone and is currently inactive. In contrast, the Calico and other faults to the west are currently still aligned with the Salton trough shear zone and have rapid slip-rates [42]. Even though a secondary zone of strain accumulation seems to connect the Eastern California Shear Zone and the Walker Lane with an average strain-rate of the order of 1×10^{-15}/s, the major plate-boundary shear zones sit beneath the Salton trough and the Central section of the San Andreas fault, separated by a 120 km-wide restraining bend. The alignment of Mojave block faults with the deep Salton trough shear zone appears to be controlling their slip history.

The rate of seismicity on mature faults is controlled, among others, by the rate of stress accumulation, the size of the unstable-weakening region, and the style of faulting, all of which challenging to unravel. Plastic strain accumulation at lower-crustal and upper-mantle depths contributes to loading the seismogenic layer by elastic coupling. In addition, kinematic boundary conditions require the long-term relative displacement of blocks bounding the shear zone to be the same in the brittle and ductile layers. We investigate this relationship by comparing the long-term fault slip-rates on vertical or sub-vertical strike-slip faults collected in Appendix B of the Uniform California Earthquake Rupture Forecast [3] with the offsets of the underlying volume elements in the fault-perpendicular direction. There is a broad agreement between geological slip rates and integrated plastic strain-rate (Figure 8), even though mantle shear zones are diffuse, with no strict boundaries. Simple predictions based on underlying plastic strain-rate fall within the uncertainties of field observations for the Hayward, Calaveras, north San Andreas, south San Andreas, Newport-Inglewood, Elsinore, and San Jacinto faults. Large deviations occur for faults of the Eastern California Shear Zone, presumably due to off-fault deformation in the Mojave block [42] and the variable rates of the Garlock fault in the late Pleistocene, early Holocene [65]. The overall agreement

between shallow (geological) and deep deformation confirms the kinematic compatibility between the brittle and ductile layers, implying a strong mechanical coupling between upper-crust deformation and lower-crust and asthenosphere flow.

Assuming that plastic strain corresponds dominantly to simple shear, which holds true for most of California (Figure 3), integration of the plastic strain in the plate-perpendicular direction shows a relative velocity between the Pacific and the North American plates in agreement with global reconstructions of relative plate motion, i.e., 49.7±2.7 mm/yr (Figure 9). Local deviations within the uncertainties occur along the plate boundary along the extent of the 1906 Mw=7.9 San Francisco earthquake rupture, the Mojave block, and the Salton trough, which may be attributed to transient acceleration of mantle flow following large earthquakes, and in the Eastern California Shear Zone, where deformation is not in simple shear, involving a significant component of plate-perpendicular motion.

Plastic strain accumulation below the California margin is compatible with the relative motion of the North American and Pacific plates at geological time scales, with the long-term slip-rate of major faults of the San Andreas fault system, and with the contemporaneous surface deformation during the interseismic period. In addition, the trace of major faults, background seismicity, and major earthquakes are situated above regions of high plastic strain-rate (Figure 9). These results indicate that deformation at the plate boundary is mechanically coupled from the crust down to the upper mantle, despite the possible weak strengths of the lower-crust and the asthenosphere. Mantle flow and crustal faulting form a cohesive geometric assembly where deep shear zones are overlaid by a kinematically consistent network of faults in the brittle layer. Kinematic compatibility is accomplished by slip partitioning in the brittle layer and strain partitioning in the ductile substrate, with geographic overlap.

As models of crustal dynamics can now represent the coupling between brittle and ductile processes [74, 75], imaging of deep deformation opens the door to more accurate models of crustal dynamics that assimilate a realistic distribution of mantle flow. Integration of brittle and ductile deformation processes into self-consistent mechanical representations will be key to better understand the geodynamics of the California margin.

Acknowledgments

Methods

Green's functions

Any distribution of plastic strain in a viscoelastic medium causes deformation in its surrounding by elastic coupling. The details of the rheology in the ductile region is unspecified, but we assume that the flow is incompressible, driven by the deviatoric component of the stress tensor. With these assumptions, the plastic strain tensor $\boldsymbol{\epsilon}$ can be associated the moment density $\mathbf{m} = 2\mu\,\boldsymbol{\epsilon}$, where μ is the local shear modulus, and the equivalent body-force

$$\mathbf{f} = -\nabla \cdot \mathbf{m} \ . \tag{1}$$

The resulting total displacement satisfies conservation of linear and angular momentum and can be obtained at any GPS point $\mathbf{x}$ of the domain by integration with the Green's functions $\mathbf{G}(\mathbf{x}; \mathbf{y})$, where $\mathbf{y}$ is the location of the forcing term, to automatically conform with the boundary conditions, following

$$\mathbf{u}(\mathbf{x}) = \int_{\Omega} \mathbf{G}(\mathbf{x}; \mathbf{y}) \cdot \mathbf{f}(\mathbf{y}) \ \mathrm{d}\mathbf{y} \ , \tag{2}$$

where Ω is the region that deforms plastically, potentially the whole domain. To image the distribution of plastic strain accumulation in California compatible with an inverse problem, we discretize the domain into semi-infinite cuboids that extrude downwards from a burial depth of $q_3 = 20\,\mathrm{km}$. We approximate the total deformation field by superposition of the deformation caused by individual, non-overlapping

volume elements of uniform plastic strain. Considering one volume element, the displacement field is given by

$$u_i(x_1, x_2, x_3) = \int_{\partial\Omega} G_{ji}(x_1, x_2, x_3; y_1, y_2, y_3)\, m_{jk} n_k\, \mathrm{d}y_1\, \mathrm{d}y_2\, \mathrm{d}y_3 \ , \tag{3}$$

where $\partial\Omega$ represents the boundary of the volume element, and repeated indices are summed, following Einstein's convention. We use the Green's function for an elastic half-space and the coordinate system is aligned with the orientation of the semi-infinite cuboid, depth positive down. We consider cuboids of length L and thickness T in the x_1 and x_2 directions, respectively, with the origin centered in the middle of a top edge. Expanding the integration over the four vertical surfaces and the top horizontal one, we can write

$$\begin{aligned}
u_i(x_1, x_2, x_3) = &\int_{-T/2}^{T/2} \int_{q_3}^{\infty} m_{j1} \Big[G_{ji}(x_1, x_2, x_3; y_1, y_2, y_3) \Big]_{y_1=0}^{L} \, \mathrm{d}y_2\, \mathrm{d}y_3 \\
&+ \int_{0}^{L} \int_{q_3}^{\infty} m_{j2} \Big[G_{ji}(x_1, x_2, x_3; y_1, y_2, y_3) \Big]_{y_2=-T/2}^{T/2} \, \mathrm{d}y_1\, \mathrm{d}y_3 \\
&- \int_{0}^{L} \int_{-T/2}^{T/2} m_{j3}\, G_{ji}(x_1, x_2, x_3; y_1, y_2, q_3) \, \mathrm{d}y_1\, \mathrm{d}y_2 \ .
\end{aligned} \tag{4}$$

We develop solutions for the displacement field using a numerical integration involving Gauss-Legendre quadrature along the x_1 and x_2 axes and double-exponential quadrature along the depth axis. For integration over depth, we use a change of variable that maps the semi-infinite interval into finite bounds. The displacement for semi-infinite cuboids of arbitrary strike angle from north at any coordinates in the reference system is obtained by translation and rotation of the displacement field. Surface displacement are associated with plastic strain; surface velocity with plastic strain-rate. Numerical codes that evaluate the displacement field are provided (see Code Availability). The displacement fields for a buried volume element with non-zero plastic strain tensor components e_{11} and e_{12} are shown in Supplementary Figure 2. Each volume element produces a similar displacement, accordingly shifted. Imaging the plastic strain accumulation can be thought of as a deconvolution of these displacement patterns. The diffusion of the surface velocity compared to the more localized extent of the plastic deformation represents the filtering effect of the elastic crust, which eventually limits the resolution of the inverse problem.

Geodetic inversion method

We design the inversion of the surface velocity field by constructing the displacement kernels for each volume element of the mesh, complemented by other deforming surface and volume elements described in the main text. The matrices $\mathbf{L}_{11}$, $\mathbf{L}_{12}$, and $\mathbf{L}_{22}$ represent the three-component surface velocity vectors at the coordinates of the GPS stations due to the plastic strain tensor components $\dot{\mathbf{e}}_{11}$, $\dot{\mathbf{e}}_{12}$, and $\dot{\mathbf{e}}_{22}$, respectively, organized in a vector with one element per volume element in the mesh. For any distribution of horizontal plastic strain, the corresponding velocity field is given by the matrix-vector product $\mathbf{L}\,(\dot{\mathbf{e}}_{11}^T, \dot{\mathbf{e}}_{12}^T)^T$, where the Green's function matrix is defined as

$$\mathbf{L} = (\mathbf{L}_{11} - \mathbf{L}_{22} \qquad \mathbf{L}_{12}) \ , \tag{5}$$

enforcing an equal and opposite contribution of east-west and north-south uniaxial strain accumulation. The second-order deviatoric tensor $(\mathbf{e}_1 \otimes \mathbf{e}_1 - \mathbf{e}_2 \otimes \mathbf{e}_2)$ is simply a 45° rotation of the tensor $(\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1)/\sqrt{2}$, so inverting for these two orientations simultaneously provides a basis to capture horizontal pure shear in any orientation. The velocity field contains a rigid-body motion representing rotation about an unknown pole. However, rotation about any arbitrary pole can be decomposed into a rotation about the origin of the coordinate system and a residual translation, both of which can be represented by a linear operator that can be inverted. We represent the rotation about the origin by the spin vector $\mathbf{w} = -w\,\mathbf{e}_3$ and the remaining translation by the vector $\mathbf{u}$. The rigid-body component of the surface velocity is then given by the matrix-vector product $\mathbf{R}\,(w, u_1, u_2)^T$, with

$$\mathbf{R} = \begin{pmatrix} -\mathbf{x}_1 & 1 & 0 \\ +\mathbf{x}_2 & 0 & 1 \end{pmatrix} \ . \tag{6}$$

Fault creep between the seismogenic zone and the brittle-ductile transition of major California faults is captured with 55 embedded surface elements, 5 km-wide and of varying lengths, with strike-slip motion. Near-surface creep along a central segment of the San Andreas fault is captured with 48 shallow embedded

surface elements, 20 km-long and 5 km wide, with strike-slip motion. Deep and shallow creep are related to the surface displacement by the matrices $\mathbf{F}$ and $\mathbf{C}$, respectively. Additional volume elements with free plastic strain components $\dot{e}_{11}$, $\dot{e}_{12}$, $\dot{e}_{22}$ and $\dot{e}_{33}$ organized in a $7 \times 7 \times 3$ element mesh extending $70 \times 70 \times 15$ km around Long Valley caldera are used to model volcanic unrest and $\mathbf{M}$ is the matrix used to compute the associated displacements. The forward model is then given by $\mathbf{d} = \mathbf{Gm}$, where $\mathbf{d}$ is the vector of observations and the design matrix is given by $\mathbf{G} = (\mathbf{L}, \mathbf{F}, \mathbf{C}, \mathbf{M}, \mathbf{R})$. The inversion is stabilized with first-order and second-order Tikhonov regularization. A nonlinear inversion procedure is used to enforce right-lateral shear in the plate-perpendicular direction and left-lateral shear in the plate-parallel direction. The residuals with the PBO crustal velocity field is shown in Figure 4 and Supplementary Figure 1.

Data Availability

All data used in this study are publicly available. A reference to each dataset can be found in the figure captions and in the main text. The source data underlying Figs. 1, 3, 4, 5, 8a, and 9 and Supplementary Fig. 2a are provided as a Source Data file.

Code Availability

The Matlab routines to compute Green's functions for plastic strain in a semi-infinite cuboid in an elastic half-space are available at `bitbucket.org/sbarbot/natcomm20`.

Author Contributions

The Author conducted all the analysis and wrote the manuscript.

Competing Interests

The Author declares no competing interests.

References

[1] Y. Zeng and Z.-K. Shen, "A fault-based model for crustal deformation, fault slip rates, and off-fault strain rate in California," *Bull. Seism. Soc. Am.*, vol. 106, no. 2, pp. 766–784, 2016.

[2] E. Klein, Y. Bock, X. Xu, D. T. Sandwell, D. Golriz, P. Fang, and L. Su, "Transient deformation in California from two decades of GPS displacements: Implications for a three-dimensional kinematic reference frame," *J. Geophys. Res.*, vol. 124, pp. 12189–12223, 2019.

[3] E. H. Field, K. R. Milner, J. L. Hardebeck, M. T. Page, N. van der Elst, T. H. Jordan, A. J. Michael, B. E. Shaw, and M. J. Werner, "A spatiotemporal clustering model for the third Uniform California Earthquake Rupture Forecast (UCERF3-ETAS): Toward an operational earthquake forecast," *Bull. Seism. Soc. Am.*, vol. 107, no. 3, pp. 1049–1081, 2017.

[4] P. Silver and W. Holt, "The mantle flow field beneath western North America," *Science*, vol. 295, no. 5557, pp. 1054–1057, 2002.

[5] G. Zandt and E. Humphreys, "Toroidal mantle flow through the western US slab window," *Geology*, vol. 36, no. 4, pp. 295–298, 2008.

[6] M. Moschetti, M. Ritzwoller, F. Lin, and Y. Yang, "Seismic evidence for widespread western-US deep-crustal deformation caused by extension," *Nature*, vol. 464, no. 7290, pp. 885–889, 2010.

[7] S. Barbot, "Modulation of fault strength during the seismic cycle by grain-size evolution around contact junctions," *Tectonophysics*, vol. 765, pp. 129–145, 2019.

[8] S. Barbot, "Slow-slip, slow earthquakes, period-two cycles, full and partial ruptures, and deterministic chaos in a single asperity fault," *Tectonophysics*, 2019.

[9] M. L. Blanpied, D. A. Lockner, and J. D. Byerlee, "Frictional slip of granite at hydrothermal conditions," *J. Geophys. Res.*, vol. 100, no. B7, pp. 13045–13064, 1995.

[10] J. J. Nazareth and E. Hauksson, "The seismogenic thickness of the southern California crust," *Bull. Seism. Soc. Am.*, vol. 94, no. 3, pp. 940–960, 2004.

[11] E. Hauksson, "Crustal geophysics and seismicity in southern California," *Geophys. J. Int.*, vol. 186, no. 1, pp. 82–98, 2011.

[12] J. Sauber, W. Thatcher, S. C. Solomon, and M. Lisowski, "Geodetic slip rate for the eastern California shear zone and the recurrence time of Mojave Desert earthquakes," *Nature*, vol. 367, no. 6460, pp. 264–266, 1994.

[13] B. Smith and D. Sandwell, "Coulomb stress accumulation along the San Andreas Fault system," *J. Geophys. Res.*, vol. 108, p. 10.1029/2002JB002296, 2003.

[14] J. P. Platt and T. W. Becker, "Where is the real transform boundary in California?," *Geochemistry, Geophysics, Geosystems*, vol. 11, no. 6, 2010.

[15] E. O. Lindsey, V. J. Sahakian, Y. Fialko, Y. Bock, S. Barbot, and T. K. Rockwell, "Interseismic strain localization in the San Jacinto fault zone," *Pure Appl. Geophys.*, vol. 171, no. 11, pp. 2937–2954, 2014.

[16] B. R. Smith-Konter, D. T. Sandwell, and P. Shearer, "Locking depths estimated from geodesy and seismology along the San Andreas Fault System: Implications for seismic moment release," *J. Geophys. Res.*, vol. 116, no. B6, 2011.

[17] J. Jiang and N. Lapusta, "Deeper penetration of large earthquakes on seismically quiescent faults," *Science*, vol. 352, no. 6291, pp. 1293–1297, 2016.

[18] D. L. Kohlstedt, B. Evans, and S. J. Mackwell, "Strength of the lithosphere: Constraints imposed by laboratory experiments," *J. Geophys. Res.*, vol. 100, pp. 17587–17602, 1995.

[19] G. Hirth and D. L. Kohlstedt, "Rheology of the upper mantle and the mantle wedge: a view from the experimentalists," in *Inside the Subduction Factory* (J. Eiler, ed.), vol. 138 of *Geophys. Monogr.*, pp. 83–105, Washington, D. C.: Am. Geophys. Soc., 2003.

[20] P. England, G. Houseman, and L. Sonder, "Length scales for continental deformation in convergent, divergent, and strike-slip environments: Analytical and approximate solutions for a thin viscous sheet model," *J. Geophys. Res.*, vol. 90, no. B5, pp. 3551–3557, 1985.

[21] W. Thatcher and P. C. England, "Ductile shear zones beneath strike-slip faults: Implications for the thermomechanics of the San Andreas fault zone," *J. Geophys. Res.*, vol. 103, no. B1, pp. 891–905, 1998.

[22] C. S. Takeuchi and Y. Fialko, "Dynamic models of interseismic deformation and stress transfer from plate motion to continental transform faults," *J. Geophys. Res.*, vol. 117, no. B5, 2012.

[23] K. L. Allison and E. M. Dunham, "Earthquake cycle simulations with rate-and-state friction and power-law viscoelasticity," *Tectonophysics*, vol. 733, pp. 232–256, 2018.

[24] X. Zhang and T. Sagiya, "Intraplate strike-slip faulting, stress accumulation, and shear localization of a crust-upper mantle system with nonlinear viscoelastic material," *J. Geophys. Res.*, vol. 123, no. 10, pp. 9269–9285, 2018.

[25] F. F. Pollitz, G. Peltzer, and R. Bürgmann, "Mobility of continental mantle: Evidence from post-seismic geodetic observations following the 1992 landers earthquake," *J. Geophys. Res.*, vol. 105, pp. 8035–8054, Apr. 2000.

[26] A. M. Freed and R. Bürgmann, "Evidence of power-law flow in the Mojave desert mantle," *Nature*, vol. 430, pp. 548–551, 2004.

[27] A. M. Freed, R. Bürgmann, and T. Herring, "Far-reaching transient motions after Mojave earthquakes require broad mantle flow beneath a strong crust," *Geophys. Res. Lett.*, vol. 34, no. 19, 2007.

[28] L. Bruhat, S. Barbot, and J. P. Avouac, "Evidence for postseismic deformation of the lower crust following the 2004 Mw6.0 Parkfield earthquake," *J. Geophys. Res.*, vol. 116, p. 10, 2011.

[29] C. Rollins, S. Barbot, and J.-P. Avouac, "Postseismic deformation following the 2010 M=7.2 El Mayor-Cucapah earthquake: Observations, kinematic inversions, and dynamic models," *Pure Appl. Geophys.*, vol. 172, no. 5, pp. 1305–1358, 2015.

[30] F. F. Pollitz, "Transient rheology of the uppermost mantle beneath the Mojave Desert, California," *Earth Plan. Sc. L.*, vol. 215, pp. 89–104, Oct. 2003.

[31] Y.-P. Lin, L. Zhao, and S.-H. Hung, "Full-wave multiscale anisotropy tomography in Southern California," *Geophys. Res. Lett.*, vol. 41, no. 24, pp. 8809–8817, 2014.

[32] V. H. Lai, R. W. Graves, S. Wei, and D. Helmberger, "Evidence for strong lateral seismic velocity variation in the lower crust and upper mantle beneath the California margin," *Earth Planet. Sci. Lett.*, vol. 463, pp. 202–211, 2017.

[33] P. Bird, "Long-term fault slip rates, distributed deformation rates, and forecast of seismicity in the western United States from joint fitting of community geologic, geodetic, and stress direction data sets," *J. Geophys. Res.*, vol. 114, no. B11, 2009.

[34] E. H. Field, T. H. Jordan, M. T. Page, K. R. Milner, B. E. Shaw, T. E. Dawson, G. P. Biasi, T. Parsons, J. L. Hardebeck, A. J. Michael, *et al.*, "A synoptic view of the third Uniform California Earthquake Rupture Forecast (UCERF3)," *Seism. Res. Lett.*, vol. 88, no. 5, pp. 1259–1267, 2017.

[35] E. H. Hearn, "Kinematics of southern California crustal deformation: Insights from finite-element models," *Tectonophysics*, vol. 758, pp. 12–28, 2019.

[36] R. McCaffrey, "Block kinematics of the Pacific–North America plate boundary in the southwestern United States from inversion of GPS, seismological, and geologic data," *J. Geophys. Res.*, vol. 110, no. B7, 2005.

[37] B. J. Meade and B. H. Hager, "Block models of crustal motion in southern California constrained by GPS measurements," *J. Geophys. Res.*, vol. 110, no. B03403, 2005.

[38] J. C. Spinler, R. A. Bennett, M. L. Anderson, S. F. McGill, S. Hreinsdóttir, and A. McCallister, "Present-day strain accumulation and slip rates associated with southern San Andreas and eastern California shear zone faults," *J. Geophys. Res.*, vol. 115, no. B11, 2010.

[39] W. C. Hammond, G. Blewitt, and C. Kreemer, "Block modeling of crustal deformation of the northern Walker Lane and Basin and Range from GPS velocities," *J. Geophys. Res.*, vol. 116, no. B4, 2011.

[40] E. L. Evans, J. P. Loveless, and B. J. Meade, "Total variation regularization of geodetically and geologically constrained block models for the Western United States," *Geophys. J. Int.*, vol. 202, no. 2, pp. 713–727, 2015.

[41] E. L. Evans, "A comprehensive analysis of geodetic slip-rate estimates and uncertainties in California," *Bull. Seism. Soc. Am.*, vol. 108, no. 1, pp. 1–18, 2018.

[42] M. Oskin, L. Perg, E. Shelef, M. Strane, E. Gurney, B. Singer, and X. Zhang, "Elevated shear zone loading rate during an earthquake cluster in eastern California," *Geology*, vol. 36, no. 6, pp. 507–510, 2008.

[43] J. Chéry, M. D. Zoback, and R. Hassani, "An integrated mechanical model of the San Andreas fault in central and northern California," *J. Geophys. Res.*, vol. 106, no. B10, pp. 22051–22066, 2001.

[44] J. Chéry, "Geodetic strain across the San Andreas fault reflects elastic plate thickness variations (rather than fault slip rate)," *Earth Planet. Sci. Lett.*, vol. 269, no. 3-4, pp. 352–365, 2008.

[45] N. P. Fay and E. D. Humphreys, "Fault slip rates, effects of elastic heterogeneity on geodetic data, and the strength of the lower crust in the Salton Trough region, southern California," *J. Geophys. Res.*, vol. 110, no. B09401, 2005.

[46] R. Y. Chuang and K. M. Johnson, "Reconciling geologic and geodetic model fault slip-rate discrepancies in Southern California: Consideration of nonsteady mantle flow and lower crustal fault creep," *Geology*, vol. 39, no. 7, pp. 627–630, 2011.

[47] S. Barbot and Y. Fialko, "Fourier-domain Green's function for an elastic semi-infinite solid under gravity, with applications to earthquake and volcano deformation," *Geophys. J. Int.*, vol. 182, no. 2, pp. 568–582, 2010.

[48] S. Barbot and Y. Fialko, "A unified continuum representation of postseismic relaxation mechanisms: semi-analytic models of afterslip, poroelastic rebound and viscoelastic flow," *Geophys. J. Int.*, vol. 182, no. 3, pp. 1124–1140, 2010.

[49] S. Barbot, J. D. Moore, and V. Lambert, "Displacement and stress associated with distributed anelastic deformation in a half-space," *Bull. Seism. Soc. Am.*, vol. 107, no. 2, pp. 821–855, 2017.

[50] S. Barbot, "Deformation of a half-space from anelastic strain confined in a tetrahedral volume," *Bull. Seism. Soc. Am.*, vol. 108, no. 5A, p. 2687, 2018.

[51] Y. N. N. Lin, A. P. Kositsky, and J. P. Avouac, "PCAIM joint inversion of InSAR and ground-based geodetic time series: Application to monitoring magmatic inflation beneath the Long Valley Caldera," *Geophys. Res. Lett.*, vol. 37, 2010.

[52] L. L. Tsang, E. M. Hill, S. Barbot, Q. Qiu, L. Feng, I. Hermawan, P. Banerjee, and D. H. Natawidjaja, "Afterslip following the 2007 Mw 8.4 Bengkulu earthquake in Sumatra loaded the 2010 Mw 7.8 Mentawai tsunami earthquake rupture zone," *J. Geophys. Res.*, vol. 121, no. 12, pp. 9034–9049, 2016.

[53] Q. Qiu, J. D. P. Moore, S. Barbot, L. Feng, and E. Hill, "Transient Viscosity in the Sumatran Mantle Wedge from a Decade of Geodetic Observations," *Nature Communications*, vol. 9, no. 995, 2018.

[54] C.-H. Tang, Y.-J. Hsu, S. Barbot, J. D. Moore, and W.-L. Chang, "Lower-crustal rheology and thermal gradient in the Taiwan orogenic belt illuminated by the 1999 Chi-Chi earthquake," *Nature Communications*, vol. 5, no. 2, 2019.

[55] J. R. Weiss, Q. Qiu, S. Barbot, J. H. Foster, T. J. Wright, A. Saunders, B. A. Brooks, M. Bevis, E. Kendrick, R. S. Jr., S. R. Cimbaro, L. E. Lenzano, J. Barón, J. C. Báez, A. Echalar, and J. Avery, "Illuminating subduction zone rheological properties in the wake of a giant earthquake," *Sci. Adv.*, 2019.

[56] Y. Hang, S. Barbot, J. Dauwels, W. T., P. Nanjundiah, and Q. Qiu, "Outlier-insensitive bayesian inference for linear inverse problems (outibi) with applications to space geodetic data," *Geophys. J. Int.*, vol. 221, p. 334–350, 2020.

[57] N. Fay, R. A. Bennett, J. Spinler, and E. Humphreys, "Small-scale upper mantle convection and crustal dynamics in southern California," *Geochemistry, Geophysics, Geosystems*, vol. 9, no. 8, 2008.

[58] D. F. Argus, Y. Fu, and F. W. Landerer, "Seasonal variation in total water storage in California inferred from GPS observations of vertical land motion," *Geophys. Res. Lett.*, vol. 41, no. 6, pp. 1971–1980, 2014.

[59] S. Barbot, N. Lapusta, and J. P. Avouac, "Under the hood of the earthquake machine: Towards predictive modeling of the seismic cycle," *Science*, vol. 336, no. 6082, pp. 707–710, 2012.

[60] S. Barbot, P. Agram, and M. De Michele, "Change of Apparent Segmentation of the San Andreas Fault Around Parkfield from Space Geodetic Observations Across Multiple Periods," *J. Geophys. Res.*, vol. 118, no. 12, pp. 6311–6327, 2013.

[61] D. Bowman, G. King, and P. Tapponnier, "Slip partitioning by elastoplastic propagation of oblique slip at depth," *Science*, vol. 300, no. 5622, pp. 1121–1123, 2003.

[62] S. Daout, S. Barbot, G. Peltzer, M.-P. Doin, Z. Liu, and R. Jolivet, "Constraining the kinematics of metropolitan Los Angeles faults with a slip-partitioning model," *Geophys. Res. Lett.*, vol. 43, no. 21, 2016.

[63] C. Rollins, J.-P. Avouac, W. Landry, D. F. Argus, and S. Barbot, "Interseismic strain accumulation on faults beneath Los Angeles, California," *J. Geophys. Res.*, vol. 123, no. 8, pp. 7126–7150, 2018.

[64] D. F. Argus, M. B. Heflin, G. Peltzer, F. Crampé, and F. H. Webb, "Interseismic strain accumulation and anthropogenic motion in metropolitan Los Angeles," *J. Geophys. Res.*, vol. 110, no. B4, 2005.

[65] J. F. Dolan, L. J. McAuliffe, E. J. Rhodes, S. F. McGill, and R. Zinke, "Extreme multi-millennial slip rate variations on the Garlock fault, California: Strain super-cycles, potentially time-variable fault strength, and implications for system-level earthquake occurrence," *Earth Planet. Sci. Lett.*, vol. 446, pp. 123–136, 2016.

[66] Q. Qiu, S. Barbot, T. Wang, and S. Wei, "Slip complementarity and triggering between the fore-shock, mainshock, and afterslip of the 2019 Ridgecrest rupture sequence," *Bull. Seism. Soc. Am.*, 2020.

[67] W. Yang and E. Hauksson, "The tectonic crustal stress field and style of faulting along the Pacific North America Plate boundary in Southern California," *Geophys. J. Int.*, vol. 194, no. 1, pp. 100–117, 2013.

[68] S. Bourne, P. England, and B. Parsons, "The motion of crustal blocks driven by flow of the lower lithosphere and implications for slip rates of continental strike-slip faults," *Nature*, vol. 391, no. 6668, p. 655, 1998.

[69] M. Roy and L. H. Royden, "Crustal rheology and faulting at strike-slip plate boundaries: 1. an analytic model," *J. Geophys. Res.*, vol. 105, no. B3, pp. 5583–5597, 2000.

[70] M. D. Long and P. G. Silver, "Shear wave splitting and mantle anisotropy: Measurements, interpretations, and new directions," *Surveys in Geophysics*, vol. 30, no. 4-5, pp. 407–461, 2009.

[71] T. W. Becker, V. Schulte-Pelkum, D. K. Blackman, J. B. Kellogg, and R. J. O'Connell, "Mantle flow under the western United States from shear wave splitting," *Earth Planet. Sci. Lett.*, vol. 247, no. 3-4, pp. 235–251, 2006.

[72] G. W. Bawden, A. J. Michael, and L. H. Kellogg, "Birth of a fault: connecting the Kern County and Walker Pass, California, earthquakes," *Geology*, vol. 27, no. 7, pp. 601–604, 1999.

[73] T. H. Dixon and S. Xie, "A kinematic model for the evolution of the Eastern California Shear Zone and Garlock Fault, Mojave Desert, California," *Earth Planet. Sci. Lett.*, vol. 494, pp. 60–68, 2018.

[74] V. Lambert and S. Barbot, "Contribution of viscoelastic flow in earthquake cycles within the lithosphere-asthenosphere system," *Geophys. Res. Lett.*, vol. 43, no. 19, pp. 142–154, 2016.

[75] S. Barbot, "Asthenosphere flow modulated by megathrust earthquake cycles," *Geophys. Res. Lett.*, vol. 45, pp. 6018–6031, 2018.

[76] C. DeMets and T. H. Dixon, "New kinematic models for Pacific-North America motion from 3 Ma to present, I: Evidence for steady motion and biases in the NUVEL-1A model," *Geophys. Res. Lett.*, vol. 26, no. 13, pp. 1921–1924, 1999.

[77] G. F. Sella, T. H. Dixon, and A. Mao, "REVEL: A model for recent plate velocities from space geodesy," *J. Geophys. Res.*, vol. 107, no. B4, pp. ETG–11, 2002.

[78] C. DeMets, R. G. Gordon, and D. F. Argus, "Geologically current plate motions," *Geophys. J. Int.*, vol. 181, no. 1, pp. 1–80, 2010.

[79] Z. Altamimi, L. Métivier, and X. Collilieux, "ITRF2008 plate motion model," *J. Geophys. Res.*, vol. 117, no. B7, 2012.
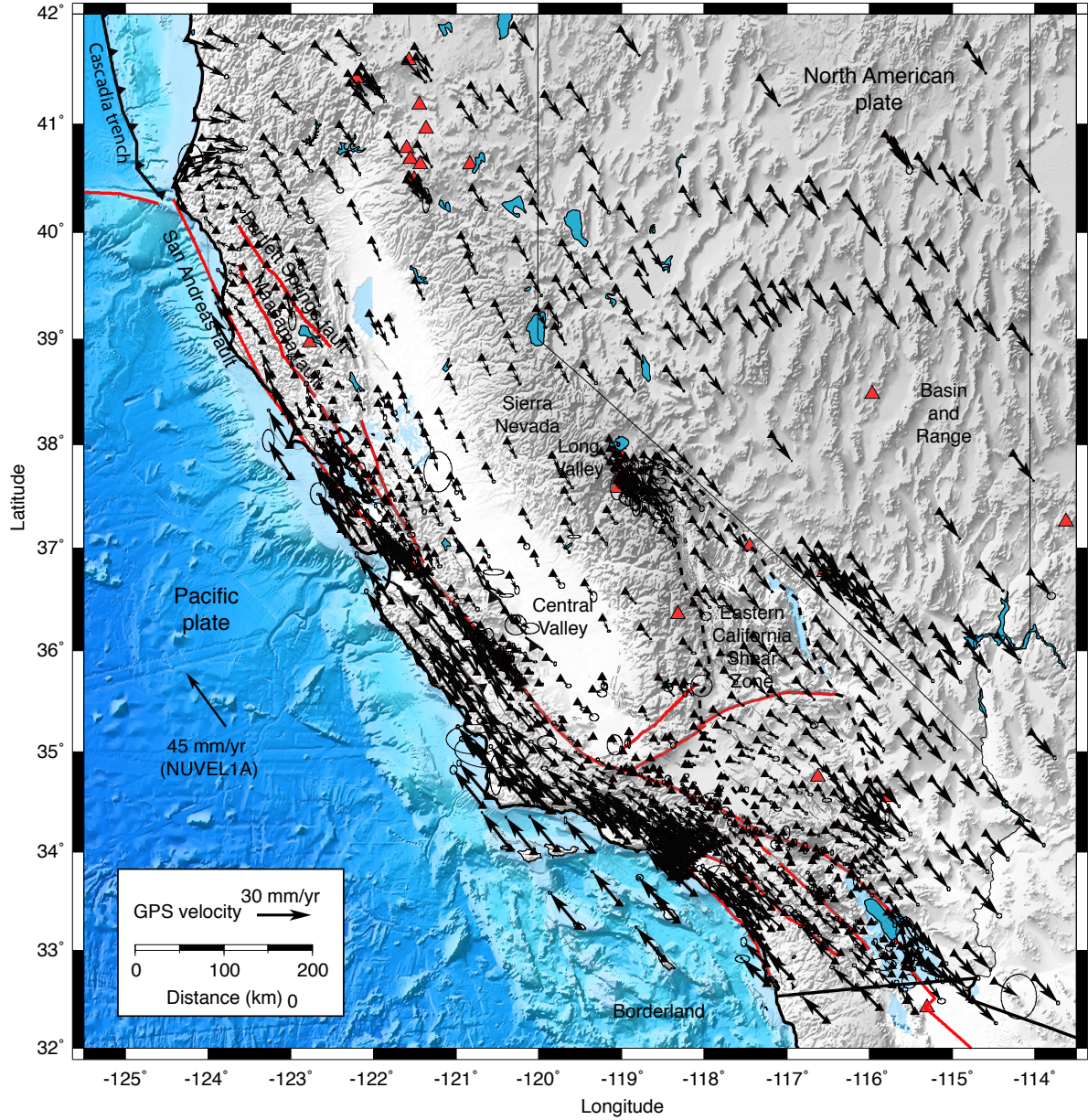
Figure 1: Distribution of the GPS velocity around the San Andreas fault. The GPS velocity field is available at `ftp://data-out.unavco.org/pub/products/velocity/`. The red triangles correspond to quaternary volcanos. The trace of major faults is from the USGS catalog of quaternary faults `https://www.usgs.gov/natural-hazards/earthquake-hazards/faults`. The NUVEL1A velocity is for the Pacific plate relative to the North American plate [76]. The dashed contour indicates the approximate extent of the Eastern California Shear Zone.
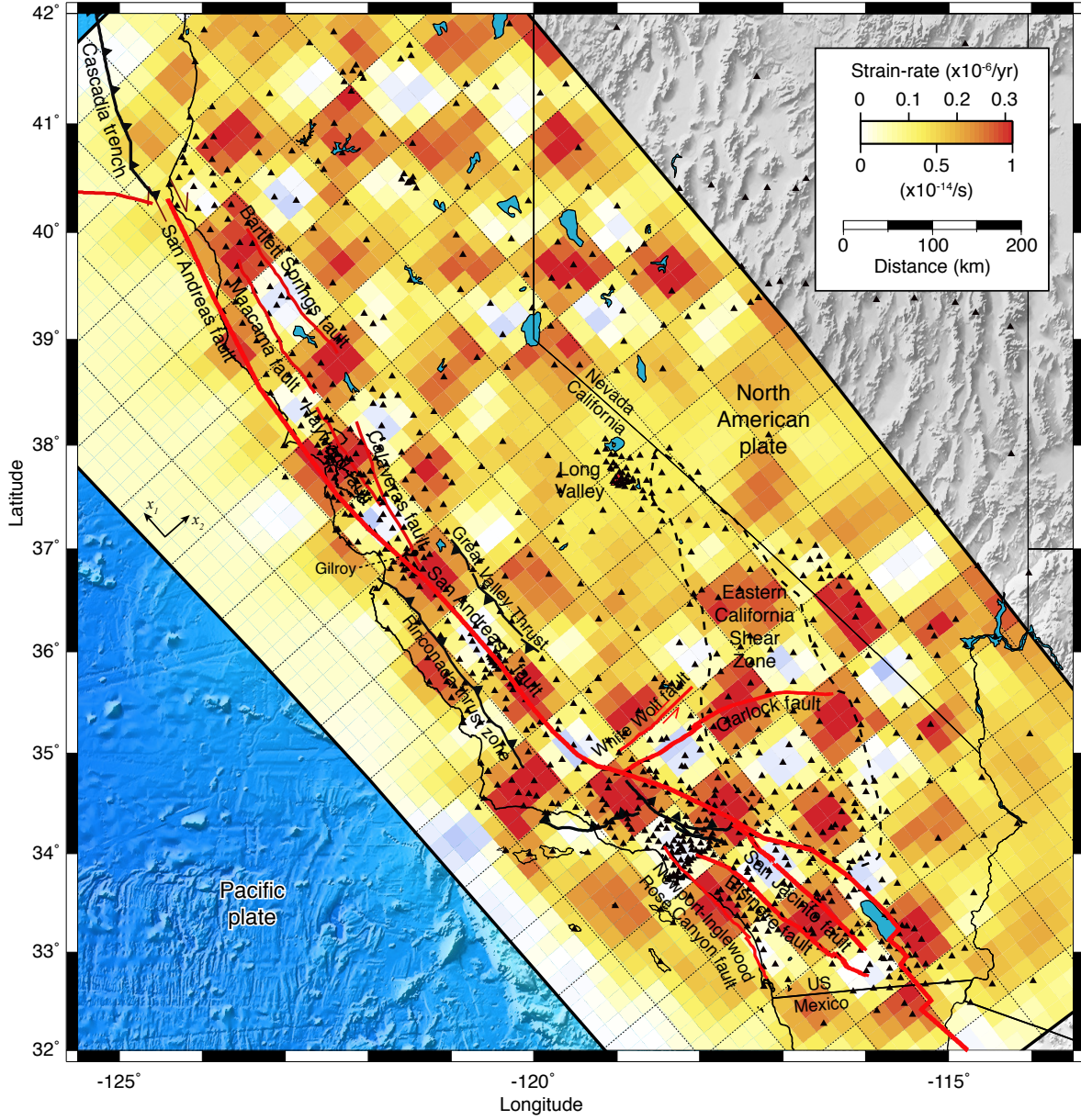
Figure 2: Checkerboard test for inversion of plastic strain-rate. The checkerboard pattern of strain-rate consisting of clusters of four by four neighboring volume elements with non-zero shear strain-rate is indicated by the dashed lines. The recovered pattern after inversion of the synthetic data (black triangles) is in the colored background. Loss of resolution occurs offshore and around the Long Valley Caldera due partly to insufficient station coverage.

16
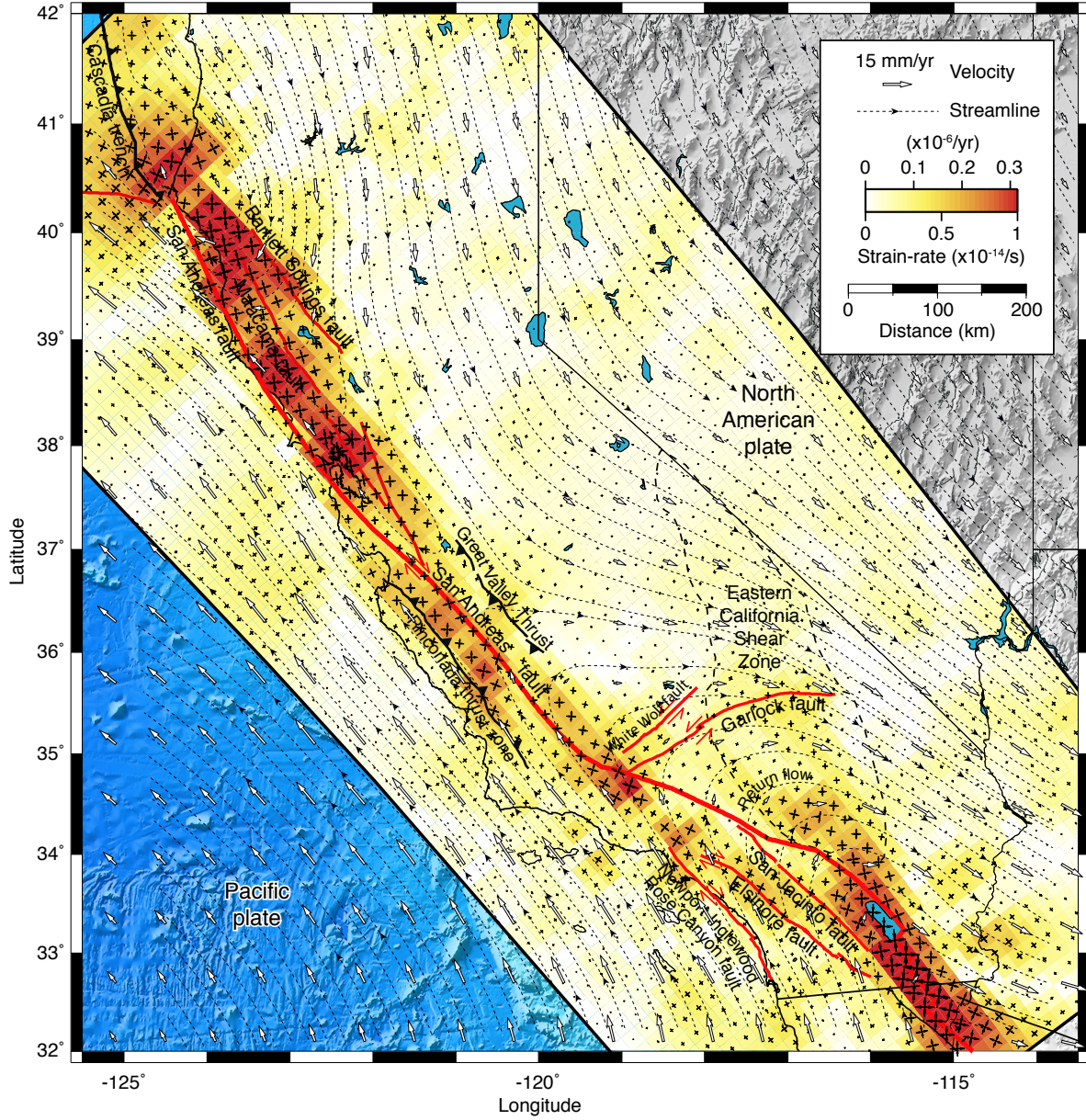
Figure 3: Contribution of viscoelastic flow on plate boundary deformation. The background color indicates the spatial distribution of strain accumulation (the norm of the strain-rate tensor) along the plate boundary. The pairs of black arrows indicate the principal axes of the strain-rate tensor. The white arrows indicate the surface velocity predicted by the model. The dashed black lines with arrows are the flow streamlines. The thick red lines indicate the surface trace of major strike-slip faults. The thick dashed contour indicates the approximative extent of the Eastern California Shear Zone.
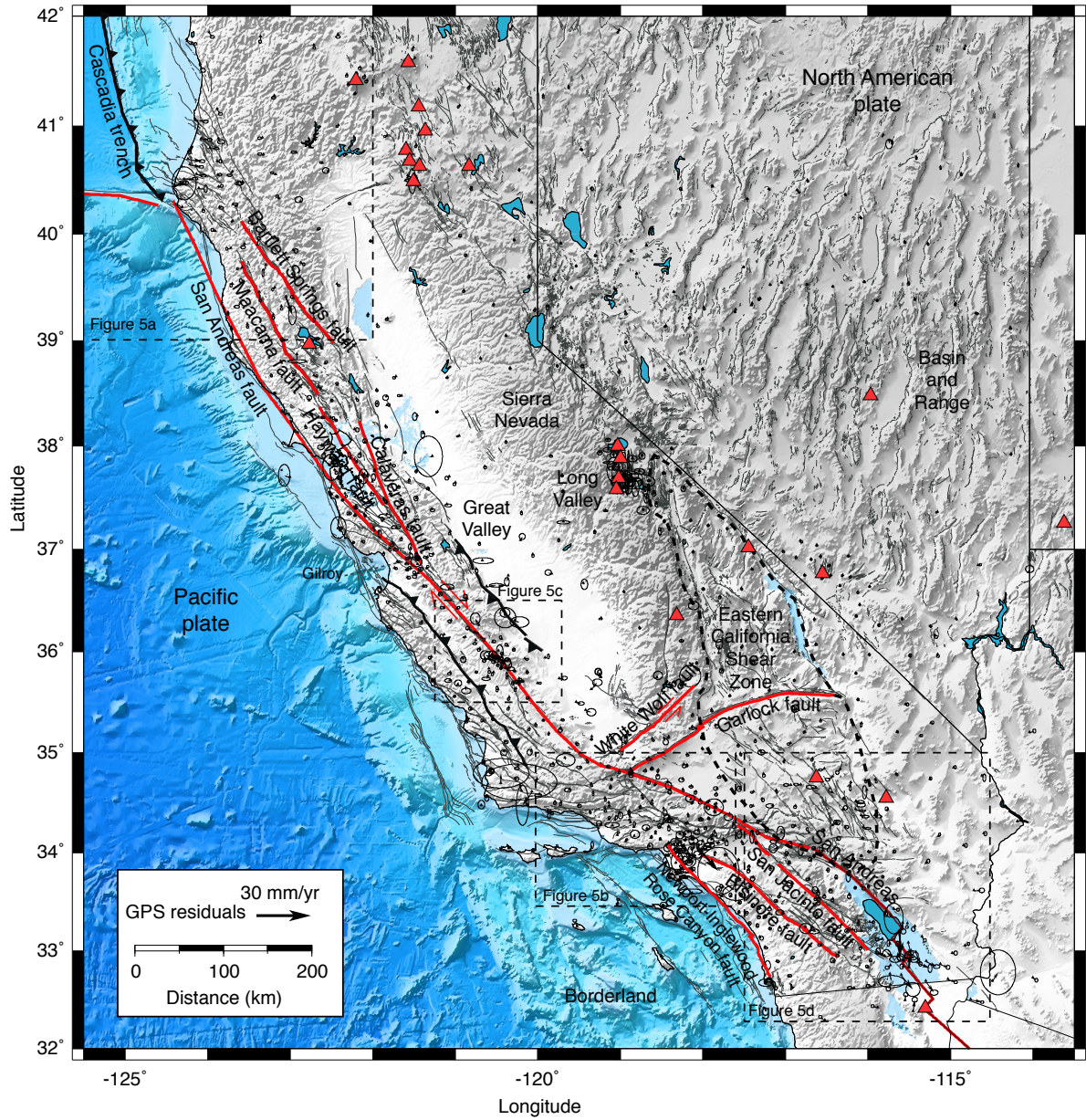
Figure 4: GPS residuals velocity. Residuals are plotted at the same scale as in Figure 2. The largest spatially coherent residuals occur off the Cascadia trench, around the Parkfield, in the Transverse Range, and around the recent 1999 Hector Mine and 2010 El Mayor-Cucapah earthquakes, shown in more detail in Figure 5. The red triangles correspond to quaternary volcanos.
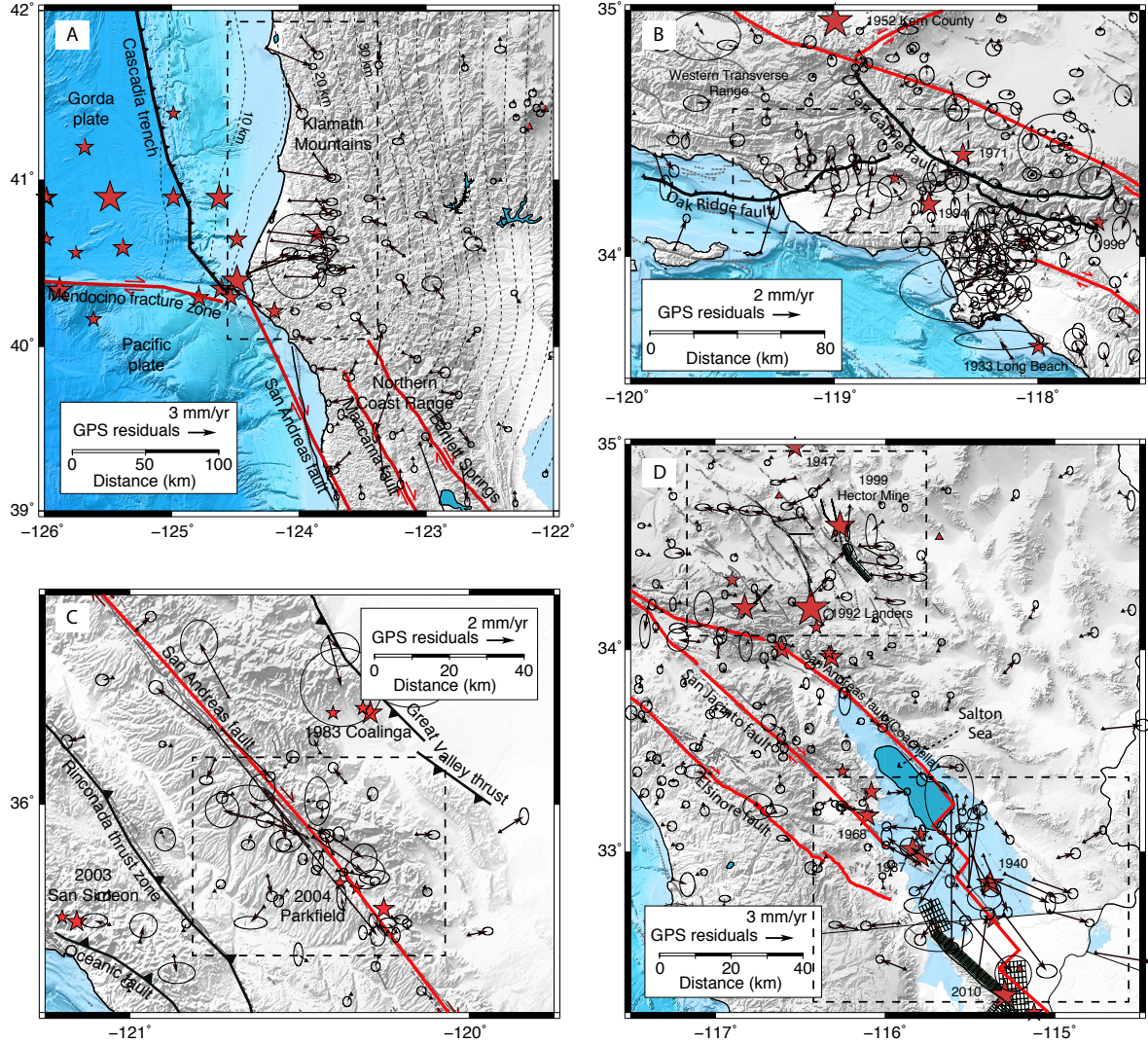
Figure 5: Spatially coherent residuals. A) Residual velocity in the northern California coast. Off the Gorda plate, the residuals are due to a shallow locked section of the Cascadia trench. The dashed line at the 5 km contours of the Cascadia megathrust (https://earthquake.usgs.gov/data/slab/models.php). Diverging velocity in the northern Coast Ranges may be due to variation of water storage in aquifers. B) Residual velocity around the Transverse Range due to shortening occurs across the Oak Ridge fault and the San Gabriel fault and in the Los Angeles Basin due to oil and gas extraction. The black lines with chevrons indicate major thrust faults. C) Velocity residuals near the Parkfield segment of the San Andreas fault due to complex shallow fault geometry, shallow creep, and afterslip from the 2004 Mw=6.0 earthquake. D) GPS residual velocity around the 1999 Mw=7.1 Hector Mine earthquake and the 2010 Mw=7.2 El Mayor-Cucapah earthquake in the Salton trough. The red stars indicate the hypocenter of historical earthquakes of magnitude greater than 6 since 1906.
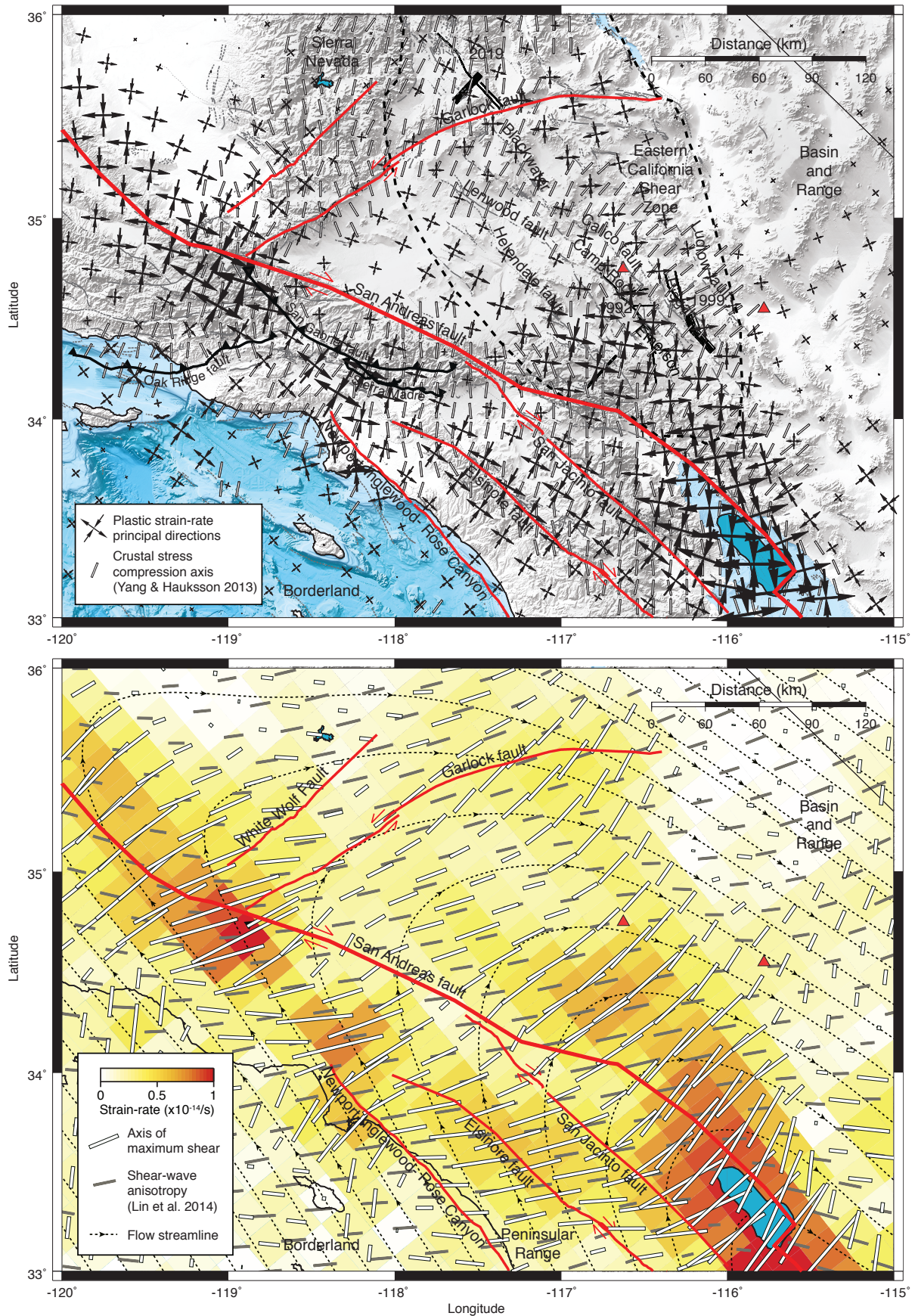
Figure 6: Principal strains, faults, seismicity, and seismic anisotropy. The top panel shows the principal strain direction of plastic strain-rate in relation to the orientation of strike-slip and thrust faults, and in relation with the compression axis of seismicity [67]. The bottom panel direction of maximum shear of the plastic strain-rate and flow streamlines in relation to the orientation of seismic anisotropy [31]. Red lines indicate strike-slip faults. Black lines with chevrons indicate major thrust faults. Red triangles indicate quaternary volcanos. Grey lines indicate quaternary faults. The dashed contour indicates the approximative extent of the Eastern California Shear Zone.
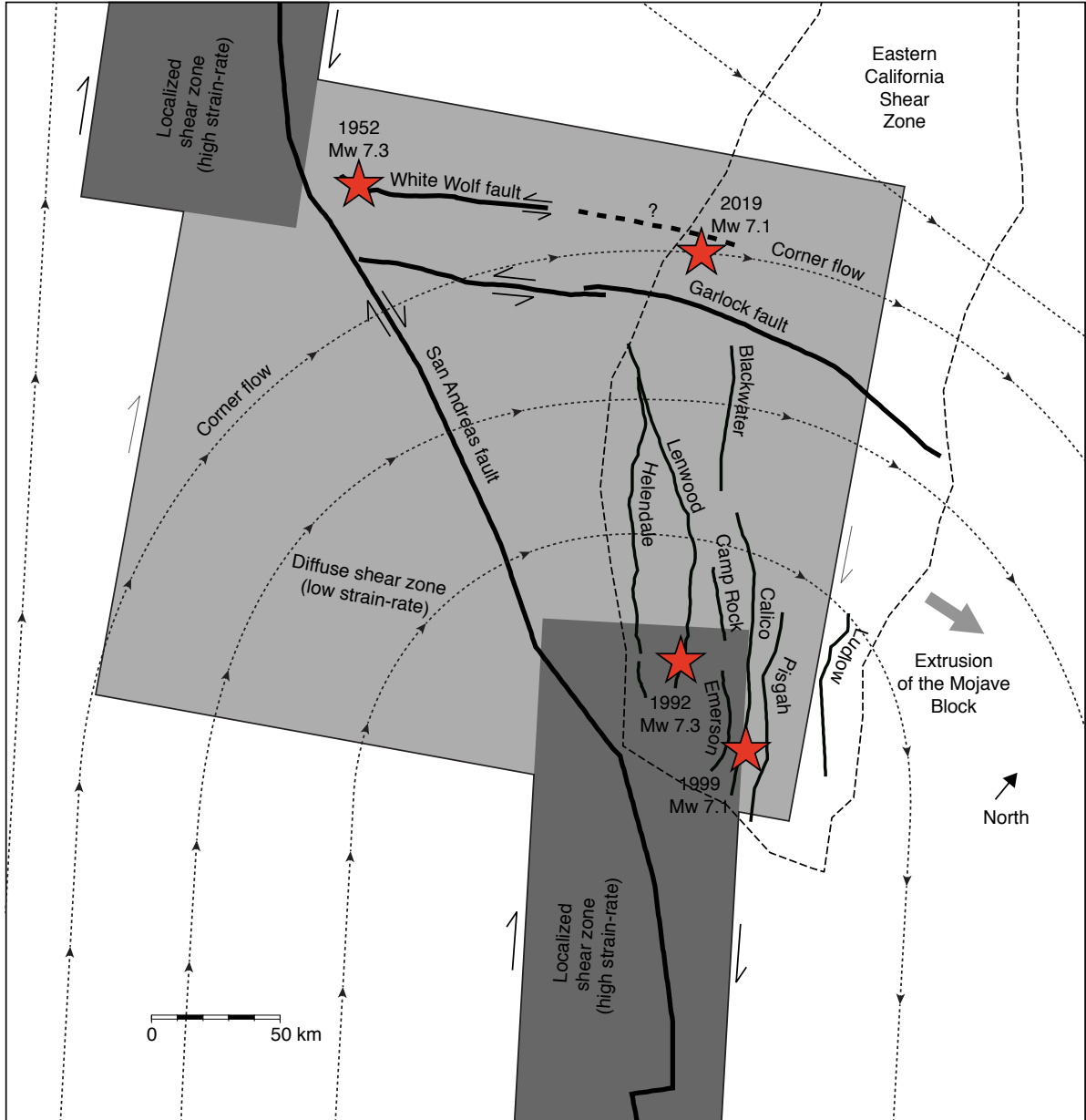
Figure 7: Evolution of fault activity due to extrusion of the Mojave block. Deformation in the Mojave is associated with corner flow distributed over a wide shear zone that accommodates the restraining step-over between more localized shear zones below the Carrizo and Coachella sections of the San Andreas fault. Faults that escaped out of the shear zone, like the Ludlow fault, become inactive. As the extrusion of the Mojave block continues, the White Wolf fault may materialize the triple junction at the big bend of the San Andreas fault. The stars represent large earthquakes in the last century.
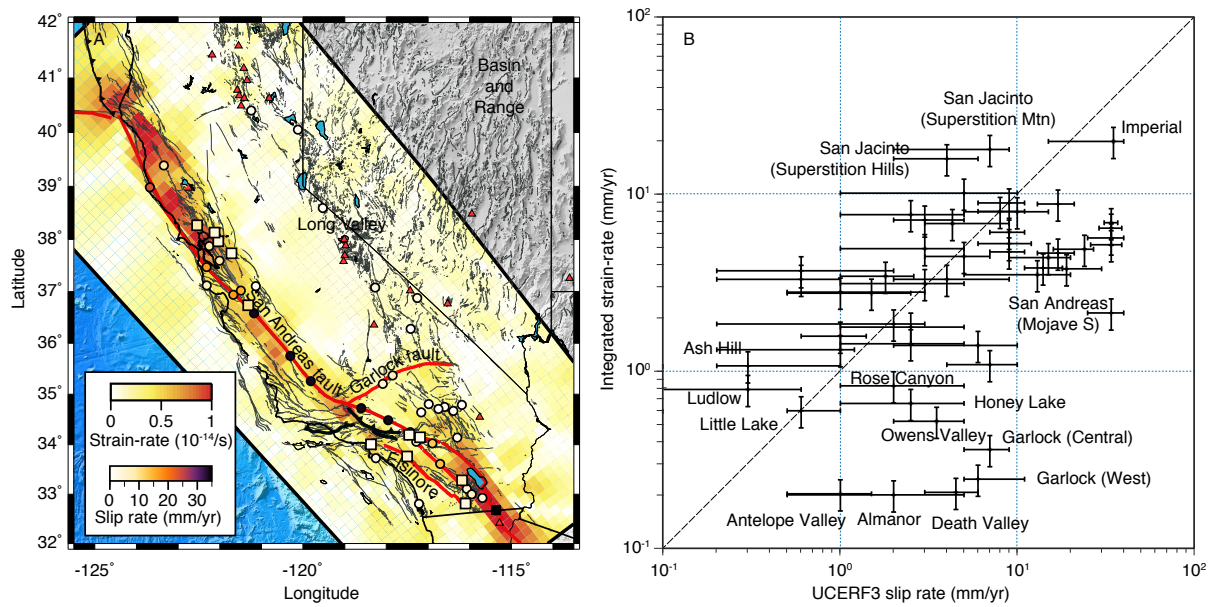
Figure 8: Anelastic strain accumulation and long-term slip-rate on California faults. A) Spatial distribution of distributed strain and the location of paleo slip-rate on strike-slip fault segments catalogued in Appendix B of the Uniform California Earthquake Rupture Forecast (UCERF3) (`https://pubs.usgs.gov/of/2013/1165/`). The squares indicate locations where the deep strain-rate immediately below the fault segment matches the long-term fault slip rate. Red triangles indicate quaternary volcanos. B) Comparison of the UCERF3 slip rates with the integrated strain-rate below the corresponding fault segment.
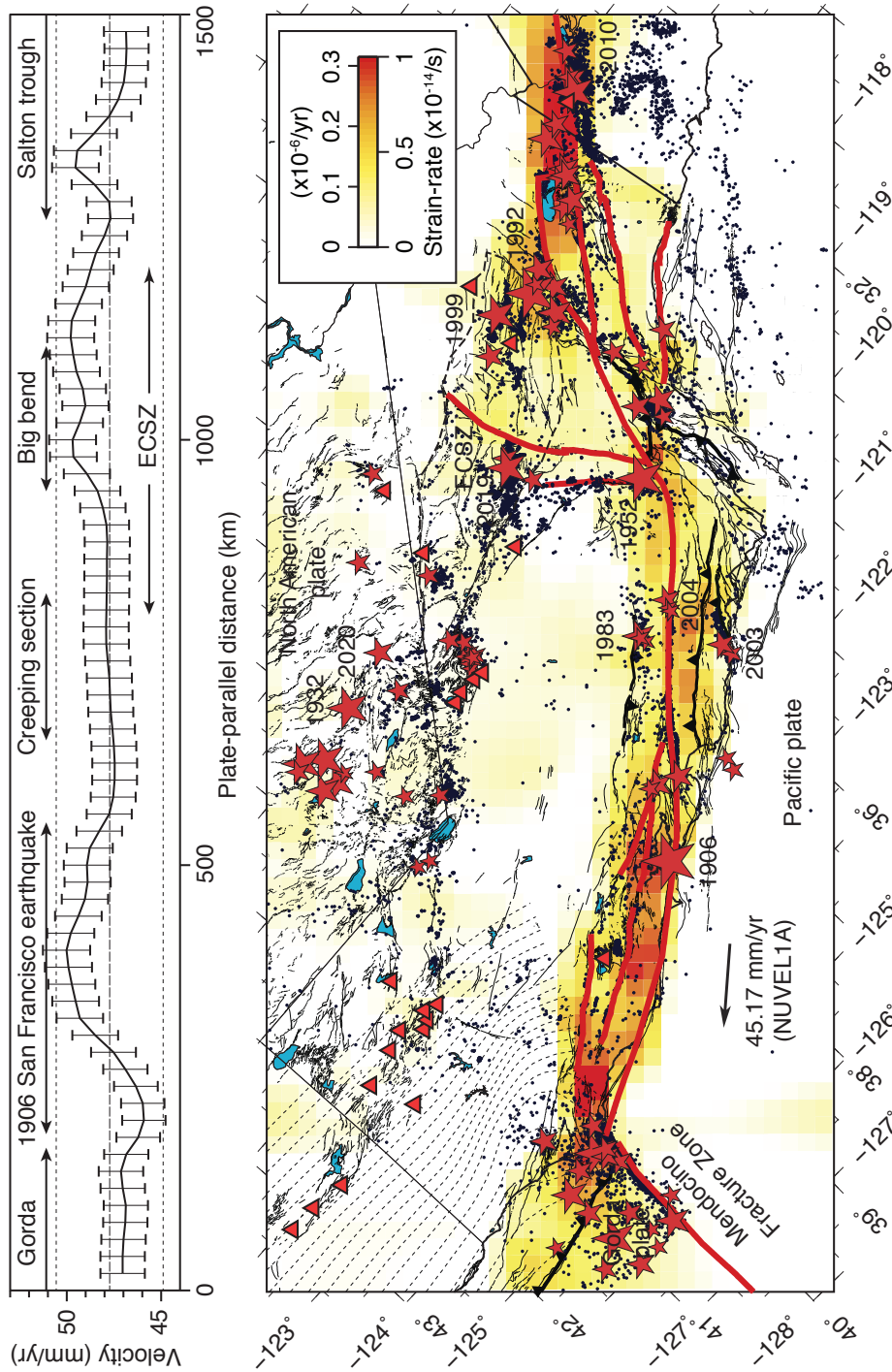
Figure 9: Strain partitioning across the San Andreas fault system. The top panel shows the plate parallel velocity from the integration of the shear component of the strain-rate tensor in the plate-perpendicular direction. The dashed lines indicate the average velocity and (1-σ) standard deviation of the Pacific plate relative to the North American plate from 18 different models, including REVEL-2000, NUVEL 1A, ITRF 2008, and MORVEL 2010 [76, 77, 78, 79]. The bottom panel shows the distribution of deep shear strain-rate in California from the Mendocino triple junction to the Salton trough, highlighting the strain partitioning primarily between the San Andreas fault and the Eastern California Shear Zone. The red stars and black circles show the hypocenter of historical earthquakes and earthquakes from moment magnitude 3 and above from 1990 to 2020 around California from the Northern California, respectively. The seismic catalogue is obtained from the Earthquake Data Center (https://ncedc.org) and the Southern California Earthquake Data Center (https://scedc.caltech.edu). The gray dashed profiles correspond to the 5 km depth contours of the Cascadia megathrust. Red triangles indicate quaternary volcanos.
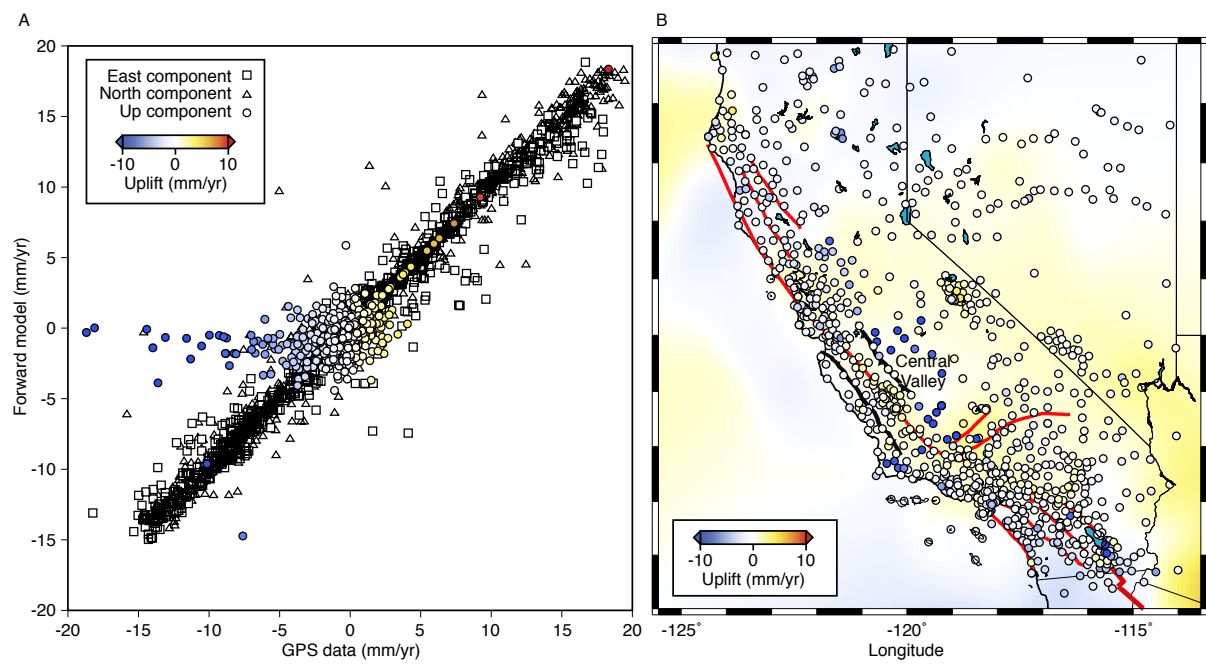
Supplementary Figure 1: Correlation between GPS data and model predictions. A) Observations and model predictions for the east (squares), north (triangles), and vertical (colored circles) components of the velocity field. B) Map view of vertical deformation for the data (colored circles) and forward model (background). Large subsidence in the Central Valley produce a few outliers.
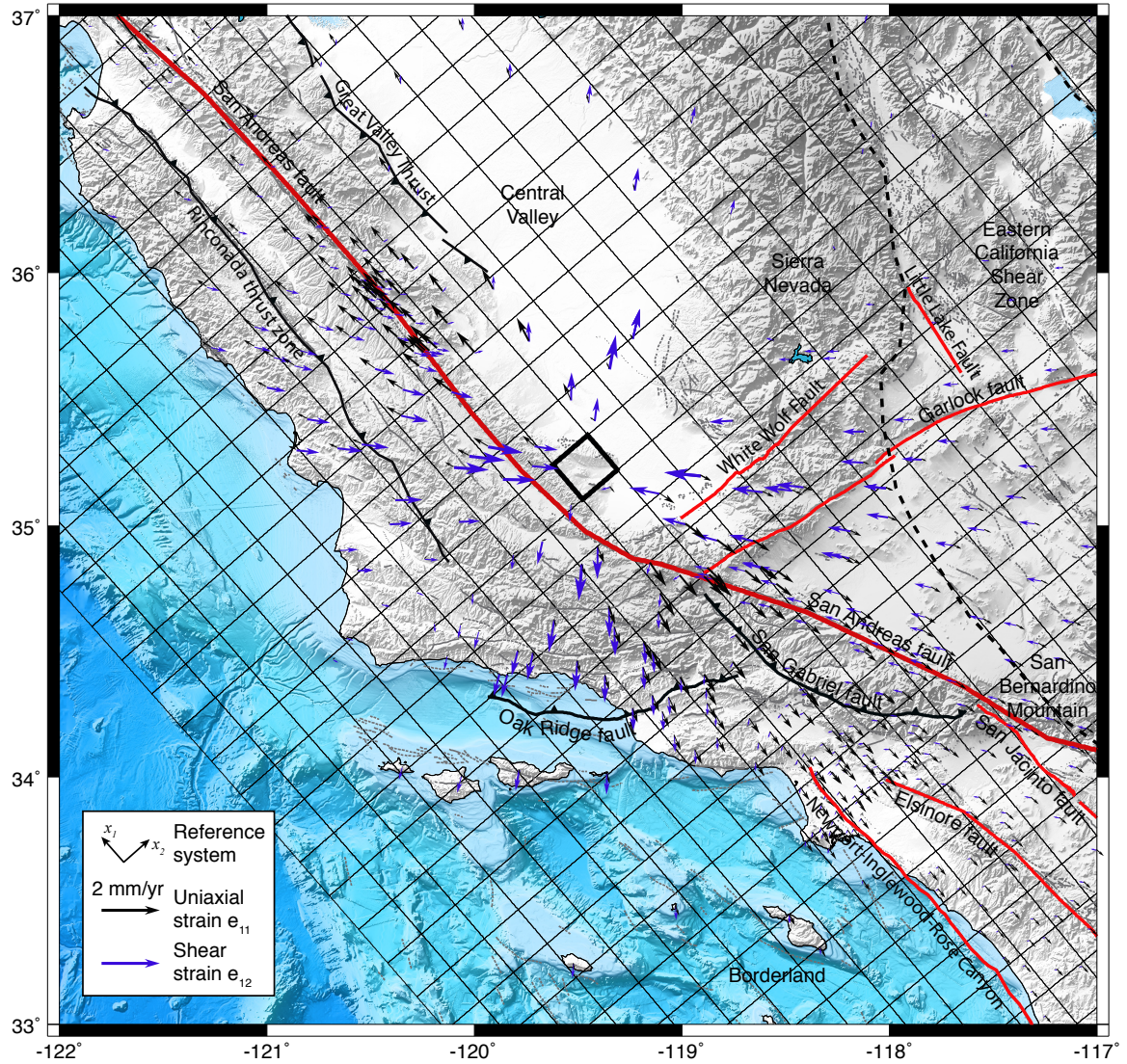
Supplementary Figure 2: Example Green's function. Surface deformation associated with distributed anelastic strain in a semi-infinite $20\times20\,\text{km}$ cuboid (black square) at $20\,\text{km}$ depth. The black vectors correspond to the surface deformation for uniaxial horizontal strain along the approximately plate-parallel direction. The blue vectors are for the surface deformation due to a non-zero ϵ_{12} shear strain in the volume element. The underlying mesh of volume elements used to model distributed viscoelastic flow in and around California is shown in the background.