Plate bending earthquakes and the strength distribution of the lithosphere

Dan Sandiford[1*] and Tim Craig[2]

[1] School of Earth, Atmosphere and Environment, Monash University, Melbourne, Victoria, Australia (formerly)

[2] COMET, Institute of Geophysics and Tectonics, School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK.

[*] e-mail: dansandifordscience@gmail.com

This is an author-produced PDF of an article submitted in *Geophysical Journal International* for peer review.

1 Summary

This study investigates the dynamics and constitutive behaviour of the oceanic lithosphere as it bends and yields during subduction. Two main observational constraints are considered: the maximum bending moment that can be supported by the lithosphere, and the inferred neutral plane depth in bending. We particularly focus on regions of old lithosphere where the 'apparent' neutral plane depth is about 30 km. We use subduction modelling approaches to investigate these flexural characteristics. We reassess bending moment estimates from a range of previous studies, and show a significant convergence towards what we call the 'intermediate' range of lithosphere strength: weaker than some classical models predict, but stronger than recent inferences at seamounts. We consider the non-uniqueness that arises due to the trade-offs in strength as well background (tectonic) stress state. We outline this problem with several end-member models, which differ in regard to relative strength in the brittle and ductile regimes. We evaluate the consistency of these models in terms of a range of constraints, primarily the seismic expression of the outer rise. We show that a 30 km neutral plane depth in old oceanic lithosphere implies net slab pull $<$ 1-2 TN/m. In contrast, models with low brittle strength imply that regions with a 30 km neutral plane depth are under moderate net axial compression. Under these conditions, reverse faulting is predicted beneath the neutral plane at depths $>$ 30 km. We show that moderate variations in background stress have a large impact on the predicted anelastic dissipation. We

suggest brittle reverse faulting is a marginal phenomenon which may be inhibited by moderate changes in background stress.

Key words: subduction; lithosphere strength; earthquakes; plate flexure

2 Introduction

The depth distribution of strength in the oceanic lithosphere is thought to arise mainly from the interaction between the normal-stress sensitivity of faults in the brittle domain and the temperature dependence of intra-crystalline creep. An important aspect of this strength distribution is the relatively low flexural strength relative to axial strength, because the strain developed in bending has an inverse character to the strength profile. The influential view that subduction bending is a major resisting force in mantle convection dynamics (e.g., Conrad & Hager, 1999) has therefore receded somewhat (Capitanio et al., 2009; Leng & Zhong, 2010; Buffett & Becker, 2012). Nevertheless, there is still significant debate about the role and interplay of various deformation mechanisms, the maximum stress levels in bending, and the interaction of bending related stress with other sources of stress — usually anticipated to be the background component associated with the driving/resisting forces of plate motion.

We start by clarifying some assumptions and, in particular, what we mean by strength. We assume that the response of the lithosphere in subduction-related loading can be adequately described by simple bending of a plate in plane strain, and that the stress state that develops is dominated by this bending, although it may be modified by background stresses. This set of assumptions seems justified because, while the oceanic lithosphere away from subduction trenches exhibits diverse focal mechanisms (normal, reverse and strike slip), the $\sim$ 100 km region near trenches is overwhelmingly dominated by normal faulting in the upper 15-35 km, with sub-horizontal tension axes oriented perpendicular to the trench, consistent with a state of effective tension induced by bending (Stein & Pelayo, 1991; Craig et al., 2014a). Under typical values of subduction-hinge curvature (e.g., 5×10^{-6} m^{-2}), elastic moduli, and laboratory-constrained deformation mechanisms, the lithosphere is expected to undergo comprehensive yielding (Goetze & Evans, 1979). In this situation the bending moment saturates, such that it does not increase with additional strain (bending). By lithospheric strength, we hence mean the saturation moment of a given rheological model.

In this study we focus particularly on lithosphere of 80 - 120 Myr age, as this region is well represented in the current global distribution of subducting plate by age, and is also where the inferred neutral plane depth can be estimated with most confidence. To aid discussion, we characterise the saturation moment of different rheological models as being 'weak', 'intermediate' and 'strong'. The proposed strength of the oceanic lithosphere in previous studies spans this range (Chapple & Forsyth, 1979; McNutt & Menard, 1982; Levitt & Sandwell, 1995; Garcia et al., 2019; Bellas & Zhong, 2021). These ranges of saturation moment are shown in Fig. 1, and are based on a simple 'family' of strength models discussed in Section 7. The regions are defined as follows: for 100 Myr-old lithosphere, weak models have saturation moments ($<$1); intermediate models (1-2) and strong models($>$3). These ranges are measured in units of

N-m/m (N) $\times 10^{17}$. In the following we use (N) to describe the bending moment per unit-length of trench.

For lithosphere in the 80 - 120 Myr range, outer rise seismicity exhibits an apparent transition from normal- to reverse-faulting mechanisms at a depth of $\sim$ 30 km (e.g. Fig. 3). It is usually (although not always) assumed that this observation reflects a stress transition across the neutral plane associated with bending (Chapple & Forsyth, 1979). The thermal thickness of plates in the 80-120 Myr range is 100 - 130 km (Parsons & Sclater, 1977; Richards et al., 2020), so that the inferred neutral plane lies at close to a quarter the plate depth. This in turn suggests that plate strength lies mainly in the upper half of the plate.

Classical yield strength envelopes (YSEs) based on experimental constitutive laws generally predict satisfactory neutral plane depths (30-40 km for $\sim$ 100 Myr lithosphere), under the assumptions that the background stress is insignificant, the friction coefficient is relatively high (i.e. comparable to Byerlee's Law), and when the exponential stress dependence of olivine (referred to here as LTP for low-temperature plasticity) is taken into account (e.g., Byerlee, 1978; Goetze & Evans, 1979; McNutt & Menard, 1982; Hunter & Watts, 2016). We note that while this classical model predicts satisfactory neutral plane depths, it typically predicts that LTP will function as the dominant deformation mechanism beneath the neutral plane, with the brittle (frictional) strength in compression several times higher than the steady-state flow stress (Goetze & Evans, 1979). This prediction conflicts with observations of deeper seismicity below the neutral plane indicating trench-perpendicular compression.

However, many of these standard assumptions might be questioned. For instance, if net axial force (see 3.4) due to slab pull is significant (i.e., much greater than typical ridge push magnitudes) and/or if outer rise normal faults have low friction coefficient (< 0.3; e.g., Craig et al., 2014b) the $\sim$ 30 km apparent neutral plane depth is difficult to reconcile, unless differential stress levels in the $\sim$ 30-60 km region are also significantly reduced relative to typical LTP models (e.g., Goetze & Evans, 1979; Mei et al., 2010). The net effect is a much weaker strength model, and would be broadly consistent with conclusions that have emerged from recent studies of seamount flexure (Zhong & Watts, 2013; Bellas et al., 2022).

The torque that acts on the plate due to the flexural topography, referred to as the 'observed moment' by previous studies, provides a rheology-independent constraint on the bending moment (Goetze & Evans, 1979; McNutt & Menard, 1982). The mathematical background is provided in Supplementary Section S1. Together, we refer to bending moment and neutral plane estimates as the 'flexural characteristics'. There is a complimentary aspect to these constraints as the neutral plane is sensitive to the relative (depth) distribution of stress in the lithosphere, while the observed moment has sensitivity to magnitude of those stresses (albeit via the first moment integral). However, the constraints are not

sufficient to uniquely determine the strength distribution with depth, and the non-uniqueness is a key focus of this study.

After preliminary sections that cover the background (3), data (4), an overview of modelling approaches (5) and modeling insights (6), the study consists of two main pieces of analysis. In Section 7 we consider a range of existing data-sets that constrain the maximum bending moment in subducting lithosphere. We argue for a reinterpretation of the observed moment data of Levitt & Sandwell (1995), in terms of a weaker strength model than originally proposed. A key conclusion is that when interpreted through a 'common lens' of the maximum bending moment, a range of studies have converged on a similar 'intermediate' strength for old oceanic lithosphere (Chapple & Forsyth, 1979; Levitt & Sandwell, 1995; Hunter & Watts, 2016; Garcia et al., 2019). In Section 8 we consider the less well-constrained part of the problem, which is the relative distribution of the strength with depth. We characterise a parameter space consisting of the relative strength of the brittle and ductile parts of the lithosphere, along with the background stress, and explore the implications of different parts of the parameter space.

3 Background and previous work

3.1 Flexural strength of the oceanic lithosphere

There is an extensive body of literature that has used the deflection of the plate surface due to loading to infer the mechanical properties of the oceanic lithosphere. The primary environs are subduction zone outer rises and seamounts. The intention here is not to attempt a summary, but to recall for the reader that: 1) studies of subduction zones have inferred a large range of strength models (from what we call weak, through to strong; see Fig. 1); and 2) that studies of seamount flexure have increasingly made the case for weak or intermediate models.

Flexural studies of subduction zones offer different perspectives on acceptable models of lithosphere strength, as underscored by the work of Chapple & Forsyth (1979); McNutt & Menard (1982); Levitt & Sandwell (1995); Hunter & Watts (2016); Garcia et al. (2019). In each case these studies address the inverse problem of generating optimal models of plate strength in terms of either: 1) direct fitting of forward models with observed topography or gravity profiles (Chapple & Forsyth, 1979; Hunter & Watts, 2016; Garcia et al., 2019), or 2) comparison between the 'observed' moment associated with flexural topography, and the predicted saturation moment of different strength models (Goetze & Evans, 1979; McNutt & Menard, 1982; Levitt & Sandwell, 1995). We note that the observed moment is typically

calculated around the first zero crossing of the flexural topography, as this eliminates the effect of the net axial force term in the moment balance (see Supplementary Section S1).

Chapple & Forsyth (1979) solve the thin-plate equations using an iterative numerical approach that incorporates the moment variation with curvature for different *ad-hoc* depth-strength profiles. The profiles are calibrated so that they approximately match the inferred neutral plane constraint, based on seismological data available at the time. Their optimal model, which is developed with reference to old lithosphere (100 – 140 Myrs) is shown in Fig. 1. It produces a saturation moment of 1.8×10^{17} N, and lies within what we have called the intermediate strength range.

Garcia et al. (2019) solve the thin-plate equations for 3D subduction segments with variable loading. Like Chapple & Forsyth (1979), the constitutive model includes anelastic deformation (yielding). In this case, they incorporate a moment-curvature relationship underpinned by a classical strength model (olivine creep and brittle failure, see Section 5.2). They vary the friction coefficient to produce a strong ($\mu = 0.6$) and a weak ($\mu = 0.3$) model (both assuming a hydrostatic pore pressure). They find that the 'weak' model can better fit data across Pacific basin subduction zones. This model produces a saturation moment of 1.3×10^{17} N for 100 Myr lithosphere, and lies at the lower end of our intermediate strength range. We note the 'weak' and 'strong' models described in Garcia et al. (2019) both lie within what we call the intermediate range (suggesting that these authors had already narrowed the focus of their investigation to that strength region).

Hunter & Watts (2016) optimise elastic deflection models for circum-Pacific subduction zones, based on both constant-thickness and variable thickness plate models. To compare these results to laboratory constrains, YSEs are computed for various strength models, then converted to equivalent elastic plate thicknesses, and compared with the optimised values from the inversion. They argue that their inversions are consistent with several combinations of friction coefficient and low-temperature plasticity, including the parameters of Mei et al. (2010), with friction coeffcients of either 0.6 or 0.3 (in both cases assuming hydrostatic pore pressure). These combinations can be shown to describe strength models that lie in the intermediate range (e.g. Fig. 1).

An alternative way to interpret the results of Hunter & Watts (2016), is to consider the bending moment predicted by the elastic flexure models at the first zero crossing. This subjects their results to an equivalent analysis as is undertaken by Goetze & Evans (1979); McNutt & Menard (1982); Levitt & Sandwell (1995), in which an elastic plane solution is used to fit the flexural topography, and then the moment at the first zero crossing is estimated. Data from Fig. 10 of Hunter & Watts (2016) was re-digitised, and the moment values at the first zero crossing are shown with red crosses in Fig. 1. These estimates reflect inversion for averaged flexural deflections ('trench-bin') typically aggregated across

several thousand kilometers. In contrast, the blue points in Fig. 1 represent estimates from individual trench profiles (from Fig. Levitt & Sandwell (1995)).

The studies of McNutt & Menard (1982); Levitt & Sandwell (1995) use similar approaches: estimating the observed moment based on uniform elastic plate fitting. However, they arrived at quite different conclusions: weak and strong models respectively (following our definition of strength). There are two reasons for this. Firstly, the estimates derived in McNutt & Menard (1982) mainly lie within the weak strength region (green points in Fig 1), while the larger dataset of Levitt & Sandwell (1995) contains moment values across weak, intermediate and strong regions (red points in Fig 1). We discuss potential reasons for this discrepancy in Section 7.3. The second issue relates to the interpretation of scatter in the data. Levitt & Sandwell (1995) argue that an acceptable lithospheric strength model should provide a ceiling to the observed moment data, which implies models in the strong region; their proposed model is shown with a dashed blue line in Fig. 1. In contrast, McNutt & Menard (1982) sought a strength model that would provide a 'best fit' to the data. This issue of interpreting scatter in the observed moment data is the focus of Section 7.

In summary, previous studies addressing lithospheric strength in subduction zones have advocated models throughout what we define as the weak, intermediate and strong regions (as shown in Fig. 1). However, we must be attentive to the assumptions, ambiguity, and interpretative aspects of some of these conclusions. For instance, the basis on which Levitt & Sandwell (1995) advocate a strong lithosphere model is tied to the interpretation of the scatter in the data.

Meanwhile, a number of studies have concluded that observations of lithosphere flexure around Hawaii and seamounts more generally, require that the lithosphere is substantially weaker than inferred at subduction zones Zhong & Watts (2013); Pleus et al. (2020); Bellas & Zhong (2021); Bellas et al. (2022). We do not consider seamount flexure directly in this study, however our analysis has implications for the the purported 'conundrum' of strong plates at subduction zones and weak plates around seamounts.

3.2 The neutral plane as a constraint on strength

For lithosphere in the 80-120 Myr range, outer rise seismicity exhibits an apparent normal-reverse mechanism transition depth of $\sim$ 30 km (e.g. Fig. 3, see Section 4). Note that all references to depths refer to depth beneath the surface of the plate (seafloor). It is usually assumed that this observation reflects a transition across a neutral plane in the bending of the plate as subduction proceeds. A contrasting view is discussed the following section.

Chapple & Forsyth (1979) noted that "the distribution, mechanisms, and depths of the earthquakes associated with the bending of the lithosphere provide the information needed to reduce greatly the non-uniqueness inherent in modeling topography alone". In Chapple & Forsyth (1979), this process involved testing *ad-hoc* strength-depth models to match the neutral plane constraint as well as flexural-topographic profiles from old subducting lithosphere. They noted that the apparent neutral plane depth for lithosphere in the 100 Myr range is $\sim$ 30 km, an observation that we discuss in the next section. The strength model derived by Chapple & Forsyth (1979) is shown with a dashed cyan-colored line in Fig. 1, and lies in the range we describe as intermediate strength.

In exploring different ways to weaken classical strength models, McNutt & Menard (1982) argued against frictional-strength reduction because the resulting model produced a neutral plane depth of $\sim$ 40 km for 100 Myr lithosphere. Their preferred strength model instead achieves strength reduction by appealing to significantly reduced creep strength (LTP) through modification of the activation energy in an existing flow law. Interestingly, this produces the opposite problem: a neutral plane that is shallower (20-25 km for 100 Myr lithosphere) than the apparent depth inferred from seismicity.

In reality, the background stress contribution also effects the depth of the neutral plane. If the typical stress state of lithosphere prior to subduction (and its variability) was well constrained, the neutral plane would potentially *greatly* reduce non-uniqueness, as Chapple & Forsyth (1979) propose. Unfortunately, because it is not, the problem remains under-constrained. We address this issue in Section 8.

3.3 The outer rise normal-reverse transition

Outer rise normal-faulting earthquakes, where sub-horizontal tension axes are oriented perpendicular to the trench, have long been associated with plate bending, (e.g., Stauder, 1968). It should be noted that the seismicity is not concentrated at the bathymetric rise/fore-bulge, but rather occurs throughout the domain of plate curvature increase, which may extend beneath the forearc (Sandiford et al., 2020). In this study, the term 'outer rise' is used to refer to the domain of curvature increase in subducting plates, and the associated dynamics.

Extensional earthquakes in the outer rise region are generally located at systematically shallower depth than less frequent reverse events (e.g., see Fig. 3). These reverse-mechanism earthquakes have usually been attributed to compressional bending stresses beneath a neutral plane (Chapple & Forsyth, 1979; Craig et al., 2014a). The distribution of outer rise reverse events is non-uniform, with few or no occurrences along some subduction zones (such as eastern

Sunda and the Aleutians). Most activity has been along the Kurile, Japan, Philippine, Solomon Islands, Tonga, Kermadec, and Chile zones (Craig et al., 2014a). Even in those regions there is notable clustering, with a concentration of events in central-to-southern Honshu, central-to-northern Kuriles, northern Ryuku, southern Philippines, and central Chile. Tonga-Kermadec represent the only region where outer rise compressional events are common across several thousand kilometers (the reader is referred to maps in Craig et al. (2014a); Ye et al. (2021).

The study that has most directly challenged the idea of a reverse faulting as a consequence of plate bending induced shortening/compression is Mueller et al. (1996). They argued that the purported association is compromised by both laboratory and seismologic observations. Of course, this perspective would invalidate attempts to infer a neutral plane on the basis of the normal-reverse seismicity transition. The argument can be summarised in the following points:

1) Brittle failure is an essential condition for seismicity. Reverse-faulting earthquakes, occurring at a depth of 30-50 km, are inconsistent with Coulomb-type frictional failure unless the effective friction coefficients drastically reduced ($\mu < 0.1$). But the flexural topography of the outer rise requires that friction coefficients cannot be this low.

2) There is insufficient evidence for material instability, i.e. fault development with unstable stress drop, in the semi-brittle or plastic regimes, as was suggested by Chapple & Forsyth (1979).

3) Attributing both thrust- and normal-faulting outer rise events to the same mechanism cannot directly account for the important observation that the former occur much less frequently than the latter.

Instead, Mueller et al. (1996) propose that reverse-faulting earthquake mechanisms in subduction outer rises are associated with regions of anomalous compression, which develop in response to strongly-localised subduction-resisting forces, such as interplate asperities or buoyant subducted crust. They argue that many such earthquakes may be shallower than previously reported, thus obviating the incompatibility with Coulomb-type frictional strength considerations.

We do not have the scope to address each of these points, but will instead focus on the seismological data (with the benefit of more than 20 years' additional accumulation). In Section 4, we review the data from a range of sources. In our assessment, these now provide a much more compelling case that in many regions the outer rise exhibits a transition from normal faulting to reverse faulting, in close proximity.

3.4 Background sources of stress in the bending lithosphere

Although flexure can clearly be modified by the background stress state, in the modelling of subduction zone flexure, there seems to be little consensus as to the overall pattern of such contributions. Some early studies include a net background component essentially as a free parameter, and advocated large compressive forces (Hanks, 1971). Parsons & Molnar (1976) showed that applied moments, which had been neglected in some previous studies, could produce equivalent deflection without requiring large axial compression. In many subsequent studies, the background stress is ignored, or plays a subordinate role (Turcotte et al., 1978; Chapple & Forsyth, 1979; Hunter & Watts, 2016; Garcia et al., 2019).

Different models for the sources of background stress, imply differences in terms of reference state, stress magnitude, and spatial and temporal variability. A useful dichotomy is articulated in Mueller et al. (1996), who state that while "'driving forces are broadly distributed and ... resisting forces [may be] relatively localized." Mueller et al. (1996) proposed that where outer rise seismicity exhibits reverse faulting mechanisms, the stress state is dominated by localised resisting forces, such as asperities in the interplate zone.

On the other hand, if the most important driving force for oceanic plates is net slab pull (of up to perhaps 10 TN/m; Conrad & Lithgow-Bertelloni, 2002; Faccenna et al., 2012), a state of effective tension seems the logical *a priori* choice for the vast majority of the global subduction system; it would seem strange not to see the 'signal' of such strong background stress in some aspect of the flexural behavior, for instance in the depth to the neutral plane.

However, the relative partitioning of plate driving forces remains debated (Becker & O'Connell, 2001; Conrad & Lithgow-Bertelloni, 2002; Sandiford et al., 2005; Copley et al., 2010; Ghosh & Holt, 2012; Husson, 2012; Stotz et al., 2018). If the mantle has an active driving role beneath the plates (Stotz et al., 2018; Liu et al., 2021), or if gravitational potential energy changes across the trench-outer rise are as large as some studies suggest (Bessat et al., 2020), a compressional rather than (effective) tensional background stress state may prevail. The numerical model we discuss in Section 6 provides some important insights into this question.

Flexure of the lithosphere is commonly modelled as a cylindrical bending problem (i.e. 2D plain strain). Yet the background deviatoric stress state is clearly 3 dimensional; under Andersonian assumptions the order of two remaining principal stresses is variable, as is the angular relationship to the trench or the plane of bending. In 2D modelling of subduction flexure, the background stress is usually incorporated in terms of a prescribed net force, with either a tensional or compressional sense. In this study we refer to this simplification (of the 3D stress) as the *net axial*

force (units of TN/m). Throughout the paper we refer to the net axial force magnitude as being 'low' ($\leq$ 3 TN/m), 'moderate' ($\sim$ 5 TN/m) and 'high' ($\geq$ 10 TN/m).

4 Data

4.1 Outer rise seismicity

Fig. 3b shows well-constrained centroid depths of outer rise earthquakes, after Craig et al. (2014a) and sources therein. While there has historically been some controversy about the veracity of the systematic depth offset (e.g., Mueller et al., 1996), this pattern is clearly evident for lithosphere $>$ 80 Myr, as shown in Fig. 3b. These data also reaffirm the conclusion of Chapple & Forsyth (1979), that the apparent normal-reverse transition depth for 80-120 Myr lithosphere is at about 30 km. The distribution becomes far more ambiguous, however, for lithosphere younger than 80 Myr, although we note that the low proportion of subducting lithosphere in the 60-80 Myr range (see Fig. 3a) is responsible for producing the substantial gap in data in this age range.

The conclusions drawn from the global distribution of moderate-magnitude outer rise earthquakes (e.g. (e.g., Craig et al., 2014a, ; Fig. 3) are generally consistent with more localised insights gained from finite-fault modelling of individual events, doublet occurrences and microseismcity (e.g., Todd & Lay, 2013b; Ye et al., 2021; Obana et al., 2012). Selected examples are summarised here.

Finite-fault modelling of the M_w 7.5 2020 Kurile earthquake (Ye et al., 2021), confirms that rupture was limited to depths beneath the anticipated neutral plane depth estimated from hypocenters for lithosphere of that age (e.g. Fig. 3). The reported aftershocks of the 2020 Kurile event are concentrated at depths of 27 to 45 km, consistent with the expectation that these events lie between the neutral plane and an isotherm close to 600 °C (Emmerson & McKenzie, 2007).

A number of instances of normal-reverse doublets provide further evidence that the stress state in the outer rise transitions from effective tension to compression with depth. These include the 2012, Honshu Japan doublet, the 2011 Kermadec doublet, the 2006-2009 Kurile doublet, and the 2012 Phillipine trench events (see Ye et al. (2021) Fig. 7 for summary). Aftershocks recorded with OBS after the 2012 Honshu doublet, showed normal-faulting seismicity down to about 30 km, relative to the plate surface (see (Obana et al., 2014), Fig. 2).

Following the 2011 Tohoku-Oki earthquake, a permanent seafloor seismic network was installed offshore Honshu (S-net). While S-net is not used to routinely determine earthquake hypocenters, a recent study us S-net data to relocate small events in the outer rise (Zhao et al. (2022) Fig. 11). The relocated hypocenters suggest clustering into shallower and deeper groups, with a separation between the two occurring about 30 km relative to the plate surface. The deeper cluster corresponds to events with reverse focal mechanisms (personal communication with author). This is broadly consistent with the 30-35 km neutral plane depth inferred based on OBS data obtained following the 2011 Tohoku-Oki earthquake (Kubota et al., 2019). S-net data has the capacity to significantly improve our future understanding of outer-rise seismic expression, including its temporal and spatial variation.

In summary, we think that a normal-reverse transition with systematic depth offset, is a consistent phenomenon in global outer rise seismicity in older plates, albeit unobserved in some regions. We think that the simplest explanation for this seismic expression relates to stress/strain conditions across the neutral plane of bending.

4.1.1 The apparent depth of the normal-reverse transition

The data and studies summarised in the previous section suggest that the apparent normal-reverse transition depth for 80-120 Myr lithosphere is around 30 km, although there is some regional variability. A key question is to what extent this apparent transition is an intrinsic property of the strength distribution of the lithosphere, or alternatively exhibits a bias due to regions where the background stress is anomalous and effects the neutral plane depth. End-member perspectives can be found in Chapple & Forsyth (1979) and Mueller et al. (1996). In the former, the apparent neutral plane depth is assumed to be the intrinsic depth for old lithosphere (i.e. the depth produced in regions of isotropic background stress). In the latter study, reverse seismicity only occurs where the background stress is anomalously compressional, and represents a localised deviation from the background trend.

It is logical to think that if there are significant global variations in the background stress in the subducting lithosphere, more reverse seismicity will occur in the regions with the highest compression (or at least the smallest extensional component). This is because, relative to a 'average' background stress state, a more compressional state will result in a shallower neutral plane, increasing the volume that exists between the neutral plane and the brittle-ductile transition (i.e. the volume of rock where brittle compressional rupture can occur). As such, it is seems plausible that the apparent normal-reverse seismicity transition may biased by regions where the background stress is more compressional (which may simply mean less tensional). In our view, however, the argument that outer rise reverse seismicity is restricted to regions of anomalous resistance, where flexure is overwhelmed by net compression (Mueller et al., 1996), is inconsistent

with a range of subsequent detailed observations (Todd & Lay, 2013b; Obana et al., 2014; Zhao et al., 2022).

4.2 Observed moment data

The observed moment is the torque around a given point, due to the load of the flexural (uncompensated) topography seaward of that point. The derivation of this relationship is outlined in Supplementary Section S1. and is independent of the details of the particular rheological model under consideration.

Fig. 1 shows moment data from several previous studies. In the case of Levitt & Sandwell (1995); McNutt & Menard (1982) these are the observed moment 'sensu stricto', and reflect the value at the first zero crossing. In the case of Garcia et al. (2019), they represent the tabulated bending moment at maximum curvature from the flexural inversion (which utilises a yielding rheology). Typically these values were within 5 - 10 % of the saturation moment for the given rheology and assumptions. In the case of Hunter & Watts (2016), the moment data were redigitised from the bending moment profiles in the original study, estimated at the first zero crossing. These data reflect the variable elastic plate inversion (A2) from Hunter & Watts (2016).

Our understanding of rock deformation mechanisms suggests that bending moments should be close to saturated at curvature values encountered in the trench slope Goetze & Evans (1979). Ideally, moment data would reflect the saturation moment of the lithosphere, in which case different datasets could be compared without additional correction for curvature (which often has significant associated uncertainty). As shown in the Supplementary Fig. S7, for the majority ($\sim$ 70 %) of the profiles in the Levitt & Sandwell (1995) data, the curvature estimates imply that the plates have reached $>$ 75 % moment saturation by the first zero crossing (see Section S5). This is also consistent with the results of the numerical model discussed in Section 6. Hence, most of the data in Levitt & Sandwell (1995) should exhibit less than 25 % moment deficit; the fact that the scatter is significantly larger than these considerations suggest, is a key issue in the current study.

5 Methodology - summary

5.1 Subduction models - overview

Two modelling approaches are used to make quantitative predictions related to the bending of the lithosphere during subduction. These are referred to as (1) yield stress envelopes (YSEs), and (2) numerical subduction modelling. The results from the numerical subduction modelling inform certain aspects of YSE construction (e.g. strain rate patterns), as well as the interpretation of moment data (e.g. moment deficit). YSEs are used to produce 'forward models' of flexural behaviour, in order to compare with observations. In both modelling approaches, the constitutive behavior follows a 'classical strength model', which combines brittle faulting with both low and high temperature plasticity (as described in Section 5.2). Fig. 2 shows an example of the YSE approach. The implementation details are described in detail in Supplementary Section S4.

The numerical subduction model was developed with the ASPECT code (Bangerth et al., 2020). The model simulates the evolution of an idealised 2D subduction system, or the one-sided sinking of a thermal boundary layer due to thermal buoyancy. The implementation details are described in detail in the Supplementary Section S2. All parameters for the numerical model are provided in Table S1. Fig. 5 shows stress and deformation patterns in the numerical subduction model, described in Section 6.

In recent studies, the applicability of laboratory-derived flow lows, particularly those describing low-temperature plasticity, has been an important issue (Hunter & Watts, 2016; Bellas et al., 2022). This question is strongly dependent on the assumptions made in the model, in particular the temperature and strain rate. YSEs are often constructed under the assumption of a constant strain rate (Hunter & Watts, 2016; Garcia et al., 2019, e.g.,). However, we assume that strain rates vary with distance from the neutral plane, a choice that is consistent with deformation patterns in the numerical model (and described in Supplementary Section S5). This assumption means that for given flow law parameters, we compute higher creep stress than those of Hunter & Watts (2016); Garcia et al. (2019). Supplementary Fig. S6 highlights these differences.

The YSEs are based on a plate cooling model, with parameters derived from Parsons & Sclater (1977) (and identical to Hunter & Watts (2016); Garcia et al. (2019)). For 100 Myr lithosphere, the inferred depth limit for well-constrained earthquakes is $\sim$ 50 km depth, and corresponds to a temperature of $\sim$ 700 °C. The Parsons & Sclater (1977) model remains relatively consistent with more recent approaches, incorporating the joint inversion of plate subsidence and heat

flow, and based on temperature and pressure dependent thermal properties (Richards et al., 2018). The temperature profile is warmer than assumed in Bellas et al. (2022), due to the higher diffusivity in that study (1.0, rather than 0.8 ($\times 10^{-6} \mathrm{m}^2\,\mathrm{s}^{-1}$)). This higher diffusivity is equivalent to a 25 % increase in apparent age (i.e. $\sqrt{\kappa t}$ = constant). For typical strength models discussed in this paper, this would produces a $\sim$ 25 % increase in saturation moment. A comparison of temperature models is shown in Supplementary Fig. S8.

Parameter name	Value	Symbol	Units
Young's modulus	1×10^{11}	E	Pa
Acceleration due to gravity	9.8	g	m s−2
Poisson's Ratio	0.25	ν	-
Thermal plate thickness	125	-	km
Temperature at seafloor	0	-	°C
Mantle potential temperature	1350	T_m	°C
Thermal diffusivity	0.8×10^{-6}	κ	$\mathrm{m}^2\,\mathrm{s}^{-1}$
Mantle density	3300	ρ_m	$\mathrm{kg\,m}^{-3}$
Curvature	2×10^{-6}	K	m^{-1}
Curvature gradient	2×10^{-11}	$\frac{\partial K}{\partial x}$	m^{-2}
Plate velocity	10	u_x	$\mathrm{cm\,y}^{-1}$

Table 1: Standard parameters used in YSE modelling. For creep parameters see Table 2.

5.2 Constitutive behavior

5.2.1 The classical strength model

The classical strength model of the oceanic lithosphere combines elastic deformation, brittle deformation (frictional sliding on optimally-oriented faults) and separate expressions for low temperature (aka LTP or Peierls) and high temperature (aka power law or dislocation) creep (e.g., Goetze & Evans, 1979). Despite the success of this model, remaining sources of uncertainty include: 1) the significant variability in published experimentally derived equations for LTP; 2) significant variations in inferred friction coefficient based on different geophysical studies; 3) strength properties in the semi-brittle regime.

Semi-brittle behaviour is thought to operate throughout a significant depth, within the strongest part of the lithosphere (Ohnaka, 2013). A defining characteristic of semi-brittle behavior is that strength has both strong pressure and temperature sensitivity. It is often assumed that semi-brittle deformation will lead to a zone in which typical differential stresses are smaller that the peak stress given by the intersection of the brittle and ductile stresses (Ohnaka, 1992, 2013; Molnar, 2020). The paucity of constraints on stress-strain-rate-temperature relationships in this domain means

that the semi-brittle regime is routinely neglected in YSE modelling, which is also the case here. However, we note that the typical depth range of reverse faulting (typically 30-50 km for $\sim$ 100 Myr lithosphere) overlaps the anticipated semi-brittle domain (Molnar, 2020).

Moreover, the brittle strength of rocks is not restricted to the frictional strength of pre-existing faults. Other important processes include the saturation of frictional strength at high confining pressure, the temperature dependence of the fracture strength, changes in micro-crack behavior, and thermal shear instability (non-Coulombic faulting) (Ohnaka, 2013; Stesky et al., 1974; Shimada et al., 1983; Molnar, 2020; Renshaw & Schulson, 2004). Such processes may even lead to negative brittle strength-depth gradients (Ohnaka, 2013). It is plausible that outer rise reverse faulting at $\sim$ 25-50 km, may involve some of these processes.

5.2.2 Frictional sliding

The Coulomb strength criterion states that frictional strength on sliding surfaces is linearly related to the effective normal stress :

$$\tau = \mu\sigma_n \tag{1}$$

where μ is the friction coefficient. For faults of optimal orientation, the strength can be expressed in terms of the limiting value of the differential stress $\sigma_{xx} = (\sigma_1 - \sigma_3)$ as a function of depth (y):

$$\Delta\sigma_{xx} = \frac{\pm 2\mu(\rho g y(1-\lambda))}{(1+\mu^2)^{1/2} \pm \mu} \tag{2}$$

where ρ is density, g is gravity, λ is the pore pressure factor. The sign convention used here is that positive differential stress is extensional; the upper signs ($\pm$) in Eq. 2 apply to normal faults and the lower signs for reverse faults.

Predictions of Coulomb theory, incorporating standard laboratory-derived frictional coefficients (e.g. 0.6-0.8), and hydrostatic pore-pressure, are generally consistent with measured differential stress in the shallow upper crust ($<$ 1 km) (Townend & Zoback, 2000). Certain rock types and mineralogies (e.g. shales, clays and serpentines) can exhibit much lower friction coefficients (e.g. 0.1-0.3) (Ewy et al., 2003).

5.2.3 Intra-crystalline creep

At high temperatures ($T > 0.7\,T_{\text{melt}}$) and low stresses (< 100-200 MPa), olivine creep exhibits a power law relationship between strain rate and differential stress, known as dislocation creep. These conditions are relevant to the lower lithosphere and asthenoshere. The form of the power law creep is:

$$\dot{\varepsilon} = A\Delta\sigma^n \exp\left(-\frac{E+PV}{RT}\right), \tag{3}$$

where $\dot{\varepsilon}$, $\Delta\sigma$, T, P , R, are the strain rate, differential stress, temperature (K), pressure (Pa), and gas constant; A is a pre-exponential factor, E is the activation energy, V is the activation volume.

At lower temperatures ($T < 0.6\,T_{\text{melt}}$) and higher stresses, relevant for the shallow mantle lithosphere, a stronger dependence of the strain rate on stress is observed, which is usually described by an exponential law, known as low-temperature plasticity (LTP), or Peierls creeep.

The form of the LTP law generally used is:

$$\dot{\epsilon} = A\Delta\sigma^n \exp\left(-\frac{E}{RT}\left(1-\left(\frac{\sigma}{\sigma_P}\right)^p\right)^q\right), \tag{4}$$

where several additional parameters appear: σ_P is the Peierls stress constant, and p, q are additional parameters that depend on the geometry of the impediments to dislocations (Mei et al., 2010).

parameter symbol (units)	$A(\text{s}^{-1}\,\text{Pa}^{-n})$	$E(\text{J}\,\text{mol}^{-1})$	$V(\text{m}^3\,\text{mol}^{-1})$	n	σ_p(Pa)	p	q
Set 1							
Mei et al. (2010) (LTP)	1.4×10^{-19}	320×10^3	-	2	5.9×10^9	0.5	1
Hirth & Kohlstedt (2003) (HTP)	1.1×10^{-16}	530×10^3	10^{-6}	3.5	-	-	-
Set 2							
Goetze & Evans (1979)* (LTP)	5.7×10^{11}	549×10^3	-	0	8.5×10^9	1	2
Goetze & Evans (1979)* (HTP)	7×10^{-17}	523×10^3	-	3	-	-	-

Table 2: Flow law parameters used in yield strength models. Different figures in the paper refer to Set 1/Set 2. Set 2 parameters are based on the flow law formulation (Goetze & Evans, 1979) with parameter modifications following Mueller & Phillips (1995).

6 Insights from numerical models - summary

The numerical subduction modelling approach was briefly introduced in Section 5.1, while the implementation is described in detail in Supplementary Section S2. This approach, in which stress and strain (rates) emerge dynamically, provides important and complimentary insights alongside the kinematic-based YSEs. This section provides a summary of those insights, and mirrors a more detailed description in Supplementary Section S3.

6.1 Contributions to subducting plate topography

Fig. 4 shows the subducting plate topography from the numerical model, once the stress state, geometry and velocity of the subduction hinge has reached quasi-steady state (5 Myr after start of simulation, same timestep as shown in Fig. 5). We note that the numerical model has a free-surface, and that the model topography (solid red line, Fig. 4) represents the deformed uppermost elements.

The analysis of model topography is described in detail in Supplementary Section S3, and demonstrates that the subducting plate topography has 3 primary contributions. The first is the isostatic adjustment related to the density of the cooling plate, accounting for about $\sim$ 2.5 km of topographic variation from the ridge to the trench. The second is the dynamic topography, mainly associated with anomalous 'dynamic pressure' variation in the asthenosphere, producing about $\sim$ 0.5 km of topography. The (positive) pressure anomaly in the asthenosphere is consistent with the shallow return flow to the ridge Schubert et al. (1978), as well as the accommodation slab rollback; similar levels of topographic response are observed in 3d models (Holt, 2022).

The third contribution, and that of most relevance for our study, is the flexural topography, i.e. the topography related to gradients in shear stress resultant (shown in detail in Supplementary Fig. S4). The black line in Fig. 4 shows the residual (flexural) topography, once the isostatic and dynamic pressure contributions are removed.

6.2 Stress and strain rate patterns

Fig. 5 shows stress and deformation patterns in the numerical model. The scalar fields shown equate to downdip stretching/shortening, and downdip (relative) tension/compression. In the subduction hinge there are two principal regions of deformation: the subducting plate immediately seaward of the trench (the 'outer rise'), where curvature is

increasing, and the unbending zone, where curvature reduces. Stress profiles at four locations are shown in the right hand panel of Fig. 5. The blue line labelled x_0 is the first zero crossing, based on analysis of the flexural component of the topography. The black line lies at the point of maximum bending moment saturation. The difference in the moment at these two points is about 25 %, and is relevant to how we interpret bending moment data.

Although the model is clearly dominated by bending-related stress, it also provides important insights in terms of background stress patterns. The depth-integrated horizontal force balance shows that gradients in net axial force will be related to gradients in the gravitational potential energy (GPE), as well as the shear stress acting on the base of the plate (e.g., Fleitout & Froidevaux, 1983; Ghosh et al., 2009):

$$\frac{d}{dx}\left(F_{\text{net}}\right) = \frac{d}{dx}\left(\bar{\sigma}_{yy}\right) + \tau^{b}_{xy} \tag{5}$$

where the first term on the right hand side is the GPE and the second term is the basal shear stress. In the flexural domain, gradients of the vertical shear stress resultant cannot be neglected, and the vertical stress (σ_{yy}) is expected to deviate somewhat from the lithostatic pressure (P_L). F_{net} can be defined in a number of ways, using the deviatoric stress components (τ), the dynamic pressure ($P' = P - P_L$), or the differential stress ($\Delta\sigma$), assuming one of the principal stresses is vertical (e.g., Schmalholz et al., 2014):

$$F_{\text{net}} = \left(\bar{\tau}_{xx} - \bar{\tau}_{yy}\right) \approx \left(-\bar{P}' + \bar{\tau}_{xx}\right) \approx \left(\Delta\bar{\sigma}\right) \tag{6}$$

Where overbars represent depth integration. Fig. 6 shows the variations of the terms in Eq. 6, based on the numerical model results (same time step shown in Fig. 4). The values were calculated by interpolating and numerically evaluating the integrals down to a depth of 150 km. Away from the trench, the variations in net axial force are mainly controlled by the basal shear stress, while near to the trench, they are dominated by the GPE term. Even though the flexural topography is supported by the strength of the lithosphere, gradients in the vertically integrated stress (the GPE) are still important. In our model, the implied contribution of this 'trench GPE' is about $\sim$ 2.0-2.5 TN/m. It is interesting to note that this estimate of trench GPE has a very similar magnitude to the long-wavelength isostatic GPE attributable to the mid-ocean ridges relative to old oceanic lithosphere (Coblentz et al., 2015).

The trench GPE values we infer are about a third of the magnitude ($\sim$ 7 TN/m) reported by Bessat et al. (2020), by applying the density moment integral to similar numerical subduction models. However, our results suggest that the

density moment would significantly overestimate the trench GPE, due the non-isostatic nature of topography. The key insight from our model is that the net axial force varies in a way that is consistent with trench GPE, this leads to a 'less tensional' background stress state near the trench, where the net axial force almost disappears.

Strain rate patterns in the numerical model are shown in Supplementary Fig. S5. Within the strong part of the plate, strain rates show a linear increase away from the neutral plane, in agreement with the simple bending approximation. The bending rates approximately double between the first zero crossing (x_0) and the point of maximum bending moment, indicating that both curvature and curvature gradient are increasing between these two locations. When numerical model convergence velocities (14.5 cm/y) are re-scaled to a reference value of 10 cm/y, we find that strain rates of $\sim 10^{-15}\mathrm{s}^{-1}$ are typical of the strong part of the lithosphere undergoing ductile creep (40 - 50 km).

6.3 Moment deficit

An important question related to the interpretation of observed moment data is the potential amount of deficit between the bending moment at the first zero crossing, and the saturation moment (e.g., Levitt & Sandwell, 1995). As shown in Fig. 5, the numerical model predicts that the bending moment at x_0 is about 15 % less than the saturation moment. The moment deficit has been assumed to relate primarily from the potential increase in curvature between the two locations (Levitt & Sandwell, 1995; Garcia et al., 2019). Using YSEs calibrated to the same conditions as the numerical model, and knowing the strain rate change between the two points, we find that about 1/3 of the change is due to strain rate increase, while the remaining 2/3 is due to the increase in curvature and the reduction of the elastic core.

7 Interpretation and comparison of bending moment data

7.1 Defining lithosphere strength

In this section the YSE approach is used to develop a simple 'family' of strength models with varying saturation moment. These strength models are used to: 1) define regions in age-moment space (the weak, intermediate and strong regions in Fig. 1), 2) help discuss and interpret different existing age-moment data sets (which is the topic of the next section).

The family of strength models is created by generating a strong 'reference' model, and then simply scaling this model

to achieve weaker variations. The reference model is constructed using the 'Set 1' flow law parameters given in Table 2. We then choose the friction coefficient so that a 30 km neutral plane is achieved for zero net axial force (isotropic background stress) for 100 Myr lithosphere. In this exercise we assume zero pore fluid pressure for simplicity. The required friction coefficient to generate the 30 km neutral plane is unrealistically high ($\mu = 1.5$). However, the aim at this stage is simply to provide an upper limit for subsequent analysis. This reference model is shown by the outermost model in the left hand panel of Fig. 2. The upper bound to the 'strong' region (shown in red) in Fig. 1, represents the variation of strength with age for the reference model.

Having constructed this reference model, we simply scale the YSE (or the age-moment relationship) to achieve an arbitrary strength (saturation moment). Based on this scaling approach, we describe three strength regions, which are defined by the value of the saturation moment for 100 Myr lithosphere: *weak* models have saturation moments (<1); *intermediate* models (1-2) and *strong* models (>3), measured in units of N-m/m (N) $\times 10^{17}$. These regions are shown in green, tan, and red, in Fig. 1 and in Fig. 2. We also use this scaling approach as a basis for fitting observed moment data (as we describe in the next section). The solid blue line in Fig. 1 is the result of such an approach, i.e. it is simply a scaled version of the original model discussed above.

The solid black and red lines in Fig. 1 and the right hand panel of Fig. 2, show two different 'intermediate' strength models from recent subduction-flexure studies. The black line shows the preferred model from Garcia et al. (2019). The red line shows one of the models that was determined to be consistent with flexural inversion from the study of Hunter & Watts (2016) (they find that several combinations of flow law parameters and friction coefficients are viable). We use an identical thermal model to these studies. We note that the age variation of strength models is rather insensitive to the choice of parameters. The solid blue, black, and red lines in Fig. 1 have virtually identical moment-age variation, although having significant differences in strength-depth patterns, particularly in the LTP domain (as shown in Fig. 2).

7.2 Bending moment estimates: interpreting scatter

The object of this section is to bring together estimates from a range of previous studies for the bending moment of old lithosphere; ideally these estimates represent locations close to moment saturation, an assumption that is assessed in the analysis. A key focus is how the data-set of Levitt & Sandwell (1995) is interpreted in terms of competing strength models.

The observed moment data of Levitt & Sandwell (1995) show an age dependency, as well a significant component of scatter (e.g. Fig. 1). We want to interpret this data in terms of competing lithospheric strength models, which in turn requires a model for the scatter. We expect there to be 3 major contributions: 1) real variations in the saturation moment for lithosphere of given age (for instance, due to variations in the net axial force, thermal perturbations or strain rate variations); 2) a deficit between the bending moment at the point being measured (e.g. the first zero crossing) and the saturation moment ('moment deficit'); 3) estimation errors. The moment deficit effect can only lead to underestimates. The other types of effects could in principle lead to random or systematic bias in the scatter.

Levitt & Sandwell (1995) assumed that the scatter primarily derives from a moment deficit at the first zero crossing. Hence, they appealed to a strength model in which the age-moment curve provides a ceiling to the observed moment data. Models in the 'strong' range tend to satisfy this interpretation. Their preferred model has a saturation moment of about 2.5 $\times 10^{17}$ N, for 100 Myr lithosphere, and is shown with the dashed blue line in Fig. 1. Strength models in the 'strong' range are consistent with relatively high friction coefficients, comparable to those determined by experiments (Byerlee, 1978); depending on assumptions regarding the thermal model and strain rate, they are also consistent with a variety of LTP laws.

However, the interpretation of Levitt & Sandwell (1995) implies that in many cases, at the first zero crossing, the plate is considerably 'under-saturated'. For instance, many of the observed moment values of $< 1 \times 10^{17}$ N for lithosphere $>$ 100 Myr, would imply bending moments of about $\sim 1/3$ of the saturation moment. As shown in Supplementary Fig S7, for the curvature estimates corresponding to the moment data, all profiles are expected to have reached $>$ 65 % moment saturation, with the majority ($>$ 70 %) being $>$ 75 % saturated. This distribution is consistent with the behavior observed in the numerical model, which exhibits about 80 % moment saturation at the first zero-crossing. Additionally, Supplementary Fig. S7 shows that many of points that show the highest deficit have high curvature. Indeed, the trend of the moment deficit with curvature is opposite to what would be expected. Hence, we argue: that 1) where the lithosphere has a moment deficit at the first zero-crossing point, this deficit will typically be less that 25 % of the saturation moment, and 2) moment deficit is inconsistent with the trends of the scatter in the data of Levitt & Sandwell (1995). This implies that estimation errors also play a significant role.

If we assume that the combined effect (both real effects and error) is to produce random (or at least not heavily biased) scatter, then we require a rheological model that minimises the misfit. For the family of strength models described in the previous section, the misfit is minimised (RMS) with a model that has a friction coefficient of $\mu = 0.25$. This model is shown with the solid blue line in Fig. 1. It lies in the lower half of what we call the intermediate strength region, and exhibits a typical saturation moment close to 1.5 $\times 10^{17}$ N, across the range of old lithosphere (100-150

Myr).

7.3 Bending moment estimates: comparison of different datasets

In the previous section we suggested that a combination of effects (both real effects and error) could lead to to more-or-less random scatter in the Levitt & Sandwell (1995) data-set. Additional studies provide an important test of this interpretation.

As shown in Fig. 1, there is a considerable consistency between the studies of Chapple & Forsyth (1979); Hunter & Watts (2016); Garcia et al. (2019), and our reinterpretation of the Levitt & Sandwell (1995). These studies each infer strength models in the mid-intermediate strength range, and lie within about ~ 20 % of the best-fitting strength model we developed for the Levitt & Sandwell (1995) data. In terms of comparing these datasets, there are two additional features that are important to highlight.

Firstly, we note that moment estimates from Hunter & Watts (2016) (red circles Fig. 1) represent 'trench-bin' averages: each of these points contains information averaged over a typical distance of several thousand kilometers. On the other hand, the data points from Levitt & Sandwell (1995) represent flexural inversion over single bathymetry/gravity profiles. The fact that an average (or a best fit) interpretation of the Levitt & Sandwell (1995) data, leads to very similar moment estimates to the 'trench-bin' results of (Hunter & Watts, 2016), helps clarify the connection between these two studies.

Secondly, we note that the results (moment estimates) of the aforementioned studies are generated under a range of different modelling approaches. For instance, these inversions employed a range of forward flexural modelling techniques: constant-thickness elastic plate (Levitt & Sandwell, 1995), variable thickness elastic plates (Hunter & Watts, 2016), non linear flexure with ad-hoc yield limits (Chapple & Forsyth, 1979), and non-linear flexure based on classical yield strength model (Garcia et al., 2019). These modelling methodologies are summarised in Section 3. This suggest that a range of different approaches are able to consistently determine that the bending moment of the lithosphere saturates in the 'intermediate' strength range.

What is also clear in the compilation of data in Fig. 1, is that only the dataset of Levitt & Sandwell (1995) has been used to advocate a model in the 'strong' region. However that conclusion was tied to a specific interpretation of the scatter, which has proven to be inconsistent with results from additional studies (Hunter & Watts, 2016; Garcia et al., 2019). Of the data compiled in Fig. 1, the only results are not consistent with intermediate strength are those of

McNutt & Menard (1982) (green points). These estimates mainly lie in upper part of the weak region. As in Levitt & Sandwell (1995), these estimates represent results from individual profiles, but unlike Levitt & Sandwell (1995) they are based on modelling bathymetry data, rather than joint inversion that includes gravity. This is expected to make them less reliable. However, as we noted the Levitt & Sandwell (1995) data also contains points (outliers) in the weak strength region. These do not appear in trench-averaged approaches (e.g., Hunter & Watts, 2016). For these reasons we think it is appropriate to disregard the estimates of McNutt & Menard (1982). However, to the extent that there is valid information in the McNutt & Menard (1982) data, it would imply an even weaker strength model.

The model we generated to fit the Levitt & Sandwell (1995) data has a low effective friction coefficient (μ =0.25, assuming zero pore fluid pressure). We note that a range of recent studies have converged on similar values for the frictional strength (Garcia et al., 2019; Pleus et al., 2020; Bellas et al., 2022), as shown in Fig. 2. The models of (Garcia et al., 2019; Hunter & Watts, 2016) both assumed hydrostatic pore pressure, which offsets the slight higher friction coefficient (μ =0.3), relative to the best-fitting model shown in Fig. 1 (μ =0.25). While this convergence towards low frictional strength is worth highlighting, there are important trade-offs between strength in the brittle and ductile regimes. As we discuss in the following section, this means that the frictional strength is not as well constrained as it might appear on the basis of the results presented in this section.

8 Strength and non-uniqueness

8.1 Relative strength and background stress

Different combinations of brittle and ductile strength can be combined to produce a strength model with the same saturation moment (e.g. an intermediate strength model). The depth of the neutral plane seems to offer the most direct means of trying to constrain the relative strength in each regime, as was pointed out in Chapple & Forsyth (1979). However, this is hampered by the additional influence of background stress. This leaves us with 2 main sets of observational constraints: those which reflect the integrated plate strength (e.g. the saturation moment) and the neutral plane depth. However, there are at least 3 key 'variables': brittle strength, ductile strength and the net axial force (factors like the convergence rate, orientation of seafloor fabric, out-of plane stress, are expected to provide additional controls). We can think of these key variables as a parameter space. Together they lead to predictions of the saturation moment as well as the neutral plane depth (as well as additional predictions that can be compared to the seismic expression). Despite the non-uniqueness, it may be still be useful to try to investigate different regions of

the parameter space, particularly in reference to additional constraints.

Strength models proposed in previous studies span a wide range within the parameter space. We highlight 4 such models to illustrate the problem. These are shown in Fig. 7, evaluated for 100 Myr lithosphere. We refer to these as models A, B, C, D, following the labels on the Figure axes. Three of these these models (A,B,C) utilise variations of the LTP formulation originally described in Goetze & Evans (1979), the ('Set 2' parameters shown in Table 2). The ductile strength in the HTP/LTP regions is varied through changes the activation energies, as described in the Figure caption. In all cases the LTP activation energy lies in the uncertainty range specified by the original study (100-130 kcal/mole, or 418-544 kJ/mole). In this sense, all of the models within Fig. 7 have acceptable ductile strength, within the range of uncertainty given in Goetze & Evans (1979).

Each of the models shown in Fig. 7 have intermediate strength, but the relative strength in the brittle and ductile regions is different and hence the intrinsic neutral plane depth varies. Therefore, the models imply differences in terms of the background stress required to facilitate regions with a 30 km neutral plane (which produce the apparent neutral plane depth shown in Fig. 3). Each model is shown with different levels of net axial force, as labelled in the legend shown on Fig. 7B.

The maximum seismogenic limit for each of the models is shown with T_s. This estimate is based on different assumed mechanisms for brittle failure in compression. The estimate is highly speculative for Models C and B and is based on a 'stress amplification' model described in the following section.

We will try to evaluate the consistency of these different strength models in light of additional constraints. The constraints primarily relate to the seismic expression of outer rise bending, as inferred from the distribution of well-constrained centroid depths (e.g. Fig. 3), and following a similar logic to Chapple & Forsyth (1979). The key features we refer to are the depth limit of normal faulting ($\sim$ 30 km); the maximum depth of limit of (reverse) earthquake nucleation ($\sim$ 50 km); the capacity to predict a neutral plane depth of $\sim$ 30 km within assumed variations of background stress. These constraints are relevant to older plates, particularly the range 80-120 Myr and are shown schematically with the blue and red bars in Fig. 7. The general pattern of strength increase with age leads us to expect that the neutral plane depth would increase by about 5 km for very old lithosphere ($\sim$ 150 Myr), as shown in the top panel of Fig. 1.

We differ from Chapple & Forsyth (1979) in that we do not assume that these features are characteristic of an isotropic background stress state; rather we argue that a consistent strength model should be capable of generating these

features within reasonable variations in background stress. Several other constraints are also referred to: independent constraints on the frictional strength of outer rise normal faults; the capacity to explain the anomalously shallow neutral plane ($\sim$ 25 km) along Tonga-Kermadec (e.g. Fig. 3); as well as *a priori* assumptions about the background stress state. Although Fig. 7 shows specific strength models, the discussion here is intended to represent, somewhat more broadly, the regions of the parameter space represented by these models.

8.2 Brittle failure in compression

Before we evaluate the strength models, we need to address the issue of the mechanism of brittle failure in compression. Model A is the only model that can directly explain this brittle failure in terms of frictional sliding. In model D, the brittle behaviour in the lower part of the seismogenic zone ($>$ 20 km) is assumed to reflect processes other than frictional sliding, and the magnitude of the brittle strength is selected try to match the bending moment inferences (following Chapple & Forsyth (1979)).

Models B and C do not provide an explicit mechanism for brittle failure beneath the neutral plane, where the strength is instead limited by LTP. This is a common feature of most classical strength models, although it has seldom been viewed as a critical limitation. Several studies have taken a different view, and insisted that a consistent strength model needs to account for brittle failure in the depth range where outer-rise compressional earthquakes are observed (Chapple & Forsyth, 1979; Mueller et al., 1996). We tend to agree with this perspective. Even if there is a significant depth range in which the constitutive behavior is semi-brittle, it doesn't seem satisfactory to simply assume that the stress in a semi-brittle domain should be equivalent to the LTP strength – a rheological process that is itself unrepresentative of brittle failure, even if it may coexist with brittle failure mechanisms transiently.

One possibility was recently suggested by Toffol et al. (2022), termed a 'stress amplification' model. While their study is focused on intermediate depth seismicity, it also has relevance for reverse faulting in the outer rise. The model proposes that weak domains of hydrated peridotite can create large stress amplifications which are capable of exceeding the frictional strength, even at large confining pressure. An interesting feature of this model is that although it can produce large stress amplifications, the 'flexural characteristics' (neutral plane depth, bending moment) appear to be insensitive to the perturbations (see e.g. Toffol et al. (2022) Fig. 5). This is unsurprising given flexural strength depends on the integrated stress state (resultants, moments) both across the plate and laterally. Stress amplification can therefore provide a solution as to why brittle behaviour might occur within a rock mass where the macroscopic strength (e.g. the saturation moment) is determined by the LTP strength.

Stress amplification requires that strain rates can increase above the bulk plate-scale value around weak inclusions. We do not know what strain rates are appropriate, but we note that the amplification magnitudes reported in Toffol et al. (2022) vary from about 2-5 times. To characterise a seismogenic thickness (T_s) for model B and C, we assume that stress amplification can increase the LTP stress locally by 5 times, and define the T_s is the intersection of the perturbed stress and the frictional strength. As in Toffol et al. (2022), we have neglected the fact that fracture strength will place an upper limit on frictional strength, with 1 GPa being a typical limit strength based on experiments in igneous rocks in the 500 – 600 °C range (Stesky et al., 1974; Shimada et al., 1983; Ohnaka, 2013, e.g.,).

We adopt this stress amplification model as a tentative solution to brittle failure in compression for Models C and B. For model A, $T_s(1)$ refers to the seismogenic thickness based on the background stress state, whilst $T_s(2)$ refers to the seismogenic thickness based on a stress amplification model.

8.3 Evaluating strength models

Model A is very similar to the preferred model proposed by Garcia et al. (2019). The friction coefficient is low, consistent with the inferred dip angles of newly-formed outer rise faults (Craig et al., 2014b), and a relatively high pore fluid pressure is assumed ($\lambda = 0.36$). The intrinsic neutral plane depth is about 37 km. A compressional net axial force of -5 TN/m results in a neutral plane depth of 32 km (close enough to be considered consistent with the apparent neutral plane depth). Model A provides an explanation for brittle failure beneath the neutral plane in terms of frictional sliding, although the maximum depth of frictional behaviour is about 40 km ($T_m(1)$) and hence additional mechanisms of brittle failure may be required to facilitate earthquakes to $\sim$ 50 km. A stress amplification model could explain these depths ($T_m(2) \sim 50$). In regions of of moderate slab pull (5 TN/m), model A predicts normal faulting to 42 km, significantly deeper than the 30 km limit that is typically reported; this would be even deeper if the strength model was evaluated for lithosphere older than 100 Myr. Meanwhile, model A implies that regions that exhibit a $\sim$ 30 km neutral plane are in moderate net axial compression (-5 TN/m). This could be considered problematic in light of *a priori* assumptions about tectonic stress in subducting plates. Finally, even with moderate compression, the neutral plane (32 km) is significant deeper than inferred across Tonga-Kermadec ($\sim$ 25 km).

Model B is based on a strength model proposed by McNutt & Menard (1982). In this model the HTP/LTP strength is reduced by changing the activation energy, while the frictional strength remains relatively high (note also that there is no pore pressure in this model). The intrinsic neutral plane depth is about 26 km. A tensional net axial force of 5 TN/m results in a neutral plane of about 30 km. Hence, model B predicts a 30 km neutral plane under conditions

that are consistent with inferences of moderate net slab pull (Bird et al., 2008). Clearly much higher estimates are also present in the literature (e.g., Conrad & Lithgow-Bertelloni, 2002). The seismogenic depth (based on the stress amplification model) in model B is very low $T_s = 35$ km. This is difficult to reconcile with the observation of earthquake hypocenters to at least 50 km in lithosphere of 80-120 Ma. Secondly, the brittle strength ($\mu = 0.6, \lambda = 0.0$) implies dry faults, which seems unlikely given geophysical inferences of outer rise hydration (Ranero et al., 2003), as well as the preponderance for *in-situ* intraplate stress measurements to conform to hydrostatic strength expectations (Zoback & Townend, 2001).

Model C is similar to the one presented in the study of Goetze & Evans (1979) (see caption for details). It balances strength in the brittle and ductile domains so that the intrinsic neutral plane depth is $\sim$ 30 km. The assumption is hydrostatic Byerlee strength. Within the assumed range of background stress variations, model C remains consistent with the observed maximum depth of outer rise normal faults. The friction coefficient is typical of laboratory experiments, *in situ* stress measurements, and some observations of outer rise fault dip angles (Obana et al., 2023). However, it is significantly higher than the seismological constraints of Craig et al. (2014b). For Model C, the proxy for the seismogenic thickness (T_s) is 41 km, which is smaller that the maximum depth extent of outer rise seismicity, but not as dramatic as the disagreement in model B. We note a trade-off is emerging here: models which more accurately reproduce the maximum depth of normal faulting, tend to also have a T_s that is too shallow. Allowing for moderate compressional stress, model C predicts a neutral plane of 27 km, fairly close to the inferred depth along Tonga-Kermadec.

Model D is a variation on the *ad hoc* yield-strength model of Chapple & Forsyth (1979). It is therefore automatically consistent with some of key constraints, such as the maximum depth of seismicity (T_s is 50 km by construction). In the original study, the differential stress is truncated at 100/600 MPa, whereas in Model D, the truncation is 100/400 MPa. This provides a bending moment similar to the other models shown in Fig. 7. The model was originally presented and discussed in the context of an isotropic background stress state. It is useful, however, to consider the consistency of the model in relation to assumed variations in background stress. Overall, model D has the greatest sensitivity in terms of the variation of neutral plane depth as function of background stress. Under moderate net axial compression, model D can explain the 25 km neutral plane depth inferred at Tonga-Kermadec. However, like model A, under moderate net axial tension, the predicted depth of normal faulting (37 km) becomes deeper that is typically observed.

In our view Model B, based on McNutt & Menard (1982) is the least consistent in view of additional constraints. Fundamentally, the ductile strength is so low that seismic rupture beyond about 35 km seems unlikely. An implication of rejecting model B, is that regions that give rise to the apparent 30 km neutral plane depth cannot be supporting a

net axial force of $\sim$ 5 TN/m (or higher). We expand on this argument in the following section.

None of the other models (A, C, D) is completely consistent with additional constraints. However, rather than advocate a preferred model, we think it sufficient to summarise the predictions of each models, and the areas of (in)consistency, under the assumptions we have outlined. Our main intention here is to highlight the constraints that can be brought to bear on different strength models, and the challenges that remain in successfully explaining the deformation mechanisms, brittle behaviour, and relative strength properties (Chapple & Forsyth, 1979).

Of the models shown in Fig. 7, model A is unique in that it offers a direct explanation for outer rise reverse faulting in terms of frictional sliding. Previous studies have argued that a low friction coefficient would generate an inconsistent strength model and therefore could not be invoked to explain brittle failure in compression (Chapple & Forsyth, 1979; Mueller et al., 1996; McNutt & Menard, 1982). However, this mechanism is not able to explain earthquake nucleation to the full extent of the observed depth range $\sim$ 50 km.

We invoked the stress amplification model of Toffol et al. (2022) to try to quantify a relevant seismogenic depth in models where the macroscopic strength is governed by LTP (e.g. B and C). Even under rather lenient assumptions (stress amplification of a factor of 5, the maximum reported in Toffol et al. (2022)) the seismogenic thickness remains too shallow. Interestingly, the stress amplification model works more successfully for model A, which is again due to the low friction coefficient assumed in that model.

Overall, the presence of deeper compressional seismicity, and particularly the capacity for earthquake ruptures to nucleate (most clearly illustrated by the smaller-scale seismicity shown on Fig. 3) remains difficult to reconcile with existing rheological formulations for any of the models shown. Resolving this will require both additional observational data, to better understand the spatial distribution of this deeper compressional seismicity, and additional work on the nature of seismicity close to the brittle to ductile transition (e.g., Aharonov & Scholz, 2019; Toffol et al., 2022).

9 Discussion

9.1 Constraints on net slab pull

In the previous section we suggested that regions of old lithosphere which exhibit a 30 km neutral plane cannot be supporting 'typical' levels of net slab pull (e.g. 5 TN/m). This reasoning was based on the inconsistent properties

of a strength model that could give rise to a 30 km neutral plane depth under this background stress state. Now we consider constraints on net slab pull more generally.

The net axial force is given by the integral of the differential stress across the thickness of the lithosphere (e.g. Equation 6). In simple concave-down bending, net slab pull must be transmitted entirely through the upper part of the YSE (i.e. above the neutral plane). Observations of neutral plane depths, in conjunction with estimates of the brittle strength, imply limits on net slab pull. Assuming a friction coefficient of 0.6, along with hydrostatic pore pressure (λ= 0.3), leads to a maximum $F_{\text{net}} = 5$ and 10 TN/m for 30 and 40 km neutral planes respectively.

Of course the above estimates are upper limits, as they factor in no compressional stress beneath the neutral plane. Observations of seismicity, as well as laboratory estimates of olivine creep, provide compelling indication for significant strength to at least 600-700 °C, or $\sim$ 50 km (for 100 Myr lithosphere). A simple assumption we can make is that the ratio of net force in the extensional and compressional parts of the envelope scales with ratio of depth above and below the neutral plane, to 50 km. In this case, estimates of the maximum net slab pull reduce to 1.6 and 7.5 TN/m for 30 and 40 km neutral planes in $\sim$ 100 Ma lithosphere.

If the 30 km neutral plane is assumed to be globally representative for old lithosphere (e.g. Chapple & Forsyth (1979)), we could rule out net slab pull being greater than a few TN/m. However, if the apparent neutral plane depth is biased by regions of anomalous compression, then slab pull may be focused along other regions where the neutral plane is deeper (although likely to be unconstrained due to the absence of reverse seismicity). Nevertheless, this simple calculation shows the geodynamic relevance of constraining the neutral plane depth, and therefore the importance of earthquake hypocenter accuracy.

9.2 Intraplate stress variations: potential implications for outer rise seismicity

A characteristic of outer rise reverse seismicity is the uneven distribution of events relative to the shallower normal faulting (Mueller et al., 1996; Craig et al., 2014b; Ye et al., 2021). Global compilations reveal clustering in certain regions as well as the absence of moderate-size compressional events along significant ($>$ 2000 km) lengths of trench (as discussed in Section 3.3). We now consider the potential role of intra-plate stress variations in controlling such variability. To explore this idea, we follow Zoback et al. (2002) and adopt a value of $\pm$ 5 TN/m to represent the typical magnitude of the integrated intra-plate stress variations.

Although the mechanism of brittle reverse seismicity remains opaque (as discussed in the previous section), it is

instructive to consider general dynamic interactions between the net axial stress and bending. To demonstrate, we consider model C (e.g. in Fig. 7C). Increasing the net force by +5 TN/m results in an increase the neutral plane depth by about 4.5 km. This has the effect of reduces the maximum differential stress that is developed beneath the neutral plane. Furthermore, the integrated volumetric strain strain rate beneath the neutral plane will be reduced. This is because in simple-bending, the axial strain rate depends on distance from the neutral plane.

A useful way to quantify these dual effects is to consider the analastic dissipation rate, as was proposed by Lorinczi & Houseman (2009) for understanding the distribution of seismic moment. Fig. 8 shows the predicted dissipation for model C, with the same prescribed variations in net axial force. The strain rate is based on a simple bending (rate) model, being proportional to distance from the neutral plane. The results demonstrate that the dissipation beneath the neutral plane is particularly sensitive to the neutral plane depth, and hence the background stress.

In this respect we think that the typical magnitude of intra-plate stress variations, producing $\sim$ 5-10 km increase in the neutral plane depth (assuming 100 Myr lithosphere), may be sufficient to explain the absence of seismicity along some margins. In this view, outer rise reverse faulting is a marginal process: it probably requires a neutral-to-moderately compressional background stress, and may be largely suppressed with only moderate changes in the background stress. In all of these situations, however, the stress state remains dominated by bending, while the background stress plays a modifying role (e.g Fig. 7). This presents a counterpoint to the argument of Mueller et al. (1996), that links outer rise compressional seismicity to anomalous, spatially-localised subduction resistance, where axial forces dominate.

Tonga-Kermadec provides some important observations in light of the framework described above. This region stands out globally as having the most numerous and broadly-distributed pattern of outer reverse faulting (see Ye et al. (2021), Fig. S5). It also stands out as having an unusually-shallow neutral plane depth; Northern Tonga is particularly atypical, but across the entire system an apparent neutral plane depth of 25 – 30 km seems a reasonable inference (e.g. Fig. 3). These two factors support the argument above: a more compressional background stress state results in a shallower neutral plane, and much higher dissipation in the compressional part of the plate. Of course, this inference about the background stress state is consistent with the longstanding idea that the Tonga-Kermadec slab is anomalously compressional (Isacks & Molnar, 1971; Gurnis et al., 2000).

9.3 Energy dissipation in bending

Conrad & Hager (1999) suggested that bending of the lithosphere during subduction may constitute a major component

of dissipation in mantle convection, being around 60% of a slab's potential energy release. Several studies have presented a counter arguments, based on kinematic models incorporating more sophisticated rheological models (Buffett & Becker, 2012), numerical modelling based on multi-layer strength models (Capitanio et al., 2009) and convective scaling arguments (Leng & Zhong, 2010).

Buffett & Becker (2012) estimated much lower bending dissipation values: $\sim$ 10 % of the slab's potential energy. This conclusion was developed for a strength model with a saturation moment of 4.5×10^{17} N (estimated for old lithosphere). In this study we have argued that a strength range of $1 - 2 \times 10^{17}$ N is compatible with data for old lithosphere ($>$ 100 Myr). As the dissipation is linearly dependent on the saturation moment (for a given subduction hinge geometry), our study suggests further reduction in terms of the importance of bending dissipation for mantle convection.

10 Conclusions

The current study has a close affinity to the work of Chapple & Forsyth (1979); they combined flexural modelling with seismological constraints in order to develop a consistent (albeit *ad hoc*) model for the yield strength distribution of old oceanic lithosohere. The resulting strength model remains broadly consistent with a range of additional studies and data (as shown in Fig. 1).

We survey a range of historical seismicity data, including well-constrained centroid depths of outer rise earthquakes. This data increasingly supports the view that normal faulting transitions to reverse faulting within numerous segments of the global subduction zone outer rise system, consistent with a stress state dominated by flexure (as was originally interpreted by Chapple & Forsyth, 1979), but questioned by other studies (e.g., Mueller et al., 1996)).

We highlight the considerable consistency between the studies of Chapple & Forsyth (1979); Levitt & Sandwell (1995); Hunter & Watts (2016) and Garcia et al. (2019), regarding the strength of old oceanic lithosphere. All of these studies can be interpreted as favouring 'intermediate' strength models of the lithosphere. While there may still be a discrepancy in the strength inferred in seamount versus subduction-related flexure, the problem could be less acute than proposed by Bellas et al. (2022).

The way in which different deformation mechanisms interact to produce this 'intermediate' strength is less clear. In trying to constrain the relative strength distribution with depth, the neutral plane constitutes an important constraint.

In contrast to Chapple & Forsyth (1979), we do not assume that the 'apparent' neutral plane depth (30 km for 80-120 Myr lithosphere) is equivalent to the 'intrinsic' neutral plane depth. That is, we allow for the fact that the apparent neutral plane depth could be biased by regions due to systematic (non-isotropic) background stress deviations.

We characterise this problem as a parameter space, broadly defined as the relative strength in the brittle (extensional) regime, the strength in the ductile region, as well as the net axial force. In doing so do we need to make certain *a priori* assumptions about what might constitute reasonable variability in the background stress. We consider a range of end member models, based on previous studies, each with intermediate strength characteristics, but different relative strength characteristics. The problem is inherently non-unique, but additional constraints are of some value in assessing such models.

Evaluating these strength models in light of additional constraints leads to several conclusions: 1) none of the investigated models are entirely consistent with additional constraints, if the assumed variations of net axial force ($\pm$ 5 TN/m) are applicable to the global subduction system; 2) in our view only one of these models can be ruled out. This end-member is characterised by high frictional strength, and very low ductile strength, as proposed by McNutt & Menard (1982). 3) in regions of $\sim$ 100 Myr lithosphere, which exhibit a 30 km neutral plane, net slab pull should be no greater than about 1-2 TN/m; 4) models characterised by very low frictional strength (e.g., Garcia et al., 2019) are considered viable. However, they imply that regions exhibiting a $\sim$ 30 km neutral plane are under moderate net axial compression ($\sim$ -5 TN/m). Under these conditions, frictional reverse faulting is predicted beneath the neutral plane at depths of $\sim$ 30-40 km.

In the Discussion Section we address the global variability in outer rise seismicity (particularly the presence/absence and clustering of reverse seismicity). We again consider this variability in terms of the impact of background stress state. We show that the magnitude of the anelastic dissipation beneath the neutral plane is particularly sensitive to the neutral plane depth, and hence the background stress. We suggest that outer rise reverse faulting may be a marginal phenomenon, with moderate variations in the background stress being sufficient to inhibit brittle behavior. In the framework proposed here, background stress plays a modifying role, with stress patterns still dominated by flexure (c.f Mueller et al. (1996)).

The study integrates numerical modelling to support the analysis and data interpretation. Some key insights from the numerical model are: 1) the subducting plate is more than 80 % moment saturated at the first zero crossing in the outer rise; 2) in the strong part of the lithosphere, strain rates are well approximated by the simple bending model (proportional to distance from the neutral plane); 3) representative strain rates in the strong ductile part of

the lithosphere are $\mathcal{O}(10^{-15})\mathrm{s}^{-1}$ for subduction velocities of $\sim$ 10 cm/y; 4) the net axial force varies throughout the trench-outer rise domain, and is closely correlated with the estimated contribution of the GPE. This leads to negligible net axial force beneath the trench.

Acknowledgements

This research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. DS was supported by the Australian Research council (Discovery grant DP150102887). TJC was supported in this work by the Royal Society under URF\R1\180088. TJC was also supported through COMET, the UK Natural Environment Research Council's Centre for the Observation and Modelling of Earthquakes, Volcanoes, and Tectonics. The authors thank the developers of open source software packages that enabled this research (in particular ASPECT, The Geodynamic World Builder, SciPy, Numpy, Matplotlib, PyVista) Bangerth et al. (2020); Fraters et al. (2019); Virtanen et al. (2020); Harris et al. (2020); Hunter (2007); Sullivan & Kaszynski (2019). The authors thank the reviewers Fan Zhang and Shijie Zhong for their constructive reviews which have helped to improve the paper. We also thank Fern Storey of GJI for their assistance.

Data availability

Input files and a description of code modifications to reproduce the numerical model can be found at `https://github.com/dansand/subduction_GJI2022`. Data plotted in Fig. 1 were sourced from tables in previous studies (Levitt & Sandwell, 1995; Garcia et al., 2019; Craig et al., 2014a), except for data of Hunter & Watts (2016), which were redigitised by the authors. Data in Fig. 3 are from Craig et al. (2014a). Microseismicity data are from Hino et al. (2009), Gamage et al. (2009), Obana et al. (2012), (Obana et al., 2014), and Obana et al. (2018) for Honshu, and Zhu et al. (2019) and (Chen et al., 2022) for the Marianas (shallow and deeper regions, respectively).

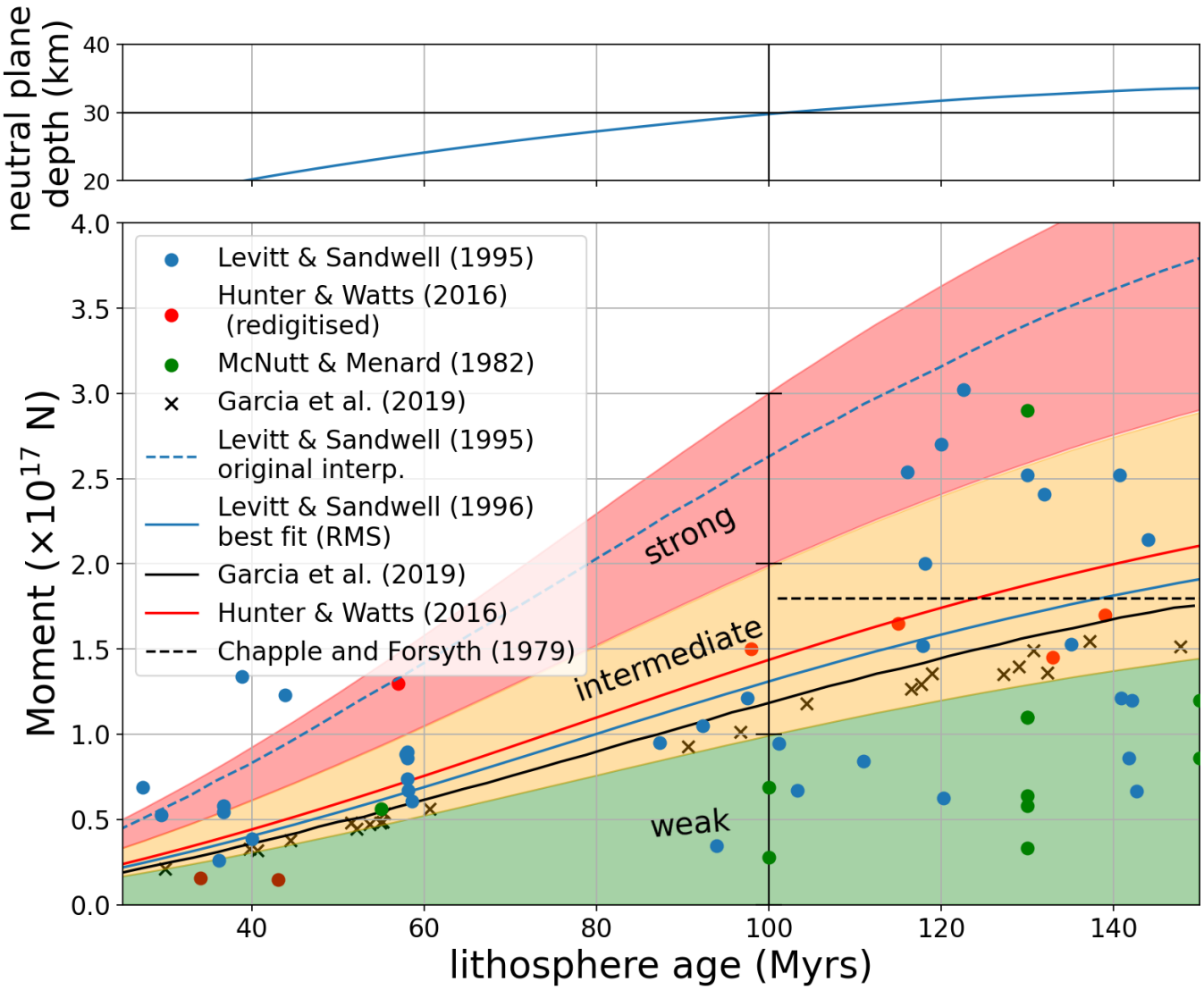

Figure 1: Main panel: Bending moment variation with age. The plot contains predicted age-saturation moment relationships for various strength models, as well as various published data. Filled circles shows observed moments from Levitt & Sandwell (1995) (blue). Black crosses show bending moments from the forward models of (Garcia et al., 2019) extracted at the point of maximum curvature; these points are within 5-10 % of the saturation moment for the 'weak' strength model investigated (shown with solid black line). Red circles show the bending moment at the first zero crossing, re-digitised from the paper of Hunter & Watts (2016), using the variable elastic thickness results (A2). As discussed in Section 7, we developed a family of models by varying the friction coefficient and the LTP strength subject to the neutral plane constraint (30 km) for 100 Myr lithosphere, with zero net axial force. These models are used to define *weak*, *intermediate* and *strong* model regions. The solid blue line shows the member of this family of models that minimises the misfit (RMS) for the data-set of Levitt & Sandwell (1995); the dashed blue line shows the strength model inferred in the original study. The dashed black line shows the strength model proposed by Chapple & Forsyth (1979), chosen to fit data for 'old' subducting plates. The top panel shows the variation in neutral plane depth as a function of age.

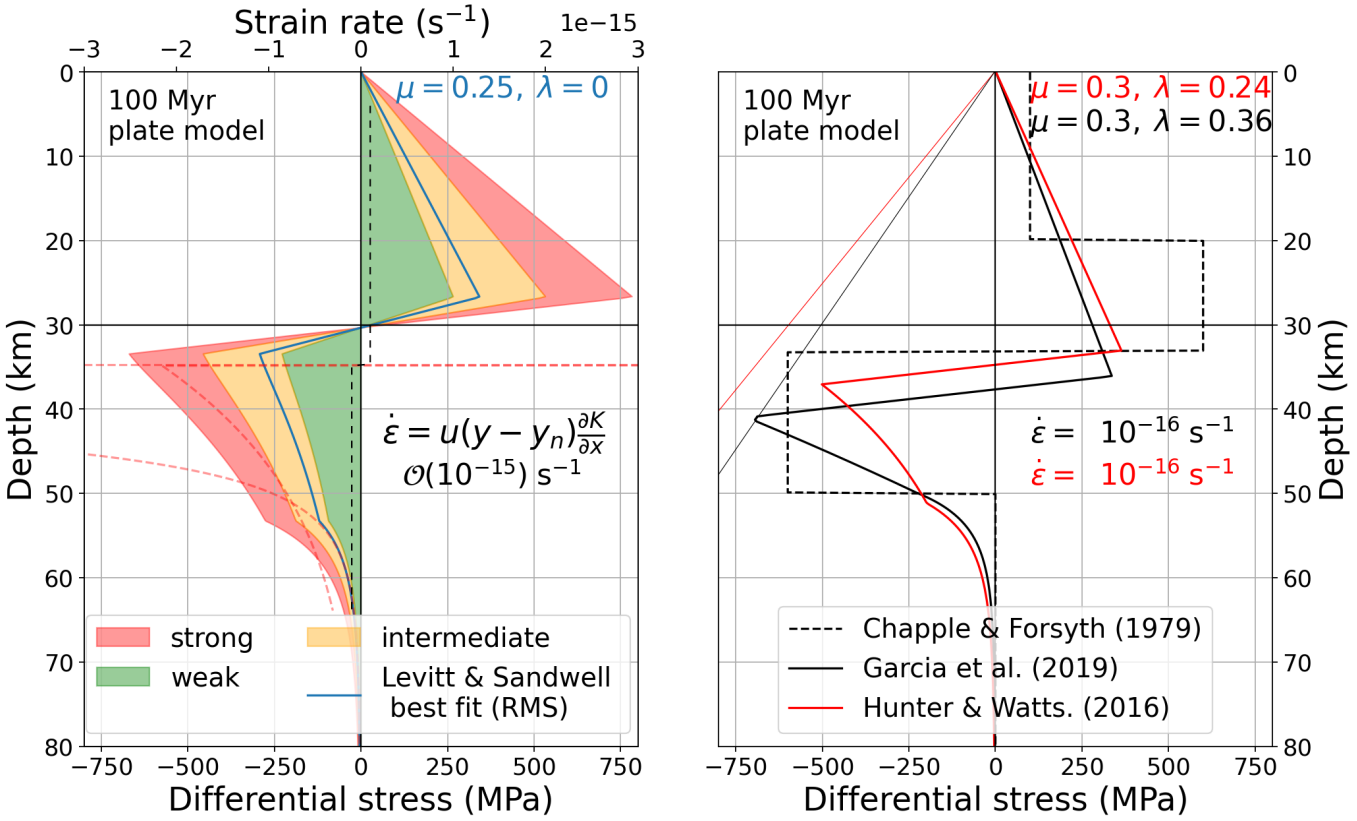

Figure 2: Yield stress envelopes evaluated at moment saturation, for 100 Ma lithosphere, corresponding to the various strength models shown in Fig. 1. The left hand panel shows the construction of a simple family of strength models, as discussed in Section 7. The strongest model are based on dry olivine flow law parameters of Mei et al. (2010) (LTP) and Hirth & Kohlstedt (2003) (HTP) (shown with dashed red lines). The strain rate is based on the bending rate assumption (discussed in the Supplementary Section S4) and is shown with the dashed black line. Successive models are simply scaled in the stress axes, to define a 'family' of models. The solid blue line shows the member of this family that minimised the RMS misfit to the observed moment data of Levitt & Sandwell (1995). The right hand panel shows additional strength models (also shown in Fig. 1). The solid black line is the preferred model of Garcia et al. (2019). The solid red line shows one of the models argued to fit subduction flexure data in Hunter & Watts (2016), based on the dry olivine flow law parameters (e.g. Set 1, Table 2) . Both of these strength models are evaluated with a constant strain rate of $10^{-16}s^{-1}$ following the original studies. The pore pressure factors (λ) were established by reanalysis of the YSEs presented in the original studies.

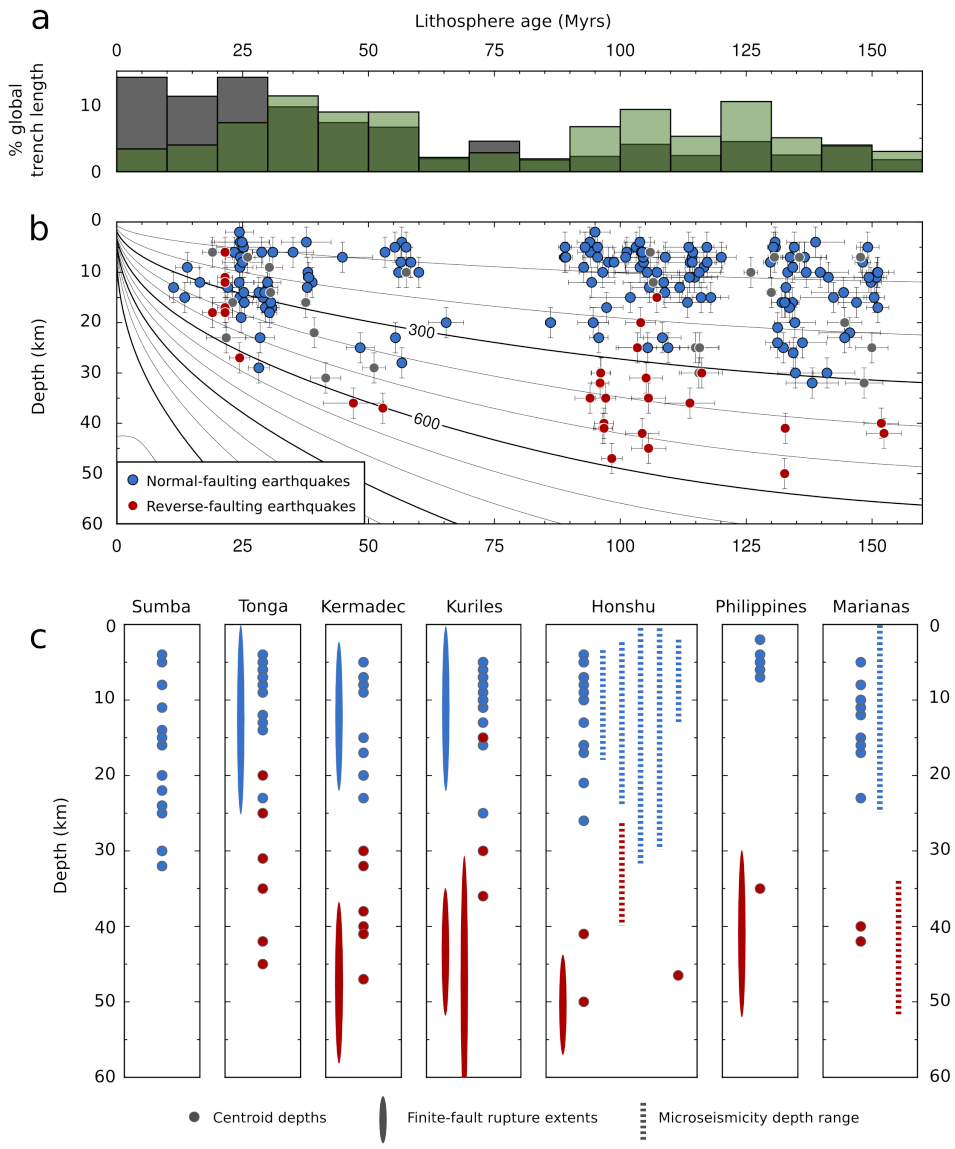

Figure 3: (a) Trench age population, based on sampling global subduction zones at 60 km increments. Dark grey shows the raw population distribution density.The green histogram shows the population distribution weighted by the product of the convergence rate and the depth to the 600°C isotherm. This weighting provides a simple proxy for expected moment release (darker green regions are where the histograms overlap). (b) Shows well-constrained centroid depths of outer rise earthquakes, plotted against the age of their host lithosphere (after Craig et al. (2014a)), overlain on thermal model of M^{c}Kenzie et al. (2005). (c) Shows seismological data from regions with incoming lithosphere in the 80 - 120 Myr age range considered. Centroid data are from Craig et al. (2014a). Finite-fault extents are from (Lay et al., 2010) for the 2009 Tonga/Samoa event; Todd & Lay (2013a) for the 2011 Kermadec doublet; Lay et al. (2010) for the 2006-2007 Kurils Islands doublet; (Ye et al., 2021) for the 2020 Paramushir earthquake; Lay et al. (2013) for the 2012 Honshu doublet; and Ye et al. (2012) for the 2012 Philippines earthquake. Microseismicity data are from (in order from left to right) Hino et al. (2009), Gamage et al. (2009), Obana et al. (2012), (Obana et al., 2014), and Obana et al. (2018) for Honshu, and Zhu et al. (2019) and (Chen et al., 2022) for the Marianas (shallow and deeper regions, respectively).

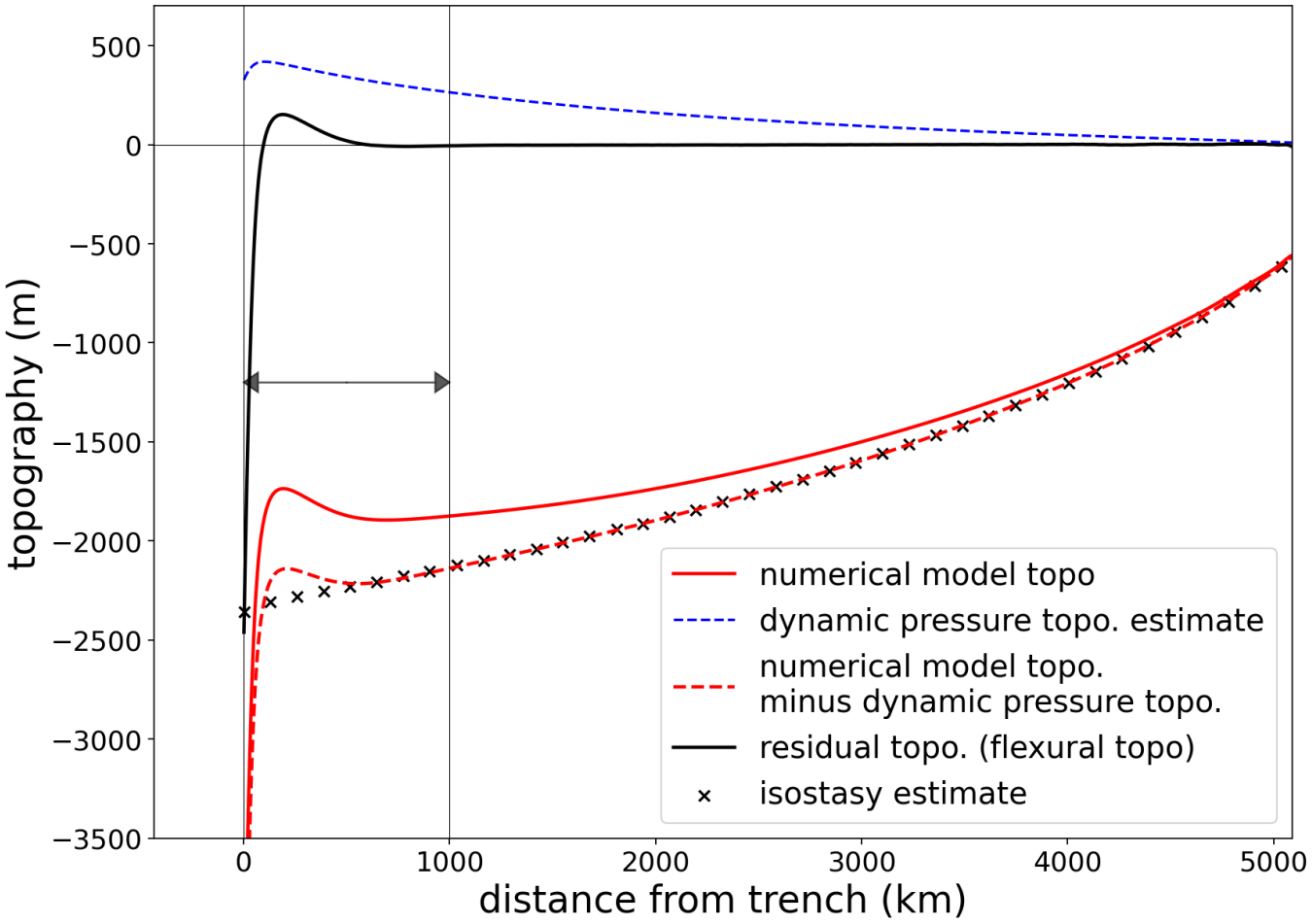

Figure 4: Topography of the subducting plate from the numerical model, same timestep as shown in Fig. 5. Analysis steps are described in detail in Supplementary Section S3. Solid red line shows the raw model topography. There is about 450 m of topography attributable to variations in the 'dynamic' pressure in the asthenosphere (shown with blue dashed line), leading to a flattening of the plate towards the trench (e.g., Holt, 2022). When this dynamic contribution is removed, the topography (dashed red line) is close to the estimated isostatic level based on the density structure of the cooling plate (black points). The solid black line shows model topography corrected for isostatic and dynamic topography, leaving only the flexural topography. The arrow shows the approximate region over which the flexural topography contributes to the observed moment.

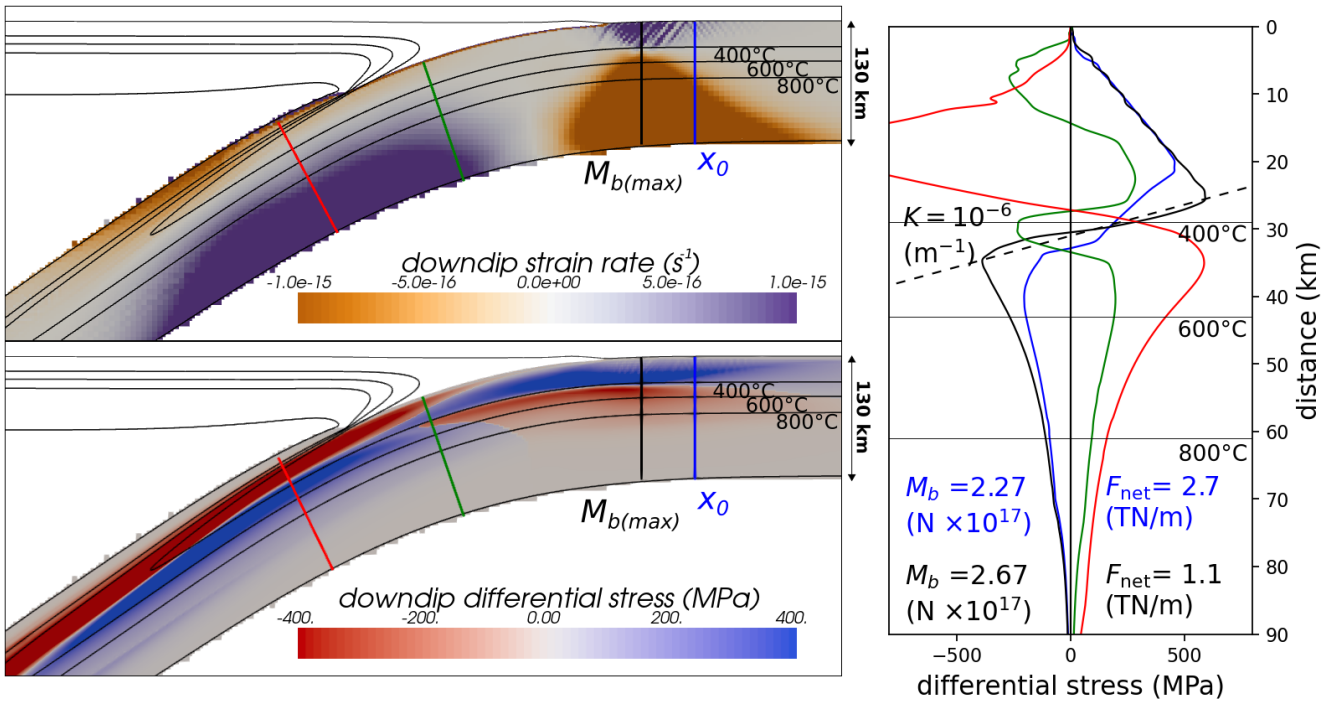

Figure 5: Downdip strain rates and downdip stress from the numerical model, focusing on features within the plate/slab. The fields show e.g. shortening/extension in the downdip direction (rather than along a Cartesian axis). These are based on a rotation of Cartesian tensor components, e.g. downdip stain rate $= \dot{\epsilon}_{ss} = (\dot{\epsilon}_{i,j} \cdot \hat{v}_j^{'}) \cdot \hat{v}_i^{'}$, where $\hat{v}^{'}$ is the unit vector in the direction of the velocity field in the upper plate reference frame (which is sufficient for approximating the downdip direction). The downdip differential stress is equal to $\sigma_1 - \sigma_3$, and the sign is determined so that negative regions are where the most-compressive eigenvector is sub-parallel to the slab dip direction. Stress profiles at four locations are shown. The blue line labelled x_0 is the first zero crossing based on analysis of the flexural component of the topography (see Fig. 4). The black line is the location of maximum bending moment. The green line shows location of partial slab unbending, note that the stress profile has a zig-zag character due to elastic effects (cf. Engdahl & Scholz (1977)). The red line location where the shape of stress has fully inverted, compared with bending. Note that the relative depth the neutral plane is significantly shallower than in the outer rise. The annotations on the right hand panel refer to the bending moment (M_b), curvature (K), and net axial force (F_{net}), with colors corresponding to the profiles.

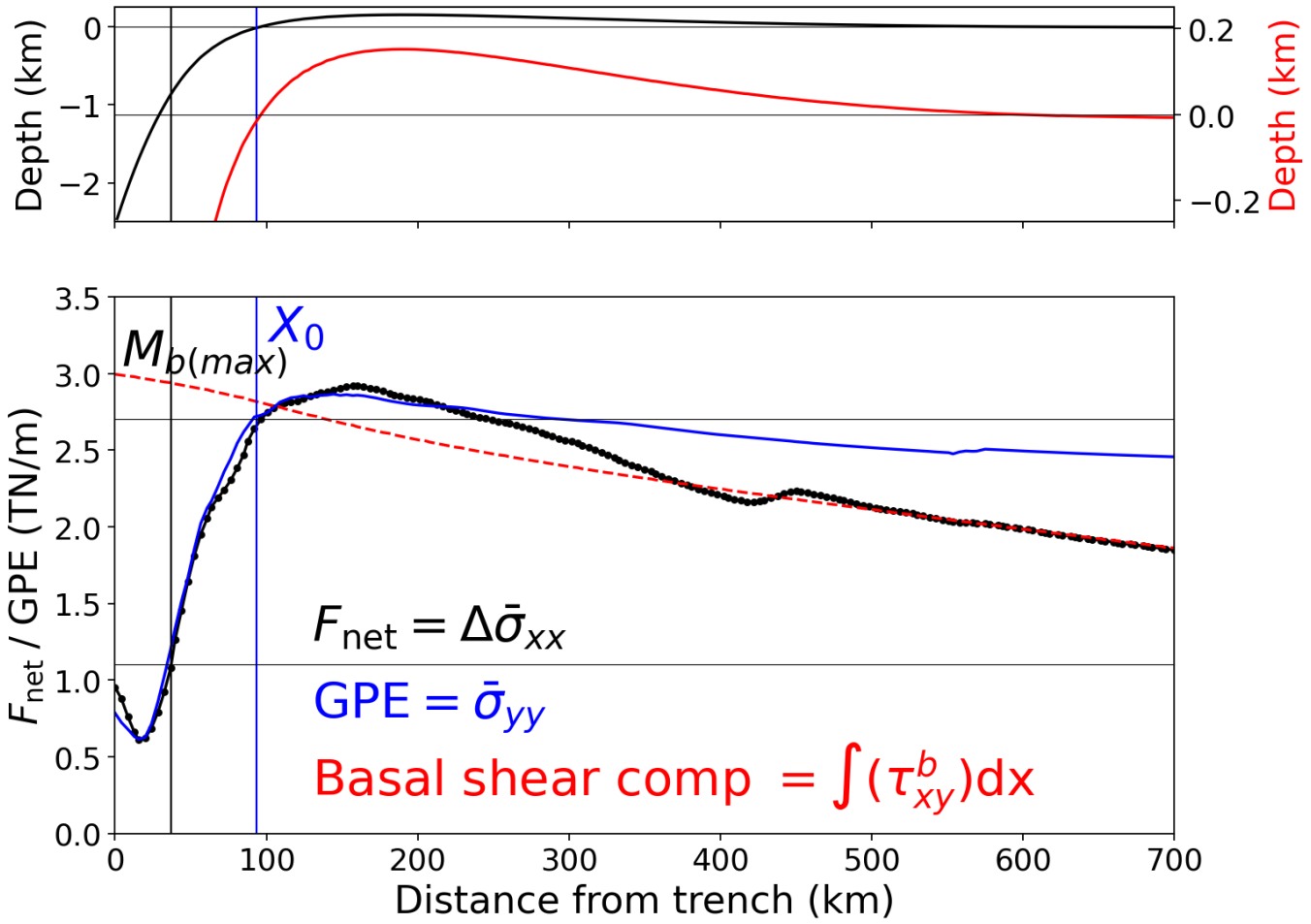

Figure 6: Spatial variation in integrated lithospheric stress, from the numerical model (same timestep as shown in Fig. 5). In the bottom panel the black line shows the net axial force F_{net}, given by Eq. 6. Positive values of F_{net} indicate a net effective tension. The blue line shows the estimated GPE, given by Eq. 6. The dashed red line shows the horizontal integral of the basal shear stress, which should equal F_{net} in the absence of significant GPE gradients (an arbitrary integration constant is chosen so that the shear stress integral is equal to F_{net} at the right hand limit of the domain). The top panel shows the flexural topography at two different vertical scales.

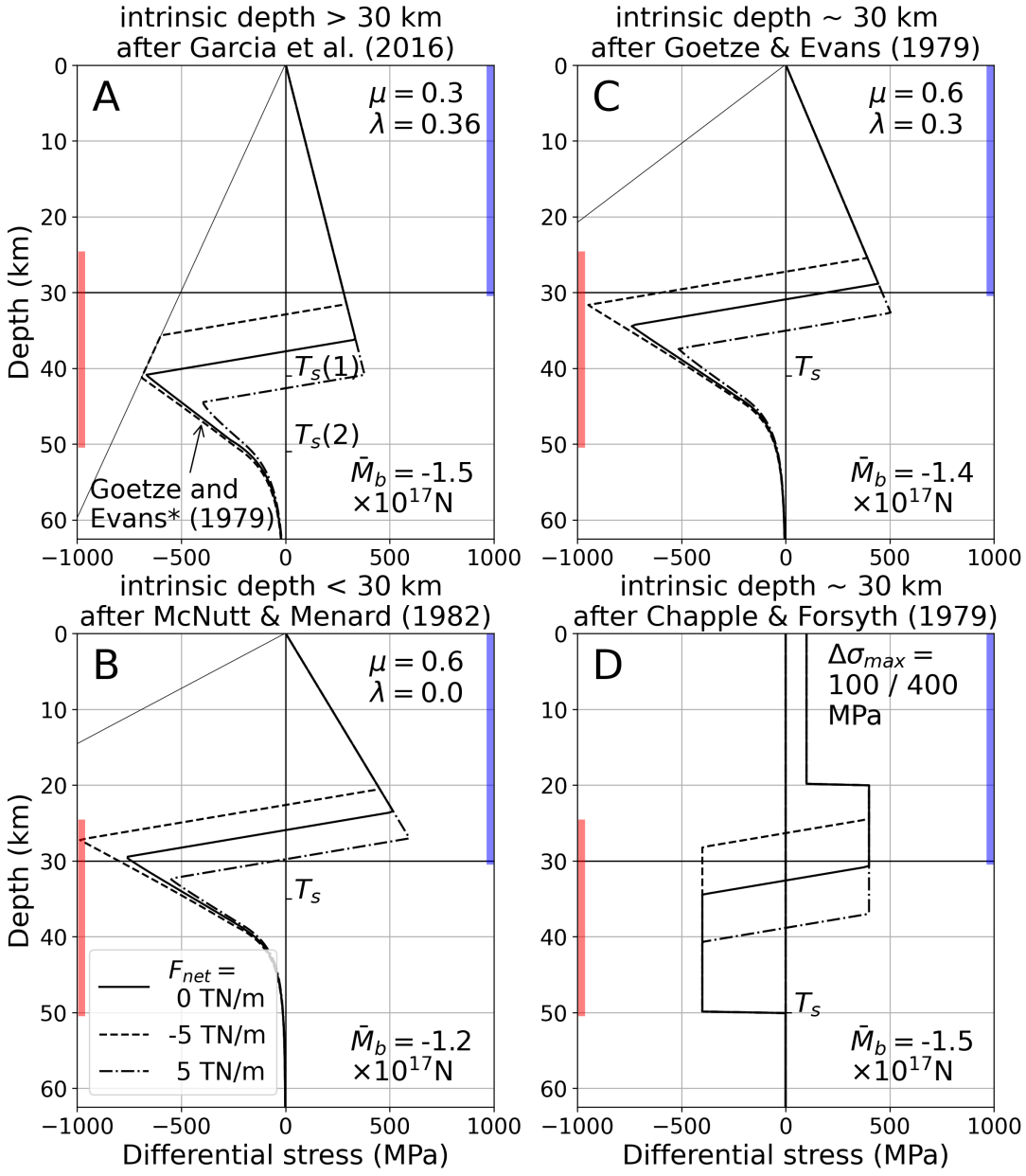

Figure 7: Variants of 'intermediate' strength models, based on previous studies. Each panel shows the YSE subject to neutral background stress (solid line), as well as under the assumption of ± 5 TN/M net axial force (dashed lines as shown in legend panel B). Models shown in panels A, B, & C represent variations on the classical strength model, differing in the relative strength in the brittle and ductile regime. Each of these models utilise the LTP/HTP parameters shown as Set 1 in Table 1. The ductile strength is varied by multiplying each of the activation energies by a factor F. Model A ($F = 0.97$), shows the preferred model from Garcia et al. (2019), characterised by low strength in the brittle (frictional) region, and relatively high strength in the ductile region. The value of F is required to match the original strength model, under our strain rate assumptions (see section 5.1). Model B ($F = 0.82$) is similar to the model proposed by McNutt & Menard (1982), characterised by low strength in the ductile regime, and relatively high strength in the brittle (frictional) region. Model C ($F = 0.89$) is similar to the strength model proposed in Goetze & Evans (1979). Model D is a variation on the *ad hoc* yield strength model proposed in Chapple & Forsyth (1979). In each panel the average bending moment (for the three YSEs) is shown with $\bar{M}_n$. T_s represents a proxy for the expected maximum depth of seismicity (see main text for details). The blue and red bars represent the limits of normal and reverse seismicity inferred from the distribution of well-constrained centroid depths.

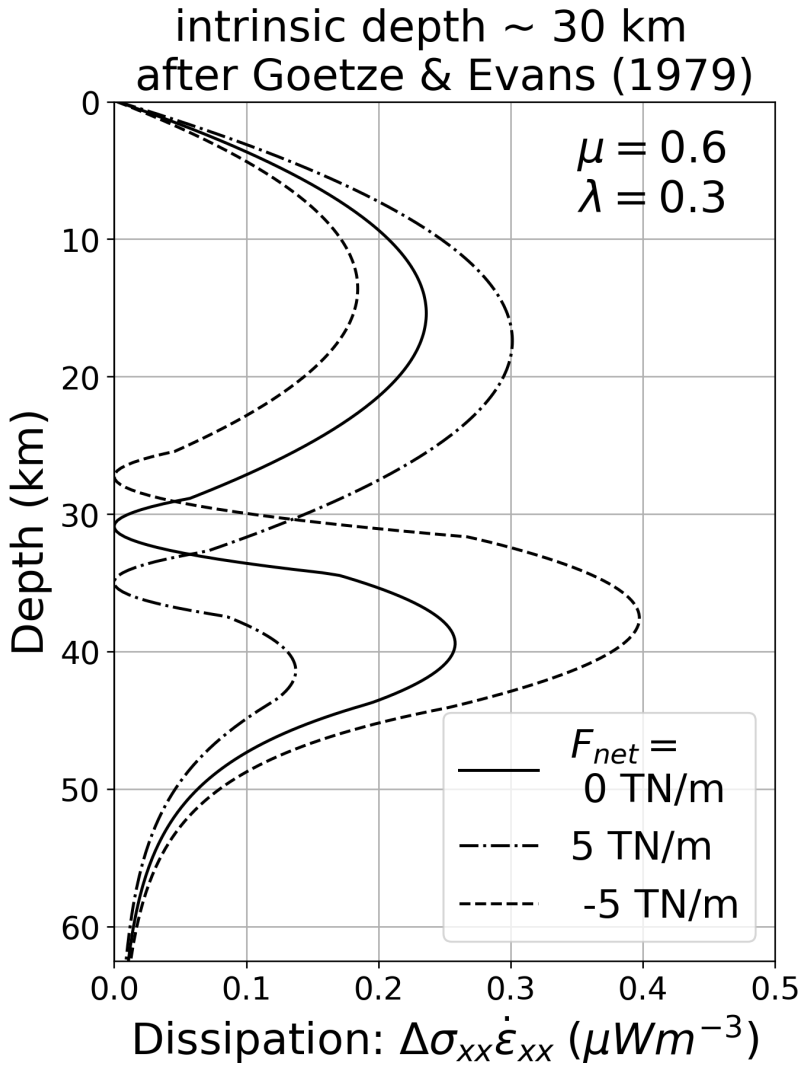

Figure 8: Analastic dissipation corresponding to the YSE shown in Fig. 7C, given by the product of the differential stress and the axial strain rate ($\Delta\sigma_{xx}\dot{\epsilon}_{xx}$). The strain rate is assumed to vary with distance from the neutral plane, based on a bending rate model (e.g. Supplementary Section S4), and consistent with results of the numerical model described in Section 6. The three lines show the same variations in the net axial force (F_{net} = -5, 0, 5 TN/m) as are shown in Fig. 7C, and labelled in the Figure legend. The magnitude of the dissipation beneath the neutral plane is particularly sensitive to changes in neutral plane depth produced by variation in the net axial force.

References

Aharonov, E. & Scholz, C., 2019. The Brittle-Ductile Transition Predicted by a Physics-Based Friction Law, *Journal of Geophysical Research*, **124**, 2721–2737.

Bangerth, W., Dannberg, J., Gassmoeller, R., & Heister, T., 2020. Aspect v2.2.0.

Becker, T. W. & O'Connell, R. J., 2001. Predicting plate velocities with mantle circulation models, *Geochemistry, Geophysics, Geosystems*, **2**(12).

Bellas, A. & Zhong, S., 2021. Seismic strain rate and flexure at the hawaiian islands constrain the frictional coefficient, *Geochemistry, Geophysics, Geosystems*, **22**(4), e2020GC009547.

Bellas, A., Zhong, S., & Watts, A. B., 2022. Reconciling lithospheric rheology between laboratory experiments, field observations and different tectonic settings, *Geophysical Journal International*, **228**(2), 857–875.

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

Bird, P., Liu, Z., & Rucker, W. K., 2008. Stresses that drive the plates from below: Definitions, computational path, model optimization, and error analysis, *Journal of Geophysical Research: Solid Earth*, **113**(B11).

Buffett, B. & Becker, T., 2012. Bending stress and dissipation in subducted lithosphere, *Journal of Geophysical Research: Solid Earth*, **117**(B5).

Byerlee, J., 1978. Friction of rocks, in *Rock friction and earthquake prediction*, pp. 615–626, Springer.

Capitanio, F. A., Morra, G., & Goes, S., 2009. Dynamics of plate bending at the trench and slab-plate coupling, *Geochemistry, Geophysics, Geosystems*, **10**(4).

Chapple, W. M. & Forsyth, D. W., 1979. Earthquakes and bending of plates at trenches, *Journal of Geophysical Research: Solid Earth*, **84**(B12), 6729–6749.

Chen, H., Yang, H., Zhu, G., Lin, J., & You, Q., 2022. Deep Outer-Rise Faults in the Southern Mariana Subduction Zone Indicated by a Machine-Learning-Based High-Resolution Earthquake Catalogue, *Geophysical Research Letters*, **49**.

Coblentz, D., van Wijk, J., Richardson, R. M., & Sandiford, M., 2015. The upper mantle geoid: Implications for continental structure and the intraplate stress field, *Geological Society of America Special Papers*, **514**, 197–214.

Conrad, C. P. & Hager, B. H., 1999. Effects of plate bending and fault strength at subduction zones on plate dynamics, *Journal of Geophysical Research: Solid Earth*, **104**(B8), 17551–17571.

Conrad, C. P. & Lithgow-Bertelloni, C., 2002. How mantle slabs drive plate tectonics, *Science*, **298**(5591), 207–209.

Copley, A., Avouac, J.-P., & Royer, J.-Y., 2010. India-asia collision and the cenozoic slowdown of the indian plate: Implications for the forces driving plate motions, *Journal of Geophysical Research: Solid Earth*, **115**(B3).

Craig, T., Copley, A., & Jackson, J., 2014a. A reassessment of outer-rise seismicity and its implications for the mechanics of oceanic lithosphere, *Geophysical Journal International*, **197**(1), 63–89.

Craig, T. J., Copley, A., & Middleton, T. A., 2014b. Constraining fault friction in oceanic lithosphere using the dip angles of newly-formed faults at outer rises, *Earth and Planetary Science Letters*, **392**, 94–99.

Emmerson, B. & McKenzie, D., 2007. Thermal structure and seismicity of subducting lithosphere, *Physics of the Earth and Planetary Interiors*, **163**(1-4), 191–208.

Engdahl, E. & Scholz, C., 1977. A double benioff zone beneath the central aleutians: An unbending of the lithosphere, *Geophysical Research Letters*, **4**(10), 473–476.

Ewy, R., Stankovich, R., & Bovberg, C., 2003. Mechanical behavior of some clays and shales from 200 m to 3800 m depth, in *39th US Rock Mechanics Symposium/12th Panamerican Conference Soil Mech & Geotech Eng. MIT, Cambridge, USA*, pp. 22–26.

Faccenna, C., Becker, T. W., Lallemand, S., & Steinberger, B., 2012. On the role of slab pull in the cenozoic motion of the pacific plate, *Geophysical Research Letters*, **39**(3).

Fleitout, L. & Froidevaux, C., 1983. Tectonic stresses in the lithosphere, *Tectonics*, **2**(3), 315–324.

Fraters, M., Thieulot, C., Van Den Berg, A., & Spakman, W., 2019. The geodynamic world builder: a solution for complex initial conditions in numerical modeling, *Solid Earth*, **10**(5), 1785–1807.

Gamage, S. S. N., Umino, N., Hasegawa, A., & Kirby, S. H., 2009. Offshore double-planed shallow seismic zone in the NE Japan forearc region revealed by sP depth phases recorded by regional networks, *Geophysical Journal International*, **178**, 195–214.

Garcia, E. S. M., Sandwell, D. T., & Bassett, D., 2019. Outer trench slope flexure and faulting at pacific basin subduction zones, *Geophysical Journal International*, **218**(1), 708–728.

Ghosh, A. & Holt, W. E., 2012. Plate motions and stresses from global dynamic models, *Science*, **335**(6070), 838–843.

Ghosh, A., Holt, W. E., & Flesch, L. M., 2009. Contribution of gravitational potential energy differences to the global stress field, *Geophysical Journal International*, **179**(2), 787–812.

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

Gurnis, M., Ritsema, J., Heijst, H.-J. V., & Zhong, S., 2000. Tonga slab deformation: The influence of a lower mantle upwelling on a slab in a young subduction zone, *Geophysical Research Letters*, **27**(16), 2373–2376.

Hanks, T. C., 1971. The kuril trench-hokkaido rise system: large shallow earthquakes and simple models of deformation, *Geophysical Journal International*, **23**(2), 173–189.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., & Oliphant, T. E., 2020. Array programming with NumPy, *Nature*, **585**(7825), 357–362.

Hino, R., Azuma, R., Ito, Y., Yamamoto, Y., Suzuki, K., Tsushima, H., Suzuki, S., Miyashita, M., Tomori, T., Arizono, M., & Tange, G., 2009. Insight into complex rupturing of the immature bending normal fault in the outer slope of the Japan Trench from aftershocks of the 2005 Sanriku earthquake ($M_W = 7.0$) located by ocean bottom seismometry, *Geochemistry, Geophysics, Geosystems*, **10**.

Hirth, G. & Kohlstedt, D., 2003. Rheology of the upper mantle and the mantle wedge: A view from the experimentalists, *Geophysical monograph-american geophysical union*, **138**, 83–106.

Holt, A. F., 2022. The topographic signature of mantle pressure build-up beneath subducting plates: Insights from spherical subduction models, *Geophysical research letters*, **49**(22), e2022GL100330.

Hunter, J. & Watts, A., 2016. Gravity anomalies, flexure and mantle rheology seaward of circum-pacific trenches, *Geophysical Journal International*, **207**(1), 288–316.

Hunter, J. D., 2007. Matplotlib: A 2d graphics environment, *Computing in Science & Engineering*, **9**(3), 90–95.

Husson, L., 2012. Trench migration and upper plate strain over a convecting mantle, *Physics of the Earth and Planetary Interiors*, **212**, 32–43.

Isacks, B. & Molnar, P., 1971. Distribution of stresses in the descending lithosphere from a global survey of focal-mechanism solutions of mantle earthquakes, *Reviews of Geophysics*, **9**(1), 103–174.

Kubota, T., Hino, R., Inazu, D., & Suzuki, S., 2019. Fault model of the 2012 doublet earthquake, near the up-dip end of the 2011 tohoku-oki earthquake, based on a near-field tsunami: implications for intraplate stress state, *Progress in Earth and Planetary Science*, **6**(1), 1–20.

Lay, T., Ammon, C. J., Kanamori, H., Rivera, L., Koper, K. D., & Hutko, A. R., 2010. The 2009 Samoa-Tonga great earthquake triggered doublet, *Nature*, **466**.

Lay, T., Duputel, Z., Ye, L., & Kanamori, H., 2013. The December 7, 2012 Japan Trench intraplate doublet (M_W 7.2, 7.1) and interactions between near-trench intraplate thrust and normal faulting, *Physics of the Earth and Planetary Interiors*, **220**, 73–78.

Leng, W. & Zhong, S., 2010. Constraints on viscous dissipation of plate bending from compressible mantle convection, *Earth and Planetary Science Letters*, **297**(1-2), 154–164.

Levitt, D. A. & Sandwell, D. T., 1995. Lithospheric bending at subduction zones based on depth soundings and satellite gravity, *Journal of Geophysical Research: Solid Earth*, **100**(B1), 379–400.

Liu, S., Forte, A., & Rowley, D., 2021. Plate motion driving forces revisited, in *AGU Fall Meeting Abstracts*, vol. 2021, pp. DI45D–0060.

Lorinczi, P. & Houseman, G., 2009. Lithospheric gravitational instability beneath the southeast carpathians, *Tectonophysics*, **474**(1-2), 322–336.

M^{c}Kenzie, D., Jackson, J., & Priestley, K., 2005. Thermal structure of oceanic and continental lithosphere, *Earth and Planetary Science Letters*, **233**, 337–349.

McNutt, M. K. & Menard, H., 1982. Constraints on yield strength in the oceanic lithosphere derived from observations of flexure, *Geophysical Journal International*, **71**(2), 363–394.

Mei, S., Suzuki, A., Kohlstedt, D., Dixon, N., & Durham, W., 2010. Experimental constraints on the strength of the lithospheric mantle, *Journal of Gmersoneophysical Research: Solid Earth*, **115**(B8).

Molnar, P., 2020. The brittle-plastic transition, earthquakes, temperatures, and strain rates, *Journal of Geophysical Research: Solid Earth*, **125**(7), e2019JB019335.

Mueller, S. & Phillips, R. J., 1995. On the reliability of lithospheric constraints derived from models of outer-rise flexure, *Geophysical Journal International*, **123**(3), 887–902.

Mueller, S., Spence, W., & Choy, G. L., 1996. Inelastic models of lithospheric stress-11. implications for outer-rise seismicity and dynamics, *Geophysical Journal International*, **125**(1), 54–72.

Obana, K., Fujie, G., Takahashi, T., Yamamoto, Y., Nakamura, Y., Kodaira, S., Takahashi, N., Kaneda, Y., & Shinohara, M., 2012. Normal-faulting earthquakes beneath the outer slope of the Japan Trench after the 2011 Tohoku earthquake: Implications for the stress regime in the incoming Pacific plate, *Geophysical Research Letters*, **39**.

Obana, K., Kodaira, S., Nakamura, Y., Sato, T., Fujie, G., Takahashi, T., & Yamamoto, Y., 2014. Aftershocks of the december 7, 2012 intraplate doublet near the japan trench axis, *Earth, Planets and Space*, **66**(1), 1–6.

Obana, K., Makamura, Y., Fujie, G., Kodaira, S., Kaiho, Y., Yamamoto, Y., & Miura, S., 2018. Seismicity in the source areas of the 1896 and 1933 Sanriku earthquake and implications for large near-trench earthquake faults, *Geophysical Journal International*, **212**.

Obana, K., Takahashi, T., Yamamoto, Y., Iinuma, T., Nakamura, Y., Fujie, G., Miura, S., & Kodaira, S., 2023. Fault geometry of m6-class normal-faulting earthquakes in the outer trench slope of japan trench from ocean bottom seismograph observations, *Progress in Earth and Planetary Science*, **10**(1), 1–19.

Ohnaka, M., 1992. Earthquake source nucleation: a physical model for short-term precursors, *Tectonophysics*, **211**(1-4), 149–178.

Ohnaka, M., 2013. *The physics of rock failure and earthquakes*, Cambridge University Press.

Parsons, B. & Molnar, P., 1976. The origin of outer topographic rises associated with trenches, *Geophysical Journal International*, **45**(3), 707–712.

Parsons, B. & Sclater, J. G., 1977. An analysis of the variation of ocean floor bathymetry and heat flow with age, *Journal of geophysical research*, **82**(5), 803–827.

Pleus, A., Ito, G., Wessel, P., & Frazer, L. N., 2020. Rheology and thermal structure of the lithosphere beneath the hawaiian ridge inferred from gravity data and models of plate flexure, *Geophysical Journal International*, **222**(1), 207–224.

Ranero, C. R., Phipps Morgan, J., McIntosh, K., & Reichert, C., 2003. Bending-related faulting and mantle serpentinization at the middle america trench, *Nature*, **425**(6956), 367–373.

Renshaw, C. E. & Schulson, E. M., 2004. Plastic faulting: Brittle-like failure under high confinement, *Journal of Geophysical Research: Solid Earth*, **109**(B9).

Richards, F., Hoggard, M., Cowton, L., & White, N., 2018. Reassessing the thermal structure of oceanic lithosphere with revised global inventories of basement depths and heat flow measurements, *Journal of Geophysical Research: Solid Earth*, **123**(10), 9136–9161.

Richards, F., Hoggard, M., Crosby, A., Ghelichkhan, S., & White, N., 2020. Structure and dynamics of the oceanic lithosphere-asthenosphere system, *Physics of the Earth and Planetary Interiors*, **309**, 106559.

Sandiford, D., Moresi, L., Sandiford, M., Farrington, R., & Yang, T., 2020. The fingerprints of flexure in slab seismicity, *Tectonics*, **39**(8), e2019TC005894.

Sandiford, M., Coblentz, D., & Schellart, W. P., 2005. Evaluating slab-plate coupling in the indo-australian plate, *Geology*, **33**(2), 113–116.

Schmalholz, S. M., Medvedev, S., Lechmann, S. M., & Podladchikov, Y., 2014. Relationship between tectonic overpressure, deviatoric stress, driving force, isostasy and gravitational potential energy, *Geophysical Journal International*, **197**(2), 680–696.

Schubert, G., Yuen, D. A., Froidevaux, C., Fleitout, L., & Souriau, M., 1978. Mantle circulation with partial shallow return flow: Effects on stresses in oceanic plates and topography of the sea floor, *Journal of Geophysical Research: Solid Earth*, **83**(B2), 745–758.

Shimada, M., Cho, A., & Yukutake, H., 1983. Fracture strength of dry silicate rocks at high confining pressures and activity of acoustic emission, *Tectonophysics*, **96**(1-2), 159–172.

Stauder, W., 1968. Tensional character of earthquake foci beneath the aleutian trench with relation to sea-floor spreading, *Journal of Geophysical Research*, **73**(24), 7693–7701.

Stein, S. & Pelayo, A., 1991. Seismological constraints on stress in the oceanic lithosphere, *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, **337**(1645), 53–72.

Stesky, R., Brace, W., Riley, D., & Robin, P.-Y., 1974. Friction in faulted rock at high temperature and pressure, *Tectonophysics*, **23**(1-2), 177–203.

Stotz, I., Iaffaldano, G., & Davies, D. R., 2018. Pressure-driven poiseuille flow: a major component of the torque-balance governing pacific plate motion, *Geophysical Research Letters*, **45**(1), 117–125.

Sullivan, C. B. & Kaszynski, A., 2019. PyVista: 3d plotting and mesh analysis through a streamlined interface for the visualization toolkit (VTK), *Journal of Open Source Software*, **4**(37), 1450.

Todd, E. K. & Lay, T., 2013a. The 2011 Northern Kermadec earthquake doublet and subduction zone faulting interactions, *Journal of Geophysical Research*, **118**, 1–13.

Todd, E. K. & Lay, T., 2013b. The 2011 northern kermadec earthquake doublet and subduction zone faulting interactions, *Journal of Geophysical Research: Solid Earth*, **118**(1), 249–261.

Toffol, G., Yang, J., Pennacchioni, G., Faccenda, M., & Scambelluri, M., 2022. How to quake a subducting dry slab at intermediate depths: Inferences from numerical modelling, *Earth and Planetary Science Letters*, **578**, 117289.

Townend, J. & Zoback, M. D., 2000. How faulting keeps the crust strong, *Geology*, **28**(5), 399–402.

Turcotte, D. L., McAdoo, D., & Caldwell, J., 1978. An elastic-perfectly plastic analysis of the bending of the lithosphere at a trench, *Tectonophysics*, **47**(3-4), 193–205.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., & SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods*, **17**, 261–272.

Ye, L., Lay, T., & Kanamori, H., 2012. Intraplate and interplate faulting interactions during the August 31, 2012, Philippine Trench earthquake (M_W 7.6) sequence, *Geophysical Research Letters*, **39**.

Ye, L., Lay, T., & Kanamori, H., 2021. The 25 march 2020 mw 7.5 paramushir, northern kuril islands earthquake and major (mw $\geq$ 7.0) near-trench intraplate compressional faulting, *Earth and Planetary Science Letters*, **556**, 116728.

Zhao, D., Katayama, Y., & Toyokuni, G., 2022. The moho, slab and tomography of the east japan forearc derived from seafloor s-net data, *Tectonophysics*, **837**, 229452.

Zhong, S. & Watts, A., 2013. Lithospheric deformation induced by loading of the hawaiian islands and its implications for mantle rheology, *Journal of Geophysical Research: Solid Earth*, **118**(11), 6025–6048.

Zhu, G., Yang, H., Lin, J., Zhou, Z., Xu, M., Sun, J., & Wan, K., 2019. Along-strike variation in slab geometry at the southern Mariana subduction zone revealed by seismicity through ocean bottom seismic experiments, *Geophysical Journal International*, **218**.

Zoback, M. D. & Townend, J., 2001. Implications of hydrostatic pore pressures and high crustal strength for the deformation of intraplate lithosphere, *Tectonophysics*, **336**(1-4), 19–30.

Zoback, M. D., Townend, J., & Grollimund, B., 2002. Steady-state failure equilibrium and deformation of intraplate lithosphere, *International Geology Review*, **44**(5), 383–401.

Supplementary Information: Plate bending earthquakes and the strength distribution of the lithosphere

Dan Sandiford[1*] and Tim Craig[2]

[1] School of Earth, Atmosphere and Environment, Monash University, Melbourne, Victoria, Australia

[2] COMET, Institute of Geophysics and Tectonics, School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK.

[*] e-mail: dan.sandiford@monash.edu

Contents

S1 Mathematical background: thin-plate flexure, topography and bending moment

The 2D thin plate equations are the basis for the 'observed moment' observations addressed in this study, and are reviewed here. The thin plate equation is the expression of static equilibrium (force and torque balance) under the assumptions of small deflection and negligible shear deformation. Following Turcotte & Schubert (2002), q is the load on the plate, with units of force per unit area (N m^{-2}). Note that q is not an external load applied to the plate; it is the load associated with the uncompensated 'flexural' topography. The downward load, per unit length in the z direction, between x and $x + dx$ is: $q(x)dx$. V is the integrated shear force per unit length in the z direction, also called the shear stress resultant (N m^{-1}). The vertical force balance (force per unit length in the z-dir) on a plate element:

$$q(x)dx + dV = 0 \tag{1}$$

$$\frac{dV}{dx} = -q \tag{2}$$

The balance between the bending moment and the torque due to the vertical shear stress is:

$$dM_b = F_{net}dw - Vdx \tag{3}$$

Where F_{net} is the horizontal stress resultant (i.e. the net axial force N m^{-1}). The bending moment over a plate of thickness L is:

$$M_b = \int_0^L \Delta\sigma_{xx}(y - y_n)dy \tag{4}$$

Where $\Delta\sigma_{xx}$ is the differential stress, y_n is the neutral plane depth, where the differential stress is zero. Taking the differential form of Eq. (3) and substituting q for V from (1) gives:

$$\frac{d^2 M_b}{dx^2} = -q + F_{net}\frac{d^2 w}{dx^2} \tag{5}$$

For a linear elastic rheology, the moment is proportional to curvature, approximately the second derivative of the topography, and a solution to Eq. 5 is a damped sinusoid:

$$w(x) = Ae^{-(x-x_0)/\alpha} sin((x-x_0)/\alpha) \tag{6}$$

where the origin of the x coordinate is chosen at the point where the load is assumed to be applied, and x_0 is the first zero crossing. α the flexural parameter, with units of distance, and defines the width of the flexure. The amplitude (A) is related to the height of the fore bulge (w_b) via, $w_b = (Ae^{-\pi/4})/\sqrt{2}$.

Eq. 6 applies to a broken plate, assuming zero moment at the trench, a condition that does not apply in the numerical model. In the analysis of the model, we investigate the use of Eq. 6 as a fitting function, primarily because this mimics the procedure that has been used in previous studies of global subduction zones Levitt & Sandwell (1995); McNutt & Menard (1982).

Integrating equation 5 twice, around an arbitrary point x_m gives:

$$M_b(x_m) = \int_{x_m}^{\infty} q(x-x_m)\ dx + F_{net} w(x_m)$$

or

$$M_b(x_m) = \int_{x_m}^{\infty} \Delta\rho g w(x)(x-x_m)\ dx + F_{net} w(x_m) \tag{7}$$

The integral on the RHS is referred to as the observed moment, the torque due to the uncompensated (flexural) topography, and should balance the bending moment at a point x_m.

In previous studies, the observed moment estimation has generally been performed around $x_m = x_0$, as the final term in Eq. 7, representing the torque due to the net axial force acting on a moment arm of length $w(x_m)$, is zero.

As pointed out by (McNutt & Menard, 1982), the integral (7) becomes extremely unstable at large values of $(x - x_0)$

if there is noise in the bathymetric or gravity profile. Noise is of course any component of the topography unrelated to plate flexure. For this reason, instead of integration of the corrected topography, several studies have proceeded by smoothing the data with the stable analytic solution to a uniform thickness elastic plate (McNutt & Menard, 1982; Levitt & Sandwell, 1995). The parameters in the elastic plate solution were fit by least-squares in McNutt & Menard (1982) and by a more complex non-linear inversion optimisation scheme in Levitt & Sandwell (1995), which also includes the gravity anomaly. When the topography is fit by the elastic plate solution, the observed moment integral in Eq. 7, at x_0, is equal to:

$$M_b(x_0) = A\frac{\Delta\rho g}{2}\alpha^2 \tag{8}$$

Exactly the same result is obtained if one simply computes the bending moment at x_0 (i.e. $M_b = DK$) based on the deflection Eq. 6, with D being the flexural rigidity $D = g\Delta\rho\alpha^4/4$, and K the curvature. Hence, one can interpret the moment estimates in studies such as in (McNutt & Menard, 1982; Levitt & Sandwell, 1995) as either the integrated moment of the topographic load, or the bending moment of the equivalent plate at x_0, based on the optimised value of α. Of course, once an elastic fitting solution is derived, the restriction of estimating the moment at x_0, where $w = 0$ is no longer required. Nevertheless, this approach is the focus of our analysis of the numerical model, because this follows the approach in several of the key data-sets of interest (e.g., McNutt & Menard, 1982; Levitt & Sandwell, 1995). It also follows from the expectation that the bending moment will have almost reached saturation at the first zero-crossing.

For the numerical model, where we have access to the stress within the plate, integration of Eq. 3, gives the bending moment in terms of vertical shear stress resultant:

$$M_b(x_0) = -\int_{x_0}^{\infty} V(x - x_0)\ dx \tag{9}$$

We compute this quantity primarily as it allows testing the validity of the thin plate equation (Eq. 3) without the complexity of topographic correction. As shown later in the study, the leading-order description of the moment balance (Eq. 9) is shown to be very accurate in the case of the numerical model (e.g. solid red line in Fig. S4).

In addition, the gradient of the vertical shear stress resultant can be used to define a pseudo-topography, that is, the topography that would be supported by the vertical shear stress:

$$w_V = \frac{dV}{dx}\left(\frac{1}{g\Delta\rho}\right) \tag{10}$$

It is important to highlight that for all rheological models discussed in this paper, not only does inelastic deformation take place, but also the bending moment is expected to be nearly-saturated in the trench-outer rise region (e.g., Levitt & Sandwell, 1995; Garcia et al., 2019). Nevertheless, as the numerical model results shows, there is still broad similarity between the flexural topography of a plate that has reached moment-saturation and the shape of the solution for a uniform thickness broken elastic plate. A visually 'close' fit to elastic plate solution does not imply that a plate is predominantly elastic (i.e. in retaining a significant non-yielding core during bending). Hence, the fact that some locations (e.g. Kuriles) may be more accurately modelled with a uniform elastic plate equation than others (e.g. Tonga, Turcotte et al. (1978)) does not be indicate that yielding is only relevant in the latter; both regions are predicted to be nearly-moment saturated in the trench-outer-rise region. The observations may well indicate that the plate at Tonga has undergone a greater degree of bending at near-moment saturation (which produces increasing departure from the form of the uniform thickness elastic plate model). The historical developments in this subject can be somewhat difficult to follow, and hence we think it worth elaborating these points.

S2 Numerical model setup, benchmarking, input files

Overview

The numerical subduction model represents the evolution of an idealised 2D convective subduction system, or the one-sided sinking of a thermal boundary layer due to thermal buoyancy. The silicate mantle is represented by an incompressible continua, with temperature and stress-dependent rheology. Buoyancy is determined solely through the temperature dependence of density, following the extended Boussinesq approximation. The flow field in the model represents the instantaneous balance between the buoyancy force and internal stresses, neglecting inertial forces (Stokes equations). Elastic shear deformation is included in the constitutive model, necessitating an additional force term in the Stokes equations (e.g. Schmalholz et al. (2001); Moresi et al. (2003); Bangerth et al. (2020)).

The model has free-slip boundaries on the edges and bottom of the rectangular domain, and a free surface at the top boundary, with a stabilisation scheme that follows (Kaus et al., 2010). The initial conditions comprise an adiabatic mantle with a potential temperature of 1350 °C and two plates, the thermal structure prescribed according the 1D

cooling profile for a half-space. One of these plates is attached to a slab that extends to 660 km depth, and has an age of 100 Myr at the trench. Imposing a slab that reaches the transition zone was found to be a more stable initial configuration in terms of potential instabilities of the free surface. The velocity field that arises from this solution is used to solve the advection-diffusion equation for the temperature field.

The ASPECT code (Bangerth et al., 2020) is used to solve the governing equations, particularly because the adaptive mesh refinement methods are suited to resolving stress in the strong part of the slab. In the numerical model, visco-plastic-elastic deformation mechanisms are implemented through an effective viscosity, as described in the Supplementary Information and previous papers (e.g., Sandiford et al., 2021).

In contrast to models based on a viscous or visco-plastic relationship, the stress in an elastic yielding material is a function of the entire strain path (including initial conditions). For the stress state in the subduction hinge to reach quasi-steady state, material needs to pass through the entire subduction hinge, which dictates running the model for approximately 5 million years.

Domain, initial conditions, boundary conditions, mesh refinement

The subduction model comprises a rectangular domain with a depth of 2900 km, and an aspect ratio of 4, as shown in Fig. S1.

The initial conditions comprise an adiabatic mantle with a potential temperature of 1350 °C and two plates, whose age and thermal structure follows the cooling 1d cooling profile for a half-space (infinite in the depth direction). One of these plates is attached to a slab that extends to 660 km depth, and has an age of 100 Myr at the trench. The upper plate is modelled with a younger thermal age, 25 Myr at the trench. Imposing an initial slab that reaches the transition zone was found to be a more stable initial configuration in terms of potential instabilities of the free surface.

Fig. S1 figure shows the geometric evolution of the slab through 0-10 Myr. During which time the trench retreats back and the slab tip advances. The subduction hinge undergoes little change in curvature.

The inset panels of Fig. S1 indicate the use of adaptive mesh refinement. 7 levels of mesh refinement were used, with the largest (Q2) elements having an edge length of 45 km, and the smallest elements have an edge length of $\sim$ 700 m.

One of the primary refinement criteria is temperature, this allows us to concentrate high mesh resolution in the colder

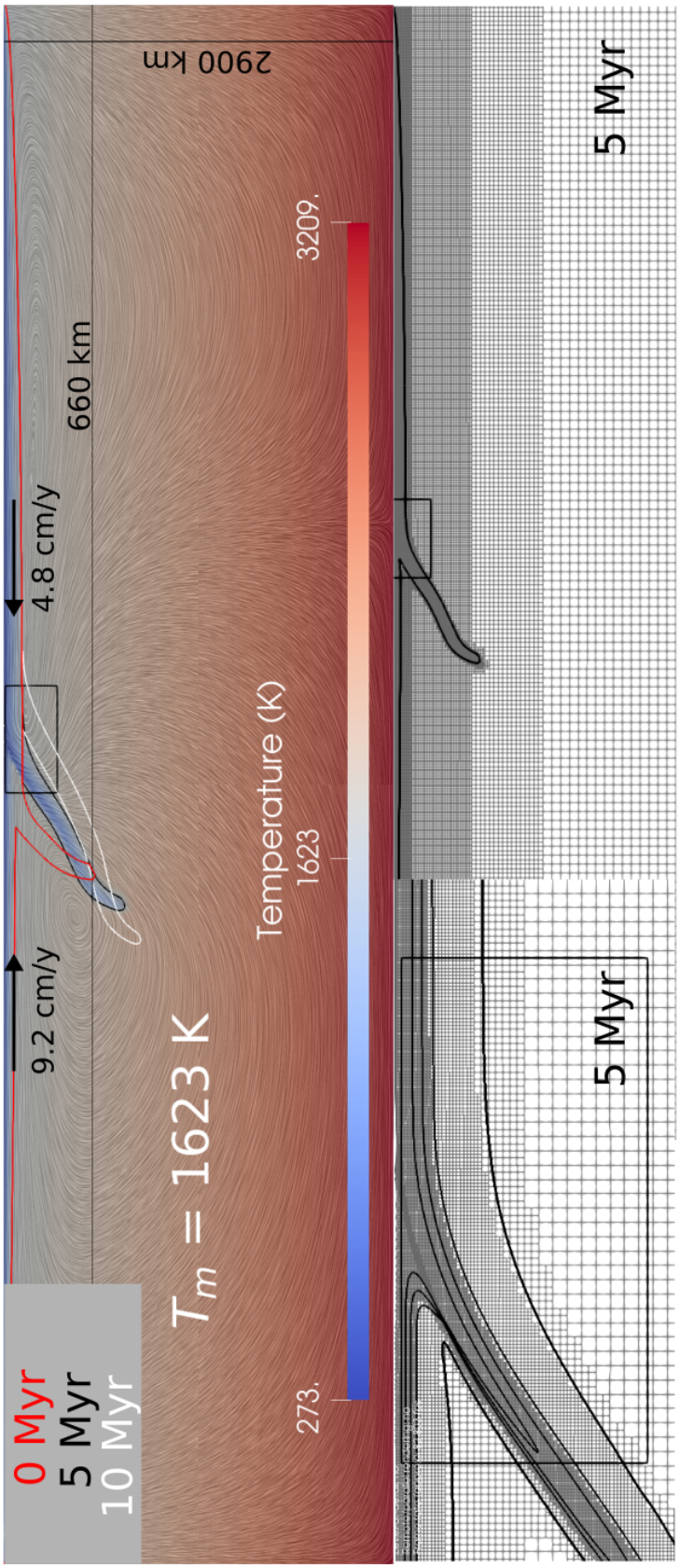

Figure S1: Numerical model domain. Model height is 2900 km, and width is 11600 km (aspect ratio of 4). Coloured contours show shape of slab at 0, 5, 10 Myr. Scalar field is temperature at 5 Myr. Convergence velocity at 5 Myr is shown with arrows, and is dominated by rollback. The texture in the main panel is a line integral convolution showing the direction of flow in the mantle. Inset panels show the adaptive mesh refinement at 5 Myr.

upper half of lithosphere, where the development of large bending stresses occurs, and where pseudo-brittle deformation occurs in localised shear bands.

The highest refinement level is used to capture the subduction interface. The interface is modelled through an entrained weak layer approach (Sandiford & Moresi, 2019). A thin layer (here 2 km thick) represented by a separate composition is imposed on the top of the subducting plate, as well as between the subducting and upper plate. This composition has a low coefficient of friction, providing a shear stress that varies between between 10 - 20 MPa throughout the interface domain. A maximum decoupling depth emerges at about 80 km, consistent with inferences on the typical depth in mature subduction settings (Wada & Wang, 2009).

Governing equations, solution and approximations

For the numerical simulations shown in this paper, ASPECT solves the 2D incompressible conservation equations assuming an infinite Prandtl number and the Extended-Boussinesq approximations described in Bangerth et al. (2020b). A successive substitution (Picard) approach is used to resolve the non-linearity in the constitutive model.

The implementation of Maxwell visco-elasticity involves a stress rate, and is discretized using a backward finite-difference scheme, over timestep Δt,. The deviatoric stress at timestep t can be written as:

$$\tau_{ij}^{t} = \eta_{\text{eff}} \left(2\dot{\epsilon_{ij}}^{t} + \frac{1}{G\Delta t}\tilde{\tau}_{ij} \right) \tag{11}$$

Where $\tilde{\tau}_{ij}$ is the (shear) stress history tensor advected and rotated into the configuration of the current timestep, G is the elastic shear modulus.

The momentum equation for visco-elasticity includes an effective viscosity, while the stress history tensor constitutes an additional force term:

$$(2\eta_{\text{eff}}\dot{\epsilon_{ij}})_{,j} - P_{,i} = \rho g_i - \frac{\eta_{\text{eff}}}{G\Delta t}\tilde{\tau}_{ij,j} \tag{12}$$

where η_{eff} is the effective visco-elasto-plastic viscosity, $\dot{\epsilon}$ is the strain rate, P is the total pressure, $\rho = \bar{\rho}\left(1 - \alpha\left(T - T_{\text{ref}}\right)\right)$ with α the thermal expansivity, g the gravity vector, which is constant across the domain. The final term on the RHS

represents the contribution of the accumulated elastic stress.

The derivation and implementation in ASPECT, are discussed in more detail in the Supplementary Information of (Sandiford et al., 2021).

Constitutive model and rheological parameters

Which stress, which strain rate?

Creep formulations in ASPECT (and many other geodynamics codes) express relationships between effective stress and effective strain rates, i.e. the square root of the second invariant of the deviatoric tensors: $\tau_{\text{eff}} = \sqrt{\text{J2}} = \sqrt{\left(\frac{1}{2}\tau_{ij} : \tau_{ij}\right)}$.

Note that the effective stress is simply a scalar representation related to the magnitude of deviatoric stress tensor (e.g., Ranalli, 1995). The effective viscosity that appears in the Stokes Equation (Eq. 12) refers to a viscosity used to capture a range of deformation mechanisms.

In contrast, YSEs are typically formulated in terms of a relationship between differential stress and the axial (or horizontal) components of the strain rate. Indeed, the parameters reported in uniaxial rock mechanics experiments (e.g. Hirth & Kohlstedt (2003); Mei et al. (2010)) are generally used directly in YSEs, where the axial strain rate is prescribed, and the differential stress is determined from the flow law (Goetze & Evans, 1979; McNutt & Menard, 1982; Garcia et al., 2019).

Consider the basic power-law stress-strain rate relationship as written in the main manuscript (with the Arrhenius term omitted):

$$\dot{\varepsilon} = A_{\text{disl}}\sigma^n \tag{13}$$

This follows the form of the olivine dislocation flow law as written Hirth & Kohlstedt (2003) (and Mei et al. (2010) for LTP). The strain rate on the LHS refers to the axial (or engineering) strain rate. The stress on the right hand side is the differntial stress.

For 2D extension/shortening (pure shear), the horizontal component of the strain rate has equal magnitude to the

effective strain rate while the deviatoric stress is twice the magnitude of the effective stress (the latter being equal to the maximum shear stress). Clearly, one cannot simply use the value of A_{disl} directly in a dynamic formulation that is based on effective tensor quantities (i.e. ASPECT).

A simple way to incorporate the standard laboratory derived (based differential stress) flow law parameters, is to halve the viscosity values that are ordinarily used to describe the relationship between effective stress and effective strain rate. Of course, it is also possible to incorporate a correction factor into the pre-exponential term, as in Ranalli (1995) (who derived the appropriate conversions for a 3D setting). This is trivial for linear or power-law creep, but is not straightforward in the case of the viscosity approximation to Peierls creep that we utilise here (as described below).

As such, for this study we compiled a version of ASPECT in which an (additional) factor of a half was applied to each of the effective viscosity relationships (diffusion, dislocation and Peierls creep). For clarity, these terms are shown in red in the viscosity equations written below. Ultimately, all this means is that for a given set of flow law parameters, the YSE model (following the standard approach) and the numerical model, will both exhibit the same relationship between the strain rate (e.g. the horizontal strain rate in the outer-rise) and the differential stress.

Intra-crystalline creep

To compute the effective viscosity η_{eff} in Eq. 12, we first consider intra-crystalline creep occurring through a composite of diffusion (linear), dislocation (power law) and exponential (Peierls/LTP) mechanisms, resulting in a creep viscosity η_v:

The effective viscosity associated with high temperature dislocation creep is modelled with a dry olivine flow law (Hirth & Kohlstedt, 2003), although the pre-exponential term is modified to result in a slightly stronger stress-strain rate relationship, which was required to achieve realistic subduction velocities (see Table S1):

$$\eta_{\text{disl}} = \left(\frac{1}{2}\right) \frac{1}{2} A_{\text{disl}}^{-\frac{1}{n}} \dot{\epsilon}_{\text{eff}}^{\frac{1-n}{n}} \exp\left(\frac{E_{\text{disl}} + P\,V_{\text{disl}}}{nRT}\right) \tag{14}$$

The effective strain rate $\dot{\epsilon}_{\text{eff}}$ is defined as $\sqrt{\dot{\epsilon}'_{\text{II}}}$, with $\dot{\epsilon}'$ the deviatoric strain rate (based on the velocity of the previous timestep or nonlinear iteration) and $\dot{\epsilon}'_{\text{II}}$ indicating the second invariant of the deviatoric strain rate tensor. T is temperature (K), R is the gas constant, n is the dislocation creep exponent.

The diffusion creep parameters are designed to provide a 'background' viscosity consistent with global geodynamic inversions. The parameters are based on the 1D model of (Steinberger & Calderwood, 2006). This model prescribes a viscosity increase at the 660 km boundary. The viscosity associated with diffusion creep is:

$$\eta_{\text{diff}} = \left(\frac{1}{2}\right)\frac{1}{2}A_{\text{diff}}^{-1}d^{m}\exp\left(\frac{E_{\text{diff}}+P\,V_{\text{diff}}}{RT}\right) \tag{15}$$

Low temperature plasticity, or exponential creep is generally modelled via the equation shown the main manuscript:

$$\dot{\epsilon} = A_{\text{p}}\sigma^{n}\exp\left(-\frac{E_{\text{p}}}{RT}\left(1-\left(\frac{\sigma}{\sigma_{p}}\right)^{p}\right)^{q}\right), \tag{16}$$

where σ_p is the Peierls stress p is the first Peierls glide parameter q is the second Peierls creep glide parameter, A_{p} is the Peierls prefactor term, n is the Peierls stress exponent, E_{p} is the Peierls activation energy, V_{p} is the Peierls activation volume.

As this equation cannot generally be inverted to define stress as a function of strain rate, an iterative approach is often required to solve for the stress-strain rate relationship. Alternatively, a more efficient method is to use a viscosity approximation (Kameyama et al., 1999). The formulation for the approximate LTP viscosity (η_{p}) used in ASPECT is:

$$\eta_{\text{ltp}} = \left(\frac{1}{2}\right) \times \text{stress term} \times \text{arrhenius term} \times \text{strain rate term} \tag{17}$$

,

where

$$\text{stress term} = \frac{\gamma\sigma_p}{2(A_p(\gamma\sigma_p)^n)^{\frac{1}{s+n}}}$$
$$\text{arrhenius term} = \exp\left[\frac{E_p + PV_p}{RT}\frac{(1-\gamma^p)^q}{(s+n)}\right]$$
$$\text{strain rate term} = \dot{\epsilon}_{\text{eff}}^{(\frac{1}{s+n}-1)}$$
$$s = \frac{E_p + PV_p}{RT}pq(1-\gamma^p)^{q-1}\gamma^p$$

where γ is the Peierls creep fitting parameter, associated with the viscosity approximation.

The LTP model is based on a the parameters of (Mei et al., 2010). There are two changes made. One is the addition of activation volume (a parameter not included in the original study. To maintain the same stress-strain relationship, within the temperature range of interest, the prefactor is is changed to offset the effect of the volume ($1.4 \times 10^{-10} \Rightarrow 1.4 \times 10^{-24}\,\text{s}^{-1}\text{Pa}^{-2}$). Secondly, the activation energy is reduced, so that the strength distribution of the lithosphere results in a $\sim$ 30 km neutral plane, based on the assumed friction coefficient ($\mu = 0.8$), and the typical net axial force (effective tension of $\sim 1.4 \times 10^{12}$ N/m).

Although the model has three independent creep mechanisms, there are some constraints placed on the domains in which these are allowed to operate. Firstly, only diffusion creep occurs beneath a depth of 660 km. This enables us to ensure that the large scale viscosity structure of the model is consistent a global average viscosity structure based on multiple constraints (Steinberger & Calderwood, 2006).

Secondly, LTP plasticity is restricted to lithosphere temperatures beneath 1000 °C. This constraint was applied for stability, and without this restriction the model produced 'runaway' viscosity reduction in high strain rate, high temperature regions of the upper mantle, i.e. regions that would be assumed to be deforming under high temperature creep mechanism. These restrictions were applied by leveraging the 'phase function' capacity of ASPECT material model; which enables variation of parameters through specified temperature and pressure boundaries.

At temperatures less that 1000 °C and depths less that 660 km, all three creep mechanisms operate. This includes the primary region of interest, the outer rise.

A few further limitations need to be noted. In the version of ASPECT used for this study, an approach was taken where the deformation rate is equal to that of the most efficient creep mechanism, i.e. we do not treat the creep

mechanisms in series; this is in keeping with the traditional YSE approach and allows a more direct comparison. Secondly, the viscosity associated with dislocation and Peierls creep contain effective strain rates. Strictly, these should represent only the part of the strain rate that is caused by the same mechanism. This is not how the code is currently implemented, and the effective strain rates include all contributions, including elastic and plastic strain rates.

$$\eta_{\mathrm{v}} = min(\eta_{\mathrm{diff}}, \eta_{\mathrm{disl}}, \eta_{\mathrm{p}}) \tag{18}$$

The effective viscosity, which models the elastic and viscous strain contributions, is:

$$\eta_{\mathrm{ve}} = \left(\frac{G\Delta t \eta_{\mathrm{v}}}{G\Delta t + \eta_{\mathrm{v}}} \right) \tag{19}$$

Pseudo-brittle behaviour

Brittle behaviour is modelled through a Drucker-Prager yield limit (τ_y) on the magnitude of the deviatoric stress. An effective strain rate is defined as:

$$\dot{\epsilon}_{\mathrm{ve\ eff}} = 2\dot{\epsilon} + \frac{1}{\mu \Delta t_e} \tilde{\tau} \tag{20}$$

The effective deviatoric stress (equal to maximum shear stress) is:

$$\tau_{\mathrm{pred}} = 2\eta_{\mathrm{ve}} \dot{\epsilon}_{\mathrm{ve\ eff}} \tag{21}$$

The yield criterion

$$F = \tau_{\mathrm{pred}} - C\cos(\phi) - p\sin(\phi) \tag{22}$$

is then evaluated, where C and ϕ are the cohesion and friction angle, related to the friction coefficient by $\mu = tan(\phi)$.

When $F >= 0$, plastic yielding occurs and the stress state is mapped back onto the yield surface by rescaling the effective viscosity:

$$\eta_{\text{eff}} = \frac{C\cos(\phi) + p\sin(\phi)}{2\dot{\epsilon}_{\text{ve eff}}}. \tag{23}$$

Otherwise $\eta_{\text{eff}} = \eta_{\text{ve}}$.

Finally, minimum and maximum bounds are applied to the averaged effective viscosity:

$$\eta_{\text{eff}} = \min(\max(\eta_{\text{eff}}, \eta_{\text{min}}), \eta_{\text{max}}) \tag{24}$$

A deformation map for the numerical model is shown the lower panels of Fig. S2. Because the implementation of the visco-elasticity follows the Maxwell model (elastic and viscous elements in series) there is always a viscous component to the deformation, albeit negligible at colder temperatures. The definition of a region behaving elastically in Fig. S2, is when the intra-crystalline mechanisms contribute $< 25\%$ of the total strain rate, and pseudo-brittle deformation is not occurring.

Benchmark for visco-elastic plastic rheology

Fig. S3 shows a benchmark for isoviscous simple shear, adapted from Farrington et al. (2014). The thicker blue line shows the analytical solution (e.g. Farrington et al. (2014) Eq. 9) for the evolution of the shear stress τ_{xy} while the material is loaded (sheared) to $t = 1$, then allowed to relax until $t = 4$. The solid black line shows the results from ASPECT, when the yield stress in the material is high enough to prevent yielding. The dashed black line shows results from ASPECT when the yield stress is equal to 2.5. In this case the model is loaded until the time ($t \sim 1.24$) where the reference model stress equals the yield stress. The stress in the yielding model follows exactly the same viscous relaxation path as the model without yielding.

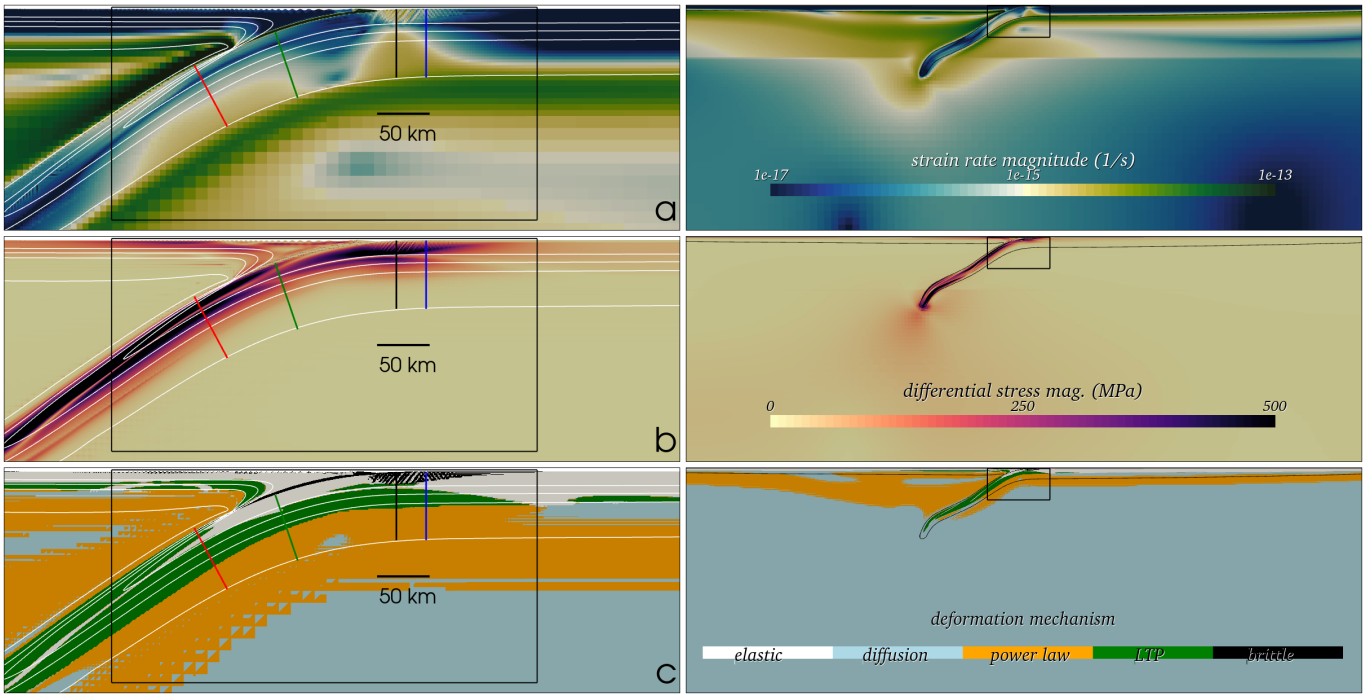

Figure S2: Snapshot of model dynamics in quasi-steady state (5 Myr after start of simulation). Left hand panels show subduction hinge, right hand panels show larger part of the model domain. (a) strain rate magnitude and (b) differential stress magnitude. Both of these scalar values are proportional to the square root of the second invariant ($J2$) of the deviatoric tensors. The colored lines perpendicular to the slab are discussed in Fig. 5 (c) deformation map; colors show the dominant deformation mechanism. Elastic deformation is defined as locations where the anelastic mechanisms contribute $< 25\%$ of the total strain rate.

ASPECT input files

Input files and a description of code modifications to reproduce the numerical model can be found at https://github.com/dansand/subduction_GJI2022.

Parameter name	**Value**	**Symbol**	**Units**
Model domain depth	2900	-	km
Model domain width	11600	-	km
Potential temperature	1623	T_m	K
Surface temperature	273	T_s	K
Viscosity minimum	5×10^{18}	-	Pa s
Viscosity maximum	1×10^{24}	-	Pa s
Dislocation Peierls volume	14×10^{-6}	V_{disl}	$m^3\,mol^{-1}$
Dislocation activation energy	520	E_{disl}	$kJ\,mol^{-1}$
Dislocation creep exponent	3.5	n	-
Dislocation creep prefactor	1.1×10^{-16}	A_{disl}	$Pa^{-n}\,s^{-1}$
Diffusion creep volume	4.4×10^{-6}	V_{diff}	$m^3\,mol^{-1}$
(Lower mantle)	5.13×10^{-6}		$m^3\,mol^{-1}$
Diffusion activation energy	300	E_{diff}	$kJ\,mol^{-1}$
Diffusion creep prefactor	3.84×10^{-11}	A_{diff}	$Pa\,s^{-1}$
(Lower mantle)	3.34×10^{-11}		$Pa\,s^{-1}$
Peierls creep prefactor	1.4×10^{-24}	A_p	$Pa^{-n}\,s^{-1}$
Peierls stress	5.9×10^{9}	σ_p	Pa
Peierls activation energy	272	E_p	$kJ\,mol^{-1}$
Peierls activation volume	7×10^{-6}	V_p	$m^3\,mol^{-1}$
first Peierls glide parameter	0.5	p	-
second Peierls glide parameter	2	q	-
Peierls creep exponent	2	n	-
Peierls fitting parameter	0.05	γ	-
Friction coefficient (mantle/weak layer)	0.8/0.005	μ	
Cohesion	20	C_0	MPa
Elastic shear modulus	10	μ	GPa
Thermal diffusivity	1×10^{-6}	-	m^2s^{-1}
Heat capacity	1250.0	C_p	$J\,K^{-1}\,kg^{-1}$
Elastic timestep	10^4	Δt_e	yr
Numerical timestep (max)	2×10^3	Δt_c	yr
Reference density	3300	ρ_0	$kg\,m^{-3}$
Thermal expansivity	3.5×10^{-5}	α	K^{-1}

Table S1: Parameters used in the numerical model.

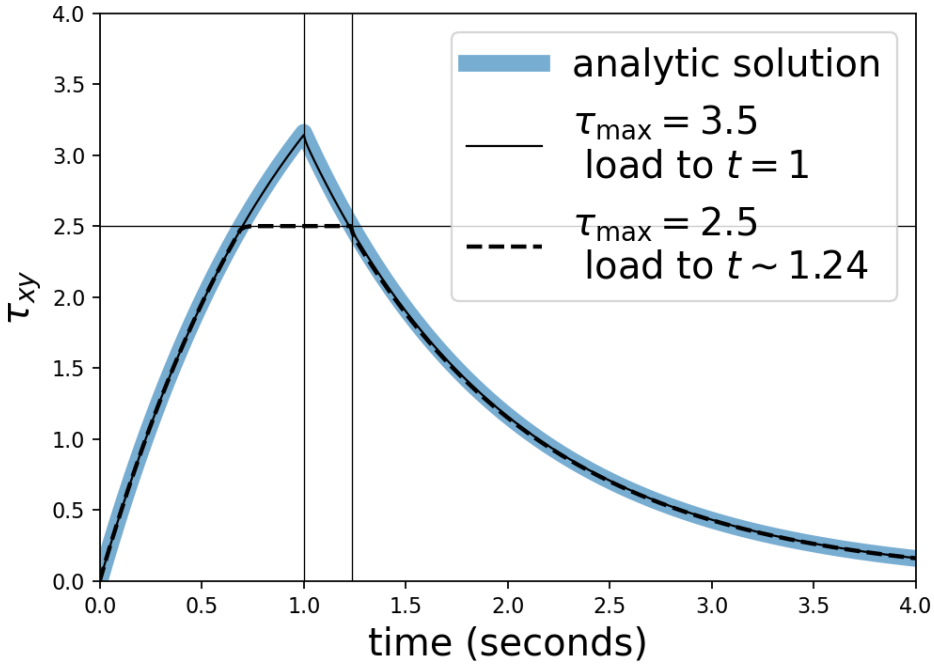

Figure S3: Benchmark for isoviscous simple shear, adapted from Farrington et al. (2014). Blue line shows analytical solution, the black lines show results from ASPECT.

S3 Insights from Numerical model

This section contains additional discussion and figures regarding the analysis and insights relating to the numerical model.

S3.1 Contributions to subducting plate topography

Fig. 4 (main manuscript) shows the topography that is developed in the numerical model subducting plate, at 5 Myr after the start of the simulation, when the subduction hinge has reached quasi-steady state (same time step as shown in Fig. S2). The model topography, shown with the red line is the combination of 3 main effects.

The first is the isostatic adjustment related to the density of the cooling plate, accounting for about $\sim$ 2.5 km of topographic variation from the ridge to the trench. To estimate the expected isostatic topography, we choose a point at the ridge axis, and integrate the density down to a specified depth below the surface (150 km), where it is assumed that compensation would be achieved. The integrated density along this profile forms a reference weight (i.e. the ridge forms the reference level). We then take points along the subducting plate surface, and take vertical profiles that start at the model surface. We compute the cumulative integral down those profiles. The point at which the cumulative integral equals the reference weight provides a depth, which is generally slightly smaller that the reference depth (150 km), because density is larger more in the plate, than at the ridge. The difference between the two depths is the estimated vertical isostatic adjustment (because it is the vertical adjustment required such that column weights are equal at a reference depth). This estimated isostatic adjustment is shown with black crosses in Fig. 4 (main manuscript). Note that the estimated adjustment is greater than the observed adjustment, particularly close to the trench. It is this deficit we next consider.

The second is the topography due to 'dynamic' pressure variations in the asthenosphere. Previous studies lead us to expect that the dynamic (or non-lithostatic) pressure variations in the asthenosphere may support several hundred meters of the topography, and hence explain part the deficit describe above (Schubert et al., 1978; Holt, 2022). To estimate the dynamic pressure, we calculate the variation of the pressure along the assumed reference depth (150 km), and subtract the value beneath the ridge as a reference level. This process yields a systematic pressure variation ΔP. In Fig. 4 (main manuscript) the pressure variation is converted to a topographic perturbation ($h_p = \Delta P/(\rho_m g)$) as shown with the dashed blue line. When we subtract this dynamic-pressure related topography from the model topography (solid red line), the resulting topography lies very close to the estimated isostatic values (dashed red line).

When we remove the isostatic and dynamic pressure contributions from the model topography, we are left with the residual topography, shown as the solid black line in Fig. 4 (main manuscript). Fig. S4 shows a closer view of the residual topography (solid black line). The dashed red line shows the pseudo-topography defined by the gradient of the shear stress resultant (e.g. Eq. 10) (discontinuities in this profile result from errors in interpolation which carry through to the numerical gradient calculation). Nevertheless, the close agreement between the residual topography and the pseudo-topography, shows that the former is, to a close approximation, the 'flexural topography', i.e. the topography supported by gradients in the shear stress resultant.

The flexural topography is therefore the third major contribution to the subducting plate topography. We note that the numerical model does not include the water column, meaning the flexural topography amplitude/wavelength cannot be directly compared to profiles from global subduction zones. However the observed moment (i.e. Eq. 7) is independent of the density contrast between rock and the overlying material; it is controlled only by the bending moment that is supported in the plate.

The blue line in Fig. S4 shows the elastic plate solution fit by least squares to the model topography, which emulates the workflow used by McNutt & Menard (1982) and Levitt & Sandwell (1995) in order to stabilise the moment integral when working with real data. This provides a sense of the limitations of trying to model the topography of a non-linear, yielding plate (at moment saturation) with the analytic solution of a uniform-thickness elastic plate. The bottom panels in Fig. S4 show the use of the topography (or the shear stress resultant) to estimate the bending moment (the value of which is given by the horizontal dashed line). The moment estimate based on the uniform elastic plate solution slightly underestimates the actual bending moment, compared with direct integration of the flexural topography (black line) or the shear stress resultant (red line).

S3.2 Stress and strain rate patterns

This section begins with a general description of dynamics of the numerical model in quasi-steady state, focusing particularly on the stress and deformation patterns within the subducting plate, as shown in Fig. S2 and Fig. 5 (main manuscript).

The magnitude of the strain rate tensor, and the magnitude of the differential stress are shown in Fig. S2a&b. The incoming subducting plate exhibits low differential stress, and strain rates consistent with long term-rigidity ($< 10^{-17} s^{-1}$). As bending occurs, the elastic core of the plate narrows with zones of brittle deformation and LTP-

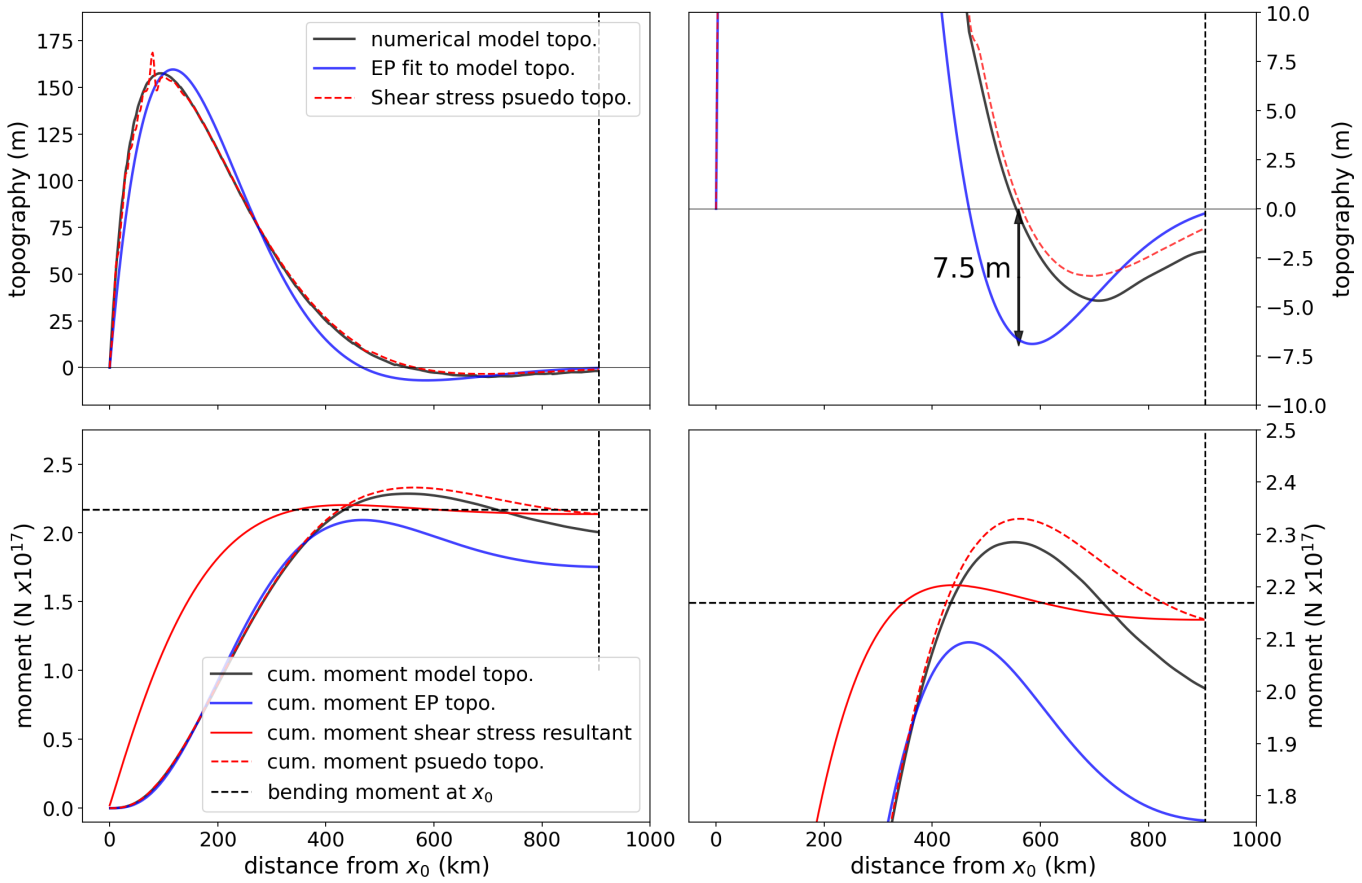

Figure S4: The left hand side shows model topography (top) and cumulative moment (bottom) based on either the flexural topography or the vertical shear stress resultant. The right hand side shows the same data at different scale. In the bottom panels the bending moment evaluated at the first zero crossing is shown with the black dashed line: this is the value that the estimates should recover. The flexural component of the model topography is shown in the black line, the best-fitting (L2) uniform-thickness elastic plate solution is shown in the blue line, as well as the pseudo-topography (the gradient of the vertical shear stress resultant) in the dashed red line. In the lower panels the horizontal dashed black line shows the bending moment calculated from the model at x_0.

dominated creep converging (black and green regions in Fig. S2c), until only a thin elastic core remains. At depths greater than about 800 °C in the plate, power law creep becomes the dominant deformation mechanism (as shown in the orange region). Peak differential stresses in the outer-rise region reach about 500 MPa. Beneath the forearc, the slab exhibits low strain rates, before it begins to unbend, reaching peak unbending rates beneath the mantle wedge. Peak differential stresses are higher in unbending, as the brittle deformation mechanism is precluded due to high confining pressure. Clearly the absence of brittle deformation is inconsistent with the abundant intraslab seismicity observed at intermediate depths in subduction zones. However the mechanism by which this intermediate-depth seismicity occurs is not of key importance to this study (nor is it really agreed on).

In Fig. 5 (main manuscript) the strain rate and differential stress are shown in a slab-relative 'downdip' reference frame. The strain rate tensor components rotated as described in the Figure caption, while the differential stress is negative when the most compressive principal stress is sub-parallel with the slab dip direction. The scalar fields shown equate to downdip stretching/shortening, and downdip (relative) tension/compression. Without this transformation, the relationship between slab geometry, and the components of the strain rate and stress tensors, is ambiguous, because the angle of slab with respect to the basis vectors changes along the slab. This can result, for instance, in the misinterpretation of a change in slab stress state at intermediate depths ('deep bending') (e.g., Sippl et al., 2022).

In the subduction hinge (i.e. the region shown in Fig. 5) there are two principal regions of deformation: in the subducting plate immediately seaward of the trench (the 'outer rise'), where curvature is increasing, and the unbending zone, where curvature reduces. Deformation rates are low in the intervening section of the slab, where curvature passes its maxima, consistent with the assertion that the geometry of the slab is in quasi-steady state, and the curvature rate is dominated by the advective component of the material derivative (e.g., Kawakatsu, 1986; Ribe, 2010; Buffett & Becker, 2012). The model results concur with the observation of Sandiford et al. (2020), that thermo-mechanical models of long term stable subduction (10s Myrs) often produce slabs that are dominated by unbending at intermediate depth, rather than by uniform stretching/shortening due to driving and resisting forces (e.g., Sandiford et al., 2020; Bessat et al., 2020; Toffol et al., 2022).

Stress profiles at four locations are shown in Fig. 5 (main manuscript). The blue line labelled x_0 is the first zero crossing based on analysis of the flexural component of the topography (as discussed in the previous section). The black line is located at the region of maximum moment saturation. As the slab begins unbending it first behaves elastically in a manner unaffected by the prior yielding, until it begins to undergo yielding in the opposite stress orientation. However, because the bending strain depends on distance from the neutral plane, a complex stress pattern can develop (e.g., Engdahl & Scholz, 1977; Wang, 2002). The green line shows the location of partial slab unbending, where the stress profile has multiple zero crossings due to these elastic-anelastic interactions. This zone of partial unbending occurs over approximately 50 km in the downdip direction, where the potential for diverse earthquake focal mechanisms would be expected (although often difficult to image, due to its overlap with the active section of the overlying megathrust and forearc). The red line shows the region of full unbending, where a simple flexural stress profile with single zero crossing has re-established.

In the outer-rise region of the model, the transition between downdip extension and shortening occurs at depths of $\sim$ 30 km and temperatures of $\sim$ 400 °C; indeed this is not a coincidence, as we used the YSE approach to calibrate the LTP flow parameters, for the given friction coefficient. As discussed in the main manuscript, the saturation moment

values exhibited by this strength model ($\sim$ 2.7 $\times 10^{17}$ N) are higher than most of the global observations for slabs in the 100 Myr range.

Strain rate patterns in the numerical model are shown in Fig. S5. The solid lines shows the axial strain rate taken from the numerical model at the first zero crossing (x_0) in blue and the maximum bending moment in black. The strain rates show a reasonably good agreement with the simple bending approximation, i.e. a linear increase away from the neutral plane. The strain rates have been scaled to account for a reference convergence rate of 10 cm/y (the model convergence rate is $\sim$ 14.5 cm/y). This scaling is based on the linear relationship between the axial strain rate and the bending rate. The strain rates approximately double between the x_0 and the point of maximum bending moment. The red dashed line shows the strain rate we assumed in the YSE calculations.

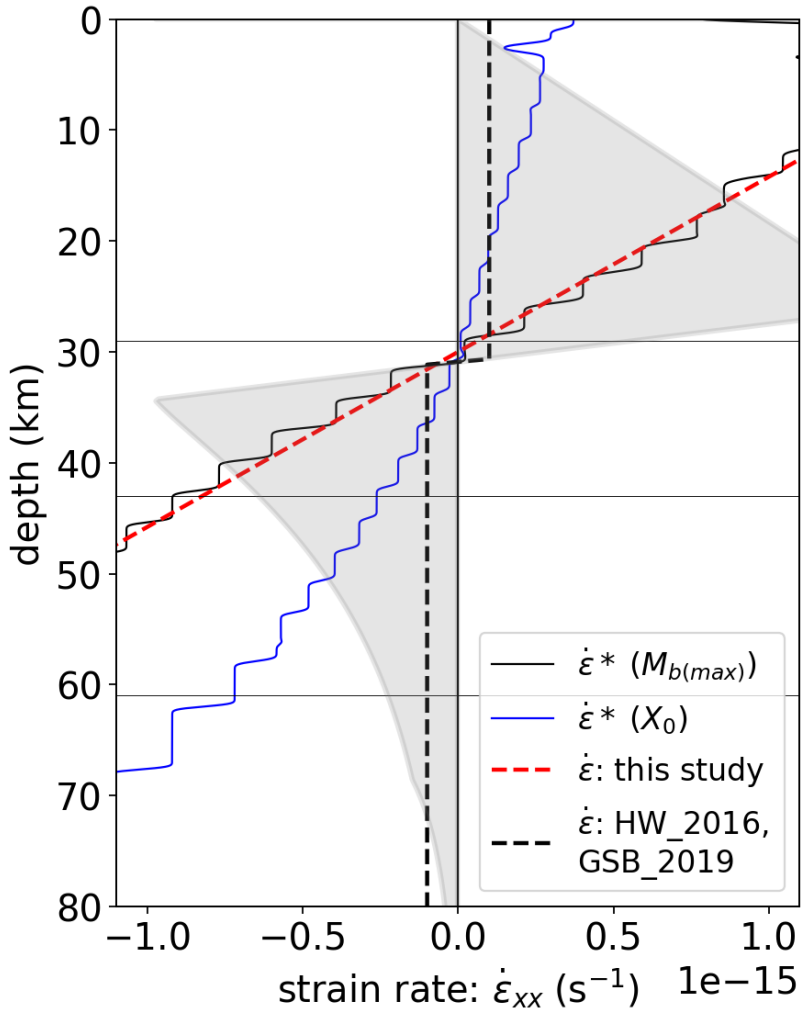

Figure S5: Strain rates in the bending lithosphere. Solid lines shows the axial strain rate in the numerical model at the first zero crossing (x_0) in blue and the maximum bending moment in black (the profiles shown in Fig. 5 (main manuscript)). The stair-step nature of the strain rates is due to the post-processing, in which strain rates are averaged across elements. The dashed lines show the strain rate assumptions we use in YSE model in this study (red dashed line) along with the much lower (constant) strain rates used in Garcia et al. (2019) . Although Hunter & Watts (2016) nominally consider bending rates in their YSE models, they resolve these as a constant strain rate, with a typical magnitude of $10^{-16}s^{-1}$ (as best as we can determine).

S4 Yield Strength Envelope calculations and parameters

In the YSE approach, the strain associated with bending of the lithosphere is prescribed kinetically. The model assumes plane strain in the trench-parallel direction, and plane stress in the vertical direction (Turcotte & Schubert, 2002). The horizontal stress is defined as the sum of the lithostatic pressure and a horizontal stress perturbation due to flexural strain. Hence we have:

$$\sigma_{yy} = \rho g y \tag{25}$$
$$\sigma_{xx} = \rho g y + \Delta\sigma_{xx}$$

Under the assumptions described above, the relation between the horizontal stress and strain is:

$$\Delta\sigma_{xx} = \left(\frac{E}{1-\nu^2}\right)\epsilon_{xx} \tag{26}$$

where E is the Young's modulus and ν is Poisson's ratio.

The axial strains ϵ_{xx} are linearly-related to the curvature, through the simple bending approximation:

$$\Delta\epsilon_{xx} = (y - y_n)\Delta K \tag{27}$$

where K indicates the curvature of the neutral surface. Stresses in the subducting plate develop due to the incremental curvature increase and are limited as various analastic deformation mechanisms are activated, depending on pressure, temperature and stress. The neutral plane depth is iteratively determined so that the integrated stress profile matches a prescribed net axial force (e.g., Mueller et al., 1996).

Rather than evaluate a YSE at a prescribed curvature and constant strain rate, we use a kinematic description of the curvature variation and strain rate based on a simple bending model. The system is assumed to be in geometric steady state, such that the horizontal strain rate is given by the advective derivative:

$$\dot{\epsilon}_{xx} = -(y - y_n)\frac{DK}{Dt} \approx -yu_x\frac{\partial K}{\partial x} \tag{28}$$

where y_n is the neutral plane depth, K is the curvature and u_x is the horizontal velocity component. The curvature gradient (2×10^{-11}) is chosen so that the strain rate at 10 cm/y is equivalent to the numerical model strain rates (the latter also being rescaled to a reference value of being rescaled 10 cm/y, as described in the previous section).

The YSE analysis uses a plate cooling model, with a thickness of 125 km based on the parameters from (Parsons & Sclater, 1977). The thermal model is identical to the studies of Hunter & Watts (2016); Garcia et al. (2019). Relevant parameters are provided in Table 1 (main manuscript). Note that the thermal diffusivity used in the YSEs is slightly lower that was adopted in the numerical model (0.8 rather than 1.0 ($\times 10^{-6}\mathrm{m}^2\,\mathrm{s}^{-1}$)).

Fig. S6 shows a comparison between the strain rate assumptions used in this study, and those of a recent study (Hunter & Watts, 2016). In both cases the creep parameters are the 'Set 1' parameters (see Table 1, main manuscript). Hunter & Watts (2016) argue that these creep parameters, with a friction coefficient of either 0.3, or 0.6 (assuming hydrostatic pore pressure) were consistent with results from their variable-thickness flexural inversion. These two models are shown in red in Fig S6, and both have saturation moments in the intermediate range (1.4 and 1.8 $\times 10^{17}$ N). Under the strain rate assumptions made in this study, the saturation moments are significantly larger (almost 30 %). The model with $\mu = 0.6$ is now in the 'strong' range. Under our assumptions the 'Set 1' parameters would be considered too strong.

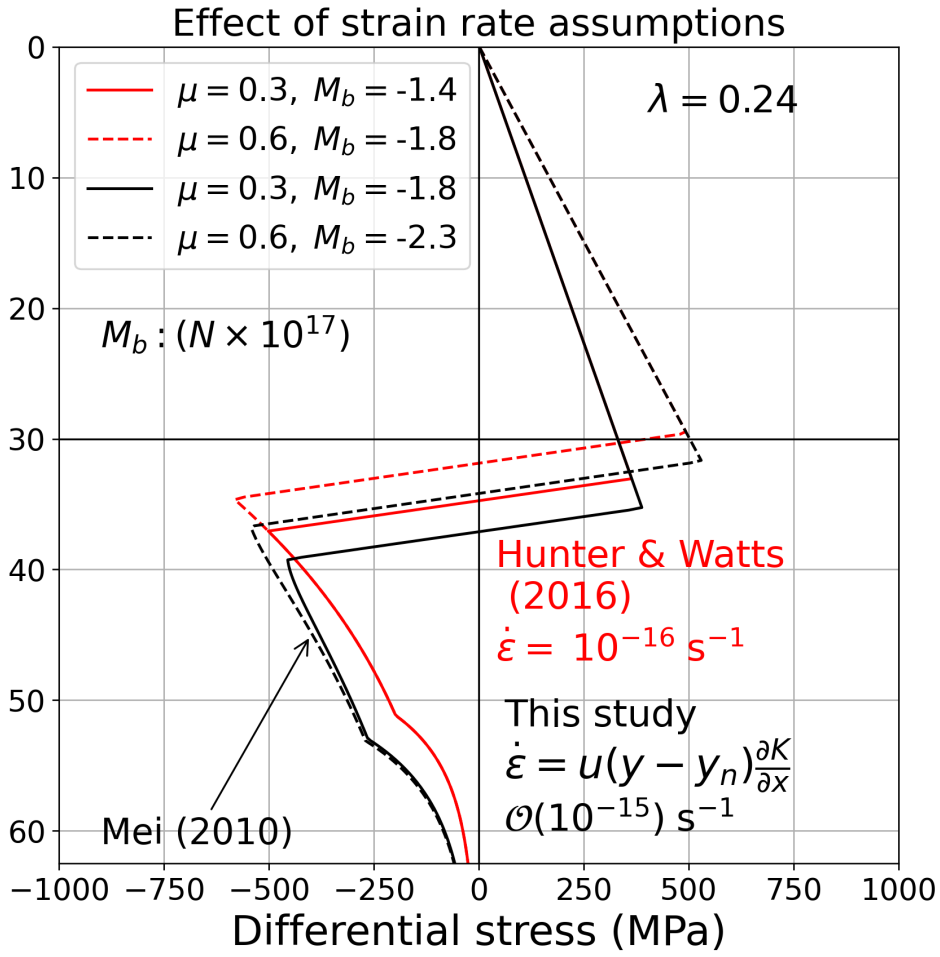

Figure S6: Comparison between the strain rate assumptions used in this study, and those of (Hunter & Watts, 2016). All YSEs have the same creep parameters (Set 1, Table 1, main manuscript). Two different friction coefficients are shown. Both resulting strength models were considered consistent with the optimised elastic plate values in Hunter & Watts (2016). The difference between the red and black lines is due to the strain rate model we adopt in this study, which is proportional to distance from the neutral plane, and has significant higher values that the constant rate ($10^{-16}s^{-1}$) assumed inHunter & Watts (2016). The result of our assumptions is a stronger model.

S5 Additional figures and text

Fig. S7 compares the curvature values (at the first zero crossing) from the data set of Levitt & Sandwell (1995), with predicted moment-curvature evolution, based on YSE analysis. Here we have adopted a 'strong' lithopsheric strength model, in keeping with the model (although not identical) advocated in Levitt & Sandwell (1995). The objective is to assess their interpretation that moment deficit, is the primary cause of scatter in the Levitt & Sandwell (1995) data.

Fig. S7 leads us to two conclusions. First, the majority ($\sim$ 70 %) of the profiles in the Levitt & Sandwell (1995) are expected to have reached $>$ 75% moment saturation. Hence, most of the data should exhibit less that $<$ 25% moment deficit if the curvature estimates are accurate. Perhaps more importantly, the moment deficit does not show the anticipated correlation with radius of curvature, as shown by the dashed line and arrow in the lower panel. In contrast, most of the data with largest moment deficits, lie in the upper end of the curvature range (small radii of curvature), and would be expected to have negligible moment deficit. These conclusions are inconsistent with the argument that moment deficit is the primary cause of scatter in the data.

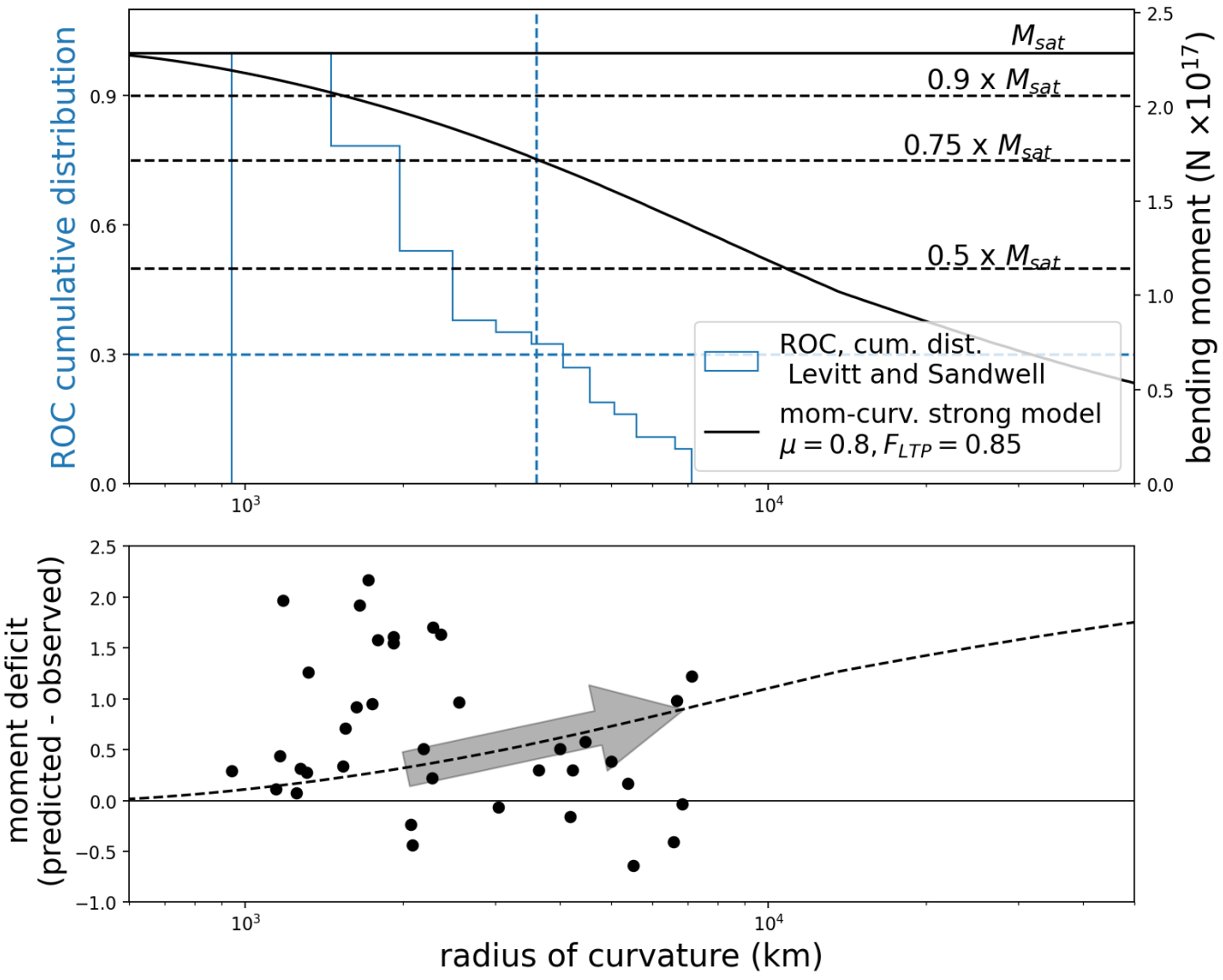

Figure S7: Top panel shows the distribution of curvature values (shown as the reciprocal - radius of curvature) from the profiles of Levitt & Sandwell (1995). Black line in top panel shows the evolution of moment as a function of curvature, for a strong lithosphere model. Bottom panel shows the predicted moment deficit, which is the predicted moment, minus the observed moment, using the rheological model as shown in top panel. The trend of the moment deficit should follow the dashed line, if moment deficit is a primary cause of the scatter, which is not the case.

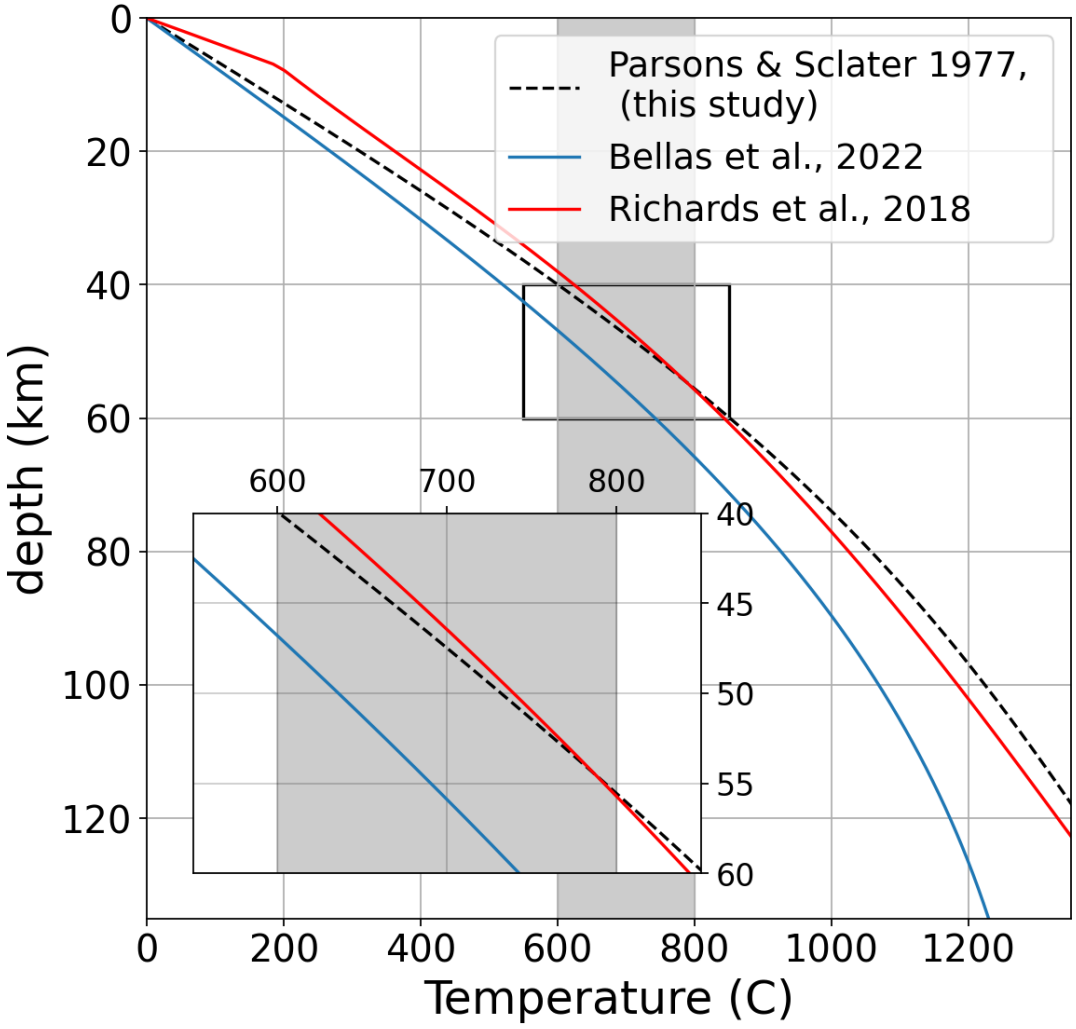

Figure S8: Temperature model comparison for 100 Myr lithosphere. Black dashed lined shows plate model parameters from Parsons & Molnar (1976). The red line shows the model from Richards et al. (2018), which is a numerical-based model based on joint inversion for subsidence and heat flow. The blue line shows the temperature model used in Bellas et al. (2022). The cooler temperature profile is due to the higher diffusivity assumed (1.0, rather than 0.8 $\times 10^{-6} \mathrm{m}^2\,\mathrm{s}^{-1}$). The grey regions shows 600-800 °C, where the temperature dependence of strength is largest.

References

Bangerth, W., Dannberg, J., Gassmoeller, R., & Heister, T., 2020. Aspect v2.2.0.

Bangerth, W., Dannberg, J., Gassmöller, R., Heister, T., et al., 2020b. ASPECT: Advanced Solver for Problems in Earth's ConvecTion, User Manual, doi:10.6084/m9.figshare.4865333.

Bellas, A., Zhong, S., & Watts, A. B., 2022. Reconciling lithospheric rheology between laboratory experiments, field observations and different tectonic settings, *Geophysical Journal International*, **228**(2), 857–875.

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

Buffett, B. & Becker, T., 2012. Bending stress and dissipation in subducted lithosphere, *Journal of Geophysical Research: Solid Earth*, **117**(B5).

Engdahl, E. & Scholz, C., 1977. A double benioff zone beneath the central aleutians: An unbending of the lithosphere, *Geophysical Research Letters*, **4**(10), 473–476.

Farrington, R., Moresi, L.-N., & Capitanio, F. A., 2014. The role of viscoelasticity in subducting plates, *Geochemistry, Geophysics, Geosystems*, **15**(11), 4291–4304.

Garcia, E. S. M., Sandwell, D. T., & Bassett, D., 2019. Outer trench slope flexure and faulting at pacific basin subduction zones, *Geophysical Journal International*, **218**(1), 708–728.

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

Hirth, G. & Kohlstedt, D., 2003. Rheology of the upper mantle and the mantle wedge: A view from the experimentalists, *Geophysical monograph-american geophysical union*, **138**, 83–106.

Holt, A. F., 2022. The topographic signature of mantle pressure build-up beneath subducting plates: Insights from spherical subduction models, *Geophysical research letters*, **49**(22), e2022GL100330.

Hunter, J. & Watts, A., 2016. Gravity anomalies, flexure and mantle rheology seaward of circum-pacific trenches, *Geophysical Journal International*, **207**(1), 288–316.

Kameyama, M., Yuen, D. A., & Karato, S.-I., 1999. Thermal-mechanical effects of low-temperature plasticity (the peierls mechanism) on the deformation of a viscoelastic shear zone, *Earth and Planetary Science Letters*, **168**(1-2), 159–172.

Kaus, B. J., Mühlhaus, H., & May, D. A., 2010. A stabilization algorithm for geodynamic numerical simulations with a free surface, *Physics of the Earth and Planetary Interiors*, **181**(1-2), 12–20.

Kawakatsu, H., 1986. Double seismic zones: Kinematics, *Journal of Geophysical Research: Solid Earth*, **91**(B5), 4811–4825.

Levitt, D. A. & Sandwell, D. T., 1995. Lithospheric bending at subduction zones based on depth soundings and satellite gravity, *Journal of Geophysical Research: Solid Earth*, **100**(B1), 379–400.

McNutt, M. K. & Menard, H., 1982. Constraints on yield strength in the oceanic lithosphere derived from observations of flexure, *Geophysical Journal International*, **71**(2), 363–394.

Mei, S., Suzuki, A., Kohlstedt, D., Dixon, N., & Durham, W., 2010. Experimental constraints on the strength of the lithospheric mantle, *Journal of Gmersoneophysical Research: Solid Earth*, **115**(B8).

Moresi, L., Dufour, F., & Mühlhaus, H.-B., 2003. A lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials, *Journal of computational physics*, **184**(2), 476–497.

Mueller, S., Choy, G. L., & Spence, W., 1996. Inelastic models of lithospheric stress-i. theory and application to outer-rise plate deformation, *Geophysical Journal International*, **125**(1), 39–53.

Parsons, B. & Molnar, P., 1976. The origin of outer topographic rises associated with trenches, *Geophysical Journal International*, **45**(3), 707–712.

Parsons, B. & Sclater, J. G., 1977. An analysis of the variation of ocean floor bathymetry and heat flow with age, *Journal of geophysical research*, **82**(5), 803–827.

Ranalli, G., 1995. *Rheology of the Earth*, Springer Science & Business Media.

Ribe, N. M., 2010. Bending mechanics and mode selection in free subduction: A thin-sheet analysis, *Geophysical Journal International*, **180**(2), 559–576.

Richards, F., Hoggard, M., Cowton, L., & White, N., 2018. Reassessing the thermal structure of oceanic lithosphere with revised global inventories of basement depths and heat flow measurements, *Journal of Geophysical Research: Solid Earth*, **123**(10), 9136–9161.

Sandiford, D. & Moresi, L., 2019. Improving subduction interface implementation in dynamic numerical models, *Solid Earth*, **10**(3), 969–985.

Sandiford, D., Moresi, L., Sandiford, M., Farrington, R., & Yang, T., 2020. The fingerprints of flexure in slab seismicity, *Tectonics*, **39**(8), e2019TC005894.

Sandiford, D., Brune, S., Glerum, A., Naliboff, J., & Whittaker, J. M., 2021. Kinematics of footwall exhumation at oceanic detachment faults: solid-block rotation and apparent unbending, *Geochemistry, Geophysics, Geosystems*, **22**(4), e2021GC009681.

Schmalholz, S., Podladchikov, Y., & Schmid, D., 2001. A spectral/finite difference method for simulating large deformations of heterogeneous, viscoelastic materials, *Geophysical Journal International*, **145**(1), 199–208.

Schubert, G., Yuen, D. A., Froidevaux, C., Fleitout, L., & Souriau, M., 1978. Mantle circulation with partial shallow return flow: Effects on stresses in oceanic plates and topography of the sea floor, *Journal of Geophysical Research: Solid Earth*, **83**(B2), 745–758.

Sippl, C., Dielforder, A., John, T., & Schmalholz, S. M., 2022. Global constraints on intermediate-depth intraslab stresses from slab geometries and mechanisms of double seismic zone earthquakes, *Geochemistry, Geophysics, Geosystems*, **23**(9), e2022GC010498, e2022GC010498 2022GC010498.

Steinberger, B. & Calderwood, A. R., 2006. Models of large-scale viscous flow in the earth's mantle with constraints from mineral physics and surface observations, *Geophysical Journal International*, **167**(3), 1461–1481.

Toffol, G., Yang, J., Pennacchioni, G., Faccenda, M., & Scambelluri, M., 2022. How to quake a subducting dry slab at intermediate depths: Inferences from numerical modelling, *Earth and Planetary Science Letters*, **578**, 117289.

Turcotte, D. L. & Schubert, G., 2002. *Geodynamics*, Cambridge university press.

Turcotte, D. L., McAdoo, D., & Caldwell, J., 1978. An elastic-perfectly plastic analysis of the bending of the lithosphere at a trench, *Tectonophysics*, **47**(3-4), 193–205.

Wada, I. & Wang, K., 2009. Common depth of slab-mantle decoupling: Reconciling diversity and uniformity of subduction zones, *Geochemistry, Geophysics, Geosystems*, **10**(10).

Wang, K., 2002. Unbending combined with dehydration embrittlement as a cause for double and triple seismic zones, *Geophysical research letters*, **29**(18), 36–1.