Initiation of the Andean orogeny by lower mantle subduction

Claudio Faccenaa,*, Onno Onckenb, Adam F. Hots, Thorsten W. Beckerd

a Dept. Sciences, University Roma TRE, Roma, Italy
b Helmholtz Centre Potsdam, GFZ German Research Centre Geosciences, Potsdam, Germany
c Dept. Earth Atmos. and Planet. Science, Massachusetts Institute of Technology, USA
d Jackson School of Geosciences, The University of Texas at Austin, USA

A R T I C L E   I N F O
Article history:
Received 11 February 2016
Received in revised form 24 November 2016
Accepted 29 January 2017

Keywords:
Andes
subduction
orogeny
mantle convection
lower mantle

A B S T R A C T

The Cordillera of the Andes is a double-vergent orogenic belt built up by thickening of South American plate crust. Several models provide plausible explanations for the evolution of the Andes, but the reason why shortening started at ~50 Ma is still unclear. We explore the evolution of the subduction zone through time by restoring the position of the Nazca trench in an absolute reference frame, comparing its position with seismic tomography models and balancing the evolution of the subducting slab. Reconstructions show that the slab enters into the lower mantle at ~50 ± 10 Ma, and then progressed, moving horizontally at shallow lower mantle depth while thickening and folding in the transition zone. We test this evolutionary scenario by numerical models, which illustrate that compression in the upper plate intensifies once the slab is anchored in the lower mantle. We conclude that onset of significant shortening and crustal thickening in the Andes and its sustained action over tens of million years is related to the penetration of the slab into the lower mantle, producing a slowdown of lateral slab migration, and dragging the upper plate against the subduction zone by large-scale return flow.

1. Introduction

The Cordillera are a 15,000 km long orogen formed during the subduction of the Farallon and Nazca plates beneath the Americas back from the Mesozoic. The recent Andean belt of South America formed during Early Tertiary since up to the Late Cretaceous Southern America was under extension.

The Andes are constituted by South American upper crust slices, forming a double-vergent belt, and by massif intrusion of igneous material, reaching a maximum crustal thickness of ~70 km along the Bolivian orocline (e.g., Kley and Monaldi, 1998). The origin of the Andean belt is still controversial, as significantly thick crust is formed during oceanic subduction, i.e. without continental collision. From a kinematic point of view, shortening is related to the westward advance of South America toward a stationary/slowly retreating Nazca trench.

So the main dynamic question is why the westward, retrograde migration of the trench slowed down in the Early Tertiary along a large portion of the Nazca slab?

Russo and Silver (1994) suggested that the Nazca trench is stationary around the Bolivian orocline because the positive dynamic pressure within the sub-slab mantle in the center of the slab prevents slab retreat. Long (along trench) subduction zones would be expected to migrate backward slower due to the increasing resistance of the sub-slab mantle (e.g., Schellart et al., 2007). However, this model is unable to explain why the Nazca subduction system rolled back during the opening of the Southern Atlantic for almost ~100 Myr, and only in the Early Tertiary decreased its retrograde motion (Husson et al., 2012).

Other models suggest that the growth of the Cordillera may be related to the variation of buoyancy of the subducted slab, following episodes of flat subduction with an increase in plate coupling (e.g., Isacks, 1988; Martinod et al., 2010). However, regions with contemporary flat subduction do not record any increase in shortening and there is poor correlation between episodes of shortening and the inferred arrival of ridges at trenches. The buoyancy of the subducting plate may matter, as could the strength/thickness of the upper plate. Numerical models show that spatial variations of the age of the subducting slab and the thickness of the upper plate may produce variations in shortening of the upper plate (Capitanio et al., 2011).

All those models provide plausible explanations for the evolution of the Andes (Oncken et al., 2006), but the reason why shortening started in the Early Tertiary is still unclear. Here, we propose that the onset of shortening in the Andes and its sustained action over tens of million years is related to the deep mantle dynamics

http://dx.doi.org/10.1016/j.epsl.2017.01.041
0012-821X/© 2017 Elsevier B.V. All rights reserved.
of the subducting slab. The concept, first explored in Husson et al. (2012) and Faccenna et al. (2013) in more general terms, is that the penetration of the slab into the lower mantle results in a twofold effect: 1) it stabilizes the trench, slowing down lateral slab migration, and, 2) the upper plate is dragged against the subduction zone by large scale return flow. Here, we provide new analysis of geophysical and geological data and a simple model to argue that subduction into the lower mantle may have indeed triggered the onset of Andean orogeny.

2. The Nazca subduction system

The trace of the Nazca subduction zone is well depicted by seismicity. The Wadati–Benioff zone is dipping to the East at ∼30° down to 300 km in the central portion of the Cordillera. This steep portion is flanked by two flat slab segments, extending from 3° to 15°S (Peruvian flat slab) and from 28° to 33°S (Chilean flat slab). Along those segments, seismicity reaches a depth of ∼200 km. A cluster of deep seismicity at ∼600 km depth is present beneath the Peruvian flat segment and beneath the Bolivian oroclone, delimiting a mid upper-mantle earthquake gap (Fig. 1). The presence of the flat slab is also marked by the paucity of recent arc volcanism. Indeed, arc volcanism stopped around ∼4 Ma, five million years after the arrival of aseismic ridges at the trench (Nazca, Carnegie and Juan Fernandez ridges). The subduction of those ridges, characterized by thick and buoyant gabbroid crust, has been considered as a possible cause for the origin of the flat slab (e.g., Gutscher et al., 2000; Martinod et al., 2010).

The subducting Nazca slab has been imaged in the mid-upper mantle by a number of regional tomography studies showing the continuity of high velocity anomalies (e.g., Bianchi et al., 2013; Scire et al., 2015, 2016). Those studies also show the presence of
a remarkable low velocity anomaly beneath the continuation of the Nazca ridges (13–16°S), a slab tear just north of the ridges, at \( \sim 10^\circ S \) (Scire et al., 2016) and a possible disruption of the slab anomaly at \( \sim 26^\circ S \) between 100 and 200 km depth (Bianchi et al., 2013).

High velocity seismic anomalies are also well defined in the aseismic portion of the slab and indicate that the anomaly gets thicker in the mantle transition zone (Li et al., 2008; Ribe et al., 2007; Scire et al., 2015). Global tomographic models also show a well-defined high velocity zone for the Nazca slab (e.g., Bijwaard et al., 1998; Ren et al., 2007; Li et al., 2008; Fukao and Obayashi, 2013). The high velocity anomaly (\( > 0.4% \)) is limited to the northern sector, i.e. north of 28°S, where subduction can be followed down to \( \sim 1400 \) km depth, and progressively fades out and gets shallower moving southward (e.g., Figs. S1–S3; Bijwaard et al., 1998; Li et al., 2008; Simmons et al., 2010; Fukao and Obayashi, 2013). Figs. 2 and 4a show a slice of the Li et al. (2008) model at 900 km depth, suggesting that the deep, subduction-related high velocity anomaly (\( > 0.4% \)) is indeed limited to extent from 5°N to 25°S, and then progressively fades out and get shallower southward (Fig. 4d), where the slab appears restricted to the upper mantle (Fukao and Obayashi, 2013). Some tomography models (e.g., Simmons et al., 2010) show a deeper high elliptical velocity anomaly (Fig. S1(g–n), 1600–1900 km depth) localized beneath Brazil from \( \sim 10^\circ \) to \( \sim 25^\circ S \) and from \( \sim 45^\circ \) to \( \sim 60^\circ E \). The importance on subduction and mantle dynamics on the Nazca-South America plate system is attested by dynamic topography calculations. Those models predicting a large scale tilting of South America while overriding the deep slab (Shephard et al., 2010; Flament et al., 2014, 2015). A remarkable negative dynamic signal is expected on the site of subducting slab, particularly evident in its northern portion (Flament et al., 2015), while the formation of flat slab and window that may locally generate dynamic uplift (Flament et al., 2015; Davila and Lithgow-Bertelloni, 2015).

3. Convergence velocity, shortening and trench restoration

At \( \sim 130 \) Ma, the South Atlantic ocean basin started opening, with the separation of South America from Africa. Spreading reached an average half rate of \( \sim 2.2 \) cm/yr with two phases of deceleration, between \( \sim 70 \) to 50 Ma, and during the last \( \sim 10 \) Ma. The first phase is due to a change in the Africa motion from NNE to NNW (e.g., Müller et al., 2008). From \( \sim 45 \) to \( \sim 25 \) Ma, the average half spreading rate rises up again up to \( \sim 2 \) cm/yr due to a modest increase in South American westward drift and to the NNE motion of Africa (Müller et al., 2008). The Nazca plate velocity oscillates and in the Tertiary attains two peaks, of 10–12 cm/yr, between 50 and 40 Ma and 30 to 15 Ma (Sdrolias and Müller, 2006; Seton et al., 2012). Average convergence rates between South America and Nazca in the Tertiary are \( \sim 10 \) cm/yr (Sdrolias and Müller, 2006; Seton et al., 2012). Error bars for spreading rates are generally of the order of few mm/yr.

The absolute plate motion of South America is more uncertain, depending upon the amount and variation of the net rotation and on the hot spot model (e.g., Williams et al., 2015). The Seton et al. (2012) model shows a westward motion of South America that sharply decreases at \( \sim 75–80 \) Ma from \( \sim 5 \) to \( \sim 2.5–3 \) cm/yr, and then from 40 Ma progressively decreases down to the present-day rate of \( \sim 1 \) cm/yr (Fig. 3, Müller et al., 2008; Seton et al., 2012). This trend is confirmed by other models. For example, the drop in velocity in the late Cretaceous is more pronounced in the Torsvik and Cocks (2016) model with respect to Seton et al. (2012) (Fig. S4), but the position of the trench at 50 Ma is similar. More pronounced differences between different models emerge when going back in the Mesozoic (e.g., Williams et al., 2015; Torsvik and Cocks, 2016).

Crustal shortening in the Andes is not uniform. It changes along strike both in onset timing and rate. For the analysis of the spatial and temporal distribution of deformation we adopted and enlarged the data base by Oncken et al. (2006, 2013). We evaluated several published balanced cross sections and compared related values with estimates from crustal area balancing in order to obtain a conservative estimate of the amount of shortening, as well as thermo-chronology and syn-tectonic deposits to estimate the timing of deformation. Our compilation of data along with original references is provided in Table 1 (revised from Oncken et al., 2006, 2013). The variability in shortening is illustrated on Fig. 4b where it is clear that post-Eocene shortening estimates peak in the orocline with up to \( \sim 270–300 \) km (15°–20°S), vary little between 17° and 21°S and then progressively fade laterally to \( \sim 150 \) km (10°S and 22°S to 32°S) and \( \sim 80 \) km (5°S or 32°S)
To the North, shortening increases again to ~150 km in the Colombian Cordilleras (e.g., Mora et al., 2009). This trend is in good agreement with previous estimates based on crustal area and cross-section balancing, but underestimates the maximum shortening in the 17–21°S region estimated by map-view restoration by 10–15% (see for example Arriagada et al., 2008; Eichelberger and McQuarrie, 2015, and references therein).

The timing of the onset of crustal shortening is not uniform either. Our data compilation along with references are given in Table 2 (revised from Oncken et al., 2013) and illustrated on Fig. 4c. The estimates are derived by defining, for each location, the start and the end of deformation activity from dating various type of deformation-related features. Although the onset of shortening in some regions may be older, significant shortening and crustal thickening initiated at ~50–45 Ma to the north of 25–30°S, and this is consistent with the earliest evidence of high elevation at ~40 Ma (De Celles et al., 2015 and reference therein). To the south of 25–30°S, the onset shortening started ~20 Ma.
Table 1
Amount of shortening along strike of the Andean belt [see also Oncken et al., 2013]. References are listed in Supplementary materials.

<table>
<thead>
<tr>
<th>S lat.</th>
<th>Pacific piedmont</th>
<th>Altiplano-Puna; West Cordillera</th>
<th>Eastern Cordillera</th>
<th>Foreland belt</th>
<th>Total</th>
<th>Total, preferred</th>
<th>Shortening from section area</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>&gt;10</td>
<td>&gt;25\sup{b}</td>
<td>27</td>
<td>&gt;62</td>
<td>70</td>
<td>60\sup{c}</td>
<td></td>
<td>Mégard, 1984; Rodriguez et al., 2001, 2003; Baby et al., 2013</td>
</tr>
<tr>
<td>7</td>
<td>&gt;10</td>
<td>30–50\sup{b}</td>
<td>84–100</td>
<td>120–150</td>
<td>140</td>
<td>120\sup{c}</td>
<td></td>
<td>Hermoza et al., 2005; Eude et al., 2015</td>
</tr>
<tr>
<td>10</td>
<td>5\sup{b}</td>
<td>30</td>
<td>50\sup{b}</td>
<td>56–84</td>
<td>140</td>
<td>140\sup{c}</td>
<td></td>
<td>Mégard, 1984; Rodriguez et al., 2003; Espurt et al., 2008, 2011; Scherrenberg et al., 2014</td>
</tr>
<tr>
<td>15</td>
<td>5\sup{b}</td>
<td>15–40</td>
<td>123</td>
<td>100–121</td>
<td>238–284</td>
<td>270</td>
<td></td>
<td>Baby et al., 1989, 1997; Roeder and Chamberlain, 1995; Rodriguez et al., 2001, 2003; McQuarrie et al., 2008</td>
</tr>
<tr>
<td>18</td>
<td>4–6</td>
<td>14–47</td>
<td>&gt;132</td>
<td>90\sup{a}–135</td>
<td>&gt;231–320</td>
<td>255–320</td>
<td></td>
<td>Sheffels, 1990; Baby et al., 1993, 1997; Roeder and Chamberlain, 1995; Munoz and Charrier, 1996; McQuarrie and DeCelles, 2001; Garcia et al., 2002</td>
</tr>
<tr>
<td>20–21</td>
<td>3</td>
<td>60</td>
<td>35–95</td>
<td>96–135</td>
<td>194–293</td>
<td>320\sup{c}</td>
<td></td>
<td>Héral et al., 1990; Baby et al., 1992; Dunn et al., 1995; Kley, 1996; Müller et al., 2002; Victor et al., 2004; Elger et al., 2005; Uba et al., 2009</td>
</tr>
<tr>
<td>24–26</td>
<td>4–8</td>
<td>30–50</td>
<td>40–60\sup{b}</td>
<td>21–30</td>
<td>95–148</td>
<td>125</td>
<td>260\sup{c}</td>
<td>Grier et al., 1993; Cladouhos et al., 1994; Cristallini et al., 1997; Costand et al., 1999; Kley and Monaldi, 2002</td>
</tr>
<tr>
<td>30–31</td>
<td>3–10</td>
<td>15–35</td>
<td>100–132</td>
<td>118–177</td>
<td>160</td>
<td>100–160\sup{c}</td>
<td></td>
<td>Allmendinger et al., 1997; Cristallini and Ramos, 2000; Ramos et al., 2002; Victor and Echtler, 2006</td>
</tr>
<tr>
<td>33–34</td>
<td>5</td>
<td>58–67\sup{a}</td>
<td>63–72</td>
<td>75</td>
<td>70–100 \sup{d}</td>
<td></td>
<td>Giambiagi et al., 2003; Victor and Echtler, 2006</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>5</td>
<td>28–53</td>
<td>33–58</td>
<td>58</td>
<td>60\sup{c}</td>
<td></td>
<td>Folguera et al., 2007; Victor and Echtler, 2006</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>5</td>
<td>17</td>
<td>22</td>
<td>22</td>
<td>25\sup{c}</td>
<td></td>
<td>Victor and Echtler, 2006</td>
<td></td>
</tr>
</tbody>
</table>

\sup{a} Measured from cross-section (no value given in ref.) or alternative estimate (this paper).
\sup{b} Estimated from average shortening ratio of unit along strike when associated to similar elevation.
\sup{c} Data from Kley and Monaldi (1998) and Victor and Echtler (2006).
\sup{d} Including Interandean zone.

We conclude that the arcuate shape of the Bolivian orocline is due to the different behavior of the northern and southern segments: relatively small shortening in northern Peru, with episodic pulses (no or slight shortening from 40 to 25 Ma) whereas in the southern segment, the onset of shortening is significantly delayed.

Using this data set (Tables 1 and 2), we were able to correct the position of the Nazca trench back in time during the last 100 Ma. This new restoration is based on Seton et al.’s (2012) absolute plate motions and accounts for shortening and subduction erosion at the trench. The variable amount of shortening is based on published balanced cross sections as listed in Tables 1–2, and the amount of subduction erosion is taken from Oncken et al. (2006, 2013). Prior to the onset of Cordillera growth, we also consider an extensional episode related to the Salta rift and the localized compressional episode related to the Peruvian phase (e.g., Kley and Monaldi, 2002; Mora et al., 2009).

The result of the restoration is shown in Fig. 2. It is comparable with previous reconstructions (Sdrolias and Müller, 2006; Ren et al., 2007; Husson et al., 2012; Facenna et al., 2013), but here we include a more detailed data set for shortening and also consider subduction erosion. Our results show that the trench moves westward, following the drift of South America and, from 50 Ma onward, increased its curvature in the central portion, during the build up of the Altiplano.

The position of the trench is plotted on top of a P-wave tomography model in Fig. 2 (Li et al., 2008). The model illustrates the presence of a high velocity anomaly running parallel to the trench, but shifted eastward beneath Amazonia. This feature is well defined in the lower mantle down to a depth of ~1200 km (Fig. 5), and similar features are seen in other models (e.g., Ren et al., 2007; Fukao and Obayashi, 2013; Fig. 51). Cross sections show a pronounced (>4%) ~40° dipping high velocity anomaly down to 1200 km. The high velocity anomaly is also overall continuous through the 660 km phase transition, where it gets significantly thicker. Similar features are imaged by other tomographic models (Fig. 52).

Inspection of Fig. 2 also shows that the western edge of the deep, fast velocity anomaly matches the position of the trench at ~50 Ma. This is in agreement with previous reconstructions (Husson et al., 2012; Facenna et al., 2013). Uncertainties on shortening estimates and on the absolute plate velocity model are expected to result in a maximum shift of ±10 Ma. Considering that the position of the trench at 50 Ma is to the west of the present-day anomaly and that the trench was always retreating westward, we deduce that the age of slab penetration into the lower mantle cannot be older than 50 Ma, otherwise the slab should have been reclined backward. Considering that the high velocity anomaly around the Bolivian orocline shows an almost direct continuation from the upper to the lower mantle (e.g. Figs. 5, 52; Scire et al., 2015, 2016), we also infer that any slab stagnation phase cannot be sustained for long time. Otherwise, we should observe a wider region of flat lying anomaly in the transition zone. The timing of penetration is hence likely not significantly delayed with respect to ~50 Ma. Therefore, the onset of significant shortening and crustal thickening at the surface, as well as the slowdown of trench retreat, started roughly at the same time.
The path of trench motion at 18°S illustrates that during the interval between 120 and 50 Ma the trench migrated at the same speed as South America (Fig. 3) with contemporaneous extension and compression in the upper plate (Peruvian-phase shortening in the arc domain and Salta rift extension in the back-arc). Only at ~50 Ma did the trench progressively slow down with respect to South America inducing the onset and progressive growth of the Cordillera. Those data indicate that from ~120 to ~50 Ma the subduction system offered little resistance to the westward drift of South America. The behavior of the slab/upper plate system switched to a new phase at 50 Ma. We infer that this change was due to slab penetration into the lower mantle and could have triggered significant shortening in the upper plate in the early Tertiary. A slightly older timing was proposed by van der Meer et al. (2010) for the Farallon slab.

In other words, the entrance of the slab into the lower mantle may have produced slab anchoring, slowing down its retrograde motion and inducing a vigorous, large-scale convection that may have dragged the upper plate against the trench (Yamato et al., 2009; Hussen et al., 2012; Facenna et al., 2013). Next, we explore this scenario using 2D numerical experiments.

### 4. Numerical model

We use fully dynamic, 2-D subduction models to investigate the state of stress on the upper plate during subduction, and how it evolves as a function of slab rollback rate and slab penetration into the higher viscosity lower mantle. Two models are presented here, a "free upper plate" model in which the upper plate, trailing edge is free to move relative to the side of the box, and a "fixed upper plate" model in which the trailing edge of the upper plate is stationary relative to the side of the box (Fig. 6a). In dynamic models that include both plates, fixing the upper plate reduces slab-rollback rates and increases the slab dip angle, and allowing the upper plate to be free permits rapid rollback and a reduced slab dip (e.g., Yamato et al., 2009; Capitanio et al., 2010). As the slab dip angle has been shown to exert a dominant control on slab penetrability into the lower mantle (e.g., Christensen, 1996), these two model end-members produce a range of slab-transition zone interactions.

The numerical setup is similar to that of Holt et al. (2015a), but with a domain size that is significantly larger (15,840 km wide, 2640 km deep), in order to capture the large-scale mantle flow cells hypothesized to be of importance. Even if features such as the return flow effects are expected to be somewhat emphasized in 2-D, we expect the overall behavior we discuss here to be similar in 3-D spherical settings, except directly at the slab edges.

We use the finite-element code CitcomCU to solve the equations governing convection in an incompressible (Boussinesq approximation), viscous fluid with negligible inertia and, in our case, zero internal heating (e.g., Moresi and Gurnis, 1996). The upper mantle has a reference viscosity of $2 \times 10^{20}$ Pas, temperature of 1573 K, and a thermal diffusivity of $10^{-6}$ m$^2$s$^{-1}$. For simplicity, the lithospheric plates are defined by a constant temperature as opposed to half-space cooling. Thus, we use a reduced thermal expansivity value of $1.5 \times 10^{-5}$ K$^{-1}$ to yield a reasonable lithosphere–asthenosphere density contrast of 65 kgm$^{-3}$ for the lithosphere–asthenosphere temperature contrast specified (1300 K).

We use a simplified rheology in order to focus on physical mechanisms that are of first order importance. To approximate the temperature-dependent viscosity of olivine, we use a strongly temperature-dependent viscosity and a maximum viscosity cutoff to set the subducting plate-asthenosphere viscosity contrast to a factor of 250 (see Holt et al., 2015a for more details). To prevent extreme lithospheric thickening, we increase the maximum
viscosity in the upper plate to give a lithosphere–asthenosphere viscosity contrast of 2500. The 1-D viscosity structure of the model is shown in Fig. 6b. To decouple the two plates, we insert a weak \(2 \times 10^{20} \) Pas, 15 km thick, crustal layer within the subducting plate. The crust is defined using compositional tracers, and is switched off at depths greater than 300 km, so as not to affect deep slab dynamics. To avoid stress discontinuities, the crust is tapered from full to zero thickness between depths of 150 and 300 km.

The subducting plates have an initial uniform thickness of 80 km. We additionally test variable thickness for the upper plate (80 and 150 km, see Figs. 6 and S8). Subduction is initiated by prescribing an asymmetric geometry in the trench region, which provides the initial instability required for free subduction. This initial notch has a radius of curvature of 200 km, and extends to a depth of 150 km. In the “free upper plate” model, the subducting and upper plates are 5000 and 4000 km long, respectively, and are each a distance of 3420 km from the box sides (Fig. 6a). Finite element dimensions vary from 4 km at depths shallower than 400 km and in proximity to the trench, and 8 km elsewhere. Mechanically, all boundaries are free slip. The upper surface has a constant temperature (273 K), and all other boundaries have zero heat flux.

To investigate slab penetration into the lower mantle, and its relation to upper plate stress regime, we set the lower mantle \((z > 660 \text{ km})\) viscosity to be a factor 50 greater than that of the upper mantle. Since the viscosity contrast between upper and
Fig. 6. Evolution of the horizontal deviatoric normal stress in the two reference models. Upper panels depict the model setup: (a) initial model geometry and viscosity structure (with region shown in subsequent figures indicated by grey box), (b) 1-D viscosity structure for the upper 1320 km, showing the subducting (black) and upper plate (red) viscosities. Low panels depict the horizontal stress evolution for the models with free (panels c, e, g, i) and fixed (d, f, h, j) upper plates, zoomed into a region around the trench. Overlaid on the stress field plots are mantle velocity vectors, slab outlines ($T' = 0.5$ contour), and compositional crust outlines (contour of composition = 0.5). The dynamic topography, computed from the (full) vertical normal stress at the surface ($z = z_{2D}/\rho_0 g$), is also plotted for each of the timesteps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
lower mantle is uncertain, we additionally test a stronger and weaker lower mantle, and find that the first order systematics can be seen for all of the tested viscosity contrasts (25: Fig. S5, 100: Fig. S6). For simplicity, we neglect other phase transitions and the buoyancy effects that tend to further inhibit penetration through the 660 km discontinuity.

**Results**

Fig. 6 shows how the horizontal (deviatoric) normal stress in the upper plate evolves as the slab penetrates into the strong lower mantle, for the free and fixed upper plate models. Additionally, the kinematics associated with these models, including trench migration and plate convergence rates, are presented in the supplementary material (Fig. S7). Prior to slab anchoring/penetration, the subduction-induced return flow cells are restricted to the upper mantle for both models. During this free-sinking stage, compression (>25 MPa) is restricted to relatively close to the trench for both upper plate styles (Fig. S7e). In addition to near-trench compression, the fixed model exhibits long wavelength upper plate extension at distances >~400 km.

The 2nd time panels show the stage at which the slab first comes into contact with the strong lower mantle, and the final two panels depict the subsequent stages as the slabs deform and propagate into the lower mantle. The variability in dip angle between the two models as the slab-tip touches the lower mantle, inherited from the variable rollback rates during the previous sinking phase (Fig. S7a), results in two distinct styles of interaction with the viscosity discontinuity. In the free upper plate model, the slab leans backward after touching the lower mantle, which further reduces its dip angle before it gradually sinks through the discontinuity with a near-horizontal deep slab segment. In addition, as in previous modeling studies (e.g., Zhong and Gurnis, 1994; Holt et al., 2015b), we observe a reduction in rollback/trench retreat rate during, and after, slab interaction with the mid-mantle viscosity increase.

The slab in the fixed upper mantle model anchors at a near vertical dip angle and then undergoes minor folding as it sinks into the lower mantle, retaining a dip angle that, aside from the local folding, is approximately vertical. In both models, as the slab penetrates through the viscosity jump, the lower mantle begins to participate in the return flow resulting in larger mantle flow cells with length scales >1000 km (e.g., Husson et al., 2012; Faccenna et al., 2013). The low slab dip angle in the free upper-plate model produces strongly asymmetric mantle flow, with large-scale return flow focused on the wedge-side of the slab. In contrast, the near vertical dip angle in the fixed case results in flow cells of comparable size and vigor either side of the subducting slab.

After slab penetration, both models exhibit a similar, near-linear increase in maximum compressive stress (Fig. S7d). However, the horizontal length of the upper plate undergoing significant compression differs significantly in the two models. In the free upper plate case, the length of the upper plate with compressive stresses >25 MPa increases rapidly after penetration at ~20 Ma (Fig. S7e), but compression remains localized near the trench in the fixed upper plate model. Similar results have been recently obtained by Yang et al. (2016) with application to SE Asia.

As proposed here for the Andes, we postulate that the decrease in trench retreat rate after slab penetration, in conjunction with the initiation of the large scale mantle flow, gives rise to the rapid increase in the horizontal extent of large compressive stresses, as the upper plate is dragged into the subducting plate hinge. The fixed upper plate model, in contrast, does not exhibit a reduction in trench retreat rate or an increase in the extent of horizontal compression. Unlike in some previous modeling studies (e.g., Guillaume et al., 2009; Gilbert et al., 2012), we do not observe periodic variations in the upper plate stress/strain rate associated with slab folding in the fixed upper plate model. This is likely due to the specific details of our model setup, notably the relatively strong and/or thick upper plate, and we would expect such an effect to occur for a model setup which would permit folding to influence stretching/rollback rates.

In addition, we examine the evolution of the subduction zone topography computed from the (full) vertical normal stress at the free slip upper boundary of the model (Fig. 6). This assumes the corresponding free surface, with zero traction would deflect according to this normal stress (e.g., Zhong and Gurnis, 1994). For both upper plate styles, we observe the negative, trench dynamic topography as well as a positive deflection of the boundary between the weak crust and the underlying lithosphere. However, beyond the forearc, the long wavelength dynamic topography in the upper plate differs strongly between the two models. In the fixed upper plate model, in addition to inducing a broad region of horizontal compression, penetration of the slab into the lower mantle induces negative dynamic topography of >1 km over distances of >1000 km.

We also explored the role of a thicker overriding plate (Figs. S8–S9), and variable lower-mantle viscosities (Figs. S5–S6), and verified the applicability of the general trends observed in both the lithospheric stress state, and subduction induced topography. Additionally, a thicker upper plate causes reduced rollback rates relative to the thin upper plate reference. Similarly, increasing (factor 100) or decreasing (factor 25) the lower/upper mantle viscosity contrast does not affect the first order model observations: A reduction in trench retreat rate, and an increase in the extent of overriding plate compression associated with slab penetration/ponding. The lower mantle strength does, however, exert a strong control on the vertical slab-sinking rate and slab dip the lower mantle (cf. Figs. S5 and S6).

**5. Discussion**

We use simplified numerical models to develop an understanding of how deep slab dynamics, particularly slab interaction with the viscous lower mantle, can affect overriding plate stress state and dynamic topography. Models show that slabs penetrate into the lower mantle within a reasonable range of viscosity stratification. The penetration is accompanied by slab thickening and folding due to the increased resistance of the lower mantle layer. The consequence of the lower mantle slab penetration is the onset of compression within the back-arc region. Models show that the surge of compression is due to a combination of decrease in trench retreat, due to slab anchoring into the lower mantle, and suction of the upper plate towards the trench related to the large scale return flow. The other surface signal is the dynamic subsidence of the upper plate, due to a large-scale downwelling, similar to that predicted by previous modeling studies (e.g., Liu et al., 2008; Flament et al., 2014, 2015; Yang et al., 2016).

Our models are not designed to investigate the specific setting and can be only broadly compared with the Nazca-South America system. However, they provide first order information on the way back-arc deforms during the progression of subduction in the deep mantle. Those general rules guide our interpretation of the subduction history beneath the Cordillera.

We select the central region of the Cordillera at 20°S to re-store the possible evolution of the subducting slab. A tomographic cross section from Li et al.’s (2008) model shows the presence of dipping slabs down to ~1200 km depth and ~40°E, and the associated slab shape gets progressively thicker in the transition zone and in the upper lower mantle (Fig. 5). Similar features have been
observed in other tomographic models (Ren et al., 2007; Fukao and Obayashi, 2013; Scire et al., 2015) and resolution tests show that the widening effect is not an artifact of the inversion (e.g., Ribe et al., 2007). A less pronounced anomaly (0.3%) spreads out at shallow lower mantle depths between 700 and 100 km further to the east. To provide an estimate of the amount of subducted material, we measured the area of the main high velocity anomalies along cross section. By dividing the area by the thickness for the oceanic lithosphere (on average around 70 km), we estimate the length of the subducted lithosphere. For example, 0.5% and 0.25% velocity perturbation will convert to ~2200 km and ~7800 km long subducted lithosphere, respectively. The expected rate of subduction material can be estimated from the South America–Nazca convergence minus by the shortening rate using the Seton et al. (2012) kinematics, for example. We deduce that the 0.5% or 0.25% anomalies account for subduction lasting for ~25 or ~80 Ma, respectively (Fig. 5c). Those numbers are similar to previous estimates (Ren et al., 2007; Quinteros and Sobolev, 2013).

Our preferred reconstruction is based on the distribution of the 0.4% high velocity anomaly, including material lying just below the 660 km discontinuity to the west of the slab (Fig. 7). This provides a record of ~60 Ma of subduction. The retro-deformation of the slab is performed considering both the amount of subduction and the trench migration. At each step we subtract the amount of subducted material as in Fig. 5c starting from the present-day configuration.

Reconstruction in Fig. 7 shows that the slab entered into the lower mantle around 50 Ma. Then, from 50 to 40 Ma, subduction was mostly accommodated by the horizontal advance of the slab tip at shallow upper mantle depth, accounting for the present-day slab, which is flat lying at a depth of 1000–1200 km. This is similar to what is observed in numerical model 1 where the slab aligns approximately horizontally at shallow upper mantle depths soon after penetration (Fig. 6c), without imposing phase transition dynamics, or any further increase in mantle viscosity within the lower mantle. The migration of the trench at this phase is very much reduced, so a large fraction of the South America westward motion is accommodated by shortening, as shown by the inversion of the Salta rift. Similarly to what is observed in our numerical model, the penetration of the slab results in compression in the back-arc region, implying the inversion of pre-existing structures. The penetration of the slab may also be associated with a large-scale episode of subsidence in the Paleogene either pre-dating, or during the early stages of, upper plate compression. This subsidence pattern and its variation in time is well defined by recent dynamic topography estimates for South America (Flament et al., 2015).

After 40 Ma, the deformation of the slab is mainly accommodated by thickening (Fig. 7d). Following our model results and previous fluid-dynamic modeling, we interpret the thickening of the slab as related to folding in the transition zone (e.g. Ribe et al., 2007; Guillaume et al., 2008). We count at least three episodes. After 15 Ma (see Table 2), deformation at the surface shifted eastward, associated to the formation of a lithospheric-crossing very weak detachment system (e.g. Oncken et al., 2013). This phase may also be associated at lithospheric depths with delamination and possibly by removal of lithospheric mantle of the South America lithosphere (Fig. 7c–b).

Following our reconstruction, we infer an alternating slab behavior, switching between penetration and then folding. Once the slab penetrates, the Nazca trench stopped retreating. This behavior can be approximately understood by the fixed upper-plate mode in numerical model 2, which favors slab folding at depth. Once into the lower mantle, the slab is anchored, stabilizing the trench against lateral migration, as observed in model 1.

Folding events at depth may explain the punctuated phase of compression in the upper plate, as proposed by Guillaume et al. (2009) on the base of simple experimental tests. It is then possible that some of the rapid pulse of deformation described for the Andes may be related to deep mantle, slab dynamics. However, we note that our numerical models do not display such episodic deformation, and so only suggest this slab folding mechanism. The application to the case of the Andes remains to be proven given the uncertainties in identifying and quantifying timing of deformation and related amplitude.

The origin of the Bolivia orocline

The Andes show a large variability of crustal deformation and deep mantle structure along strike. Shortening and crustal thickening reaches its maximum along the Bolivia orocline and decreases when moving laterally along the orocline flanks (e.g., Kley and Monaldi, 1998). A number of models have been proposed to explain this curvature, accounting for both inhomogeneity of the subducted slab (Isacks, 1988; Martinod et al., 2010), or mantle flow around stagnation points (Russo and Silver, 1994; Schellart et al., 2007). Despite the apparent symmetry, the mechanism leading to this tectonic feature varies laterally. The cumulative shortening is indeed similar north and south of the Altiplano proper. But south of 25–30°S, the shortening started in the Miocene, at ~20 Ma and persisted up to the present-day until some 37°S decaying southwards (Fig. 4c and Tables 1 and 2). Along the northern flank, north of ~15°N, shortening started at ~45 Ma, and persisted at a very slow rate, probably with some interruptions. This difference between the northern and the southern region is relevant for our model.

The behavior south of 25–30°S provides support of the lower mantle slab anchoring model. In this region, the slab is apparently not anchored in the lower mantle, as the high seismic velocity anomaly is confined to the upper mantle (Fig. 4; Fig. 5b). Therefore, shortening appears to have started later and is confined to the tip of the crustal wedge. The reason for why the slab did not penetrate into the lower mantle here is intriguing. Inspecting seafloor ages (Seton et al., 2012) the age of the seafloor appears to always have been rather young along this portion of the trench. The average age of subducting seafloor beneath the Nazca trench during the last 100 Myr decreases from ~60 Ma to ~38 Ma and ~25 Ma moving southward from 15°S to 30°S and to 45°S, respectively. This would have probably reduced the effective negative buoyancy and strength of the subducted slab along the southern portion of the trench, and hence its capability to penetrate into the lower mantle.

The behavior north of 15°S suggests that also there shortening of the crust initiated at the time of slab penetration into the lower mantle (Fig. 4). But here, the anchoring process was probably not efficient enough to slow down the lateral trench migration of the subducted slab. The low shortening rate suggests that at that time the trench retreated at the same rate as South America moved. One possible explanation is that the slab stretched and broke off at depth, leading to an inefficient slab anchoring. An early episode of slab break off in this region has also been proposed on the base of structural data and paleo-stress reconstruction (Mercier et al., 1992). Because the break off should have occurred earlier in the history, it is difficult to independently assess the validity of the model. However, tomography shows a pronounced low velocity anomaly there (at middle upper mantle depth, Fukao and Obayashi, 2013; Li et al., 2008, Fig. 5a) and a slab tear (Scire et al., 2016) associated with the flat portion of the slab and the subduction of the Nazca ridge. We speculate that those features may have influenced the subduction behavior in the Oligocene. Further, we speculate that the episode of slab break off may have reduced the slab pull transmitted from the deep slab, and so facilitated an episode of shallow flat slab.
Thus, the contrasting behavior of the two regions flanking the Bolivian orocline represents a test for the applicability of the slab anchoring model. This model implies that slab penetration into the lower mantle should be a necessary condition to drive protracted crustal shortening on the upper plate, but this may be not sufficient. Perturbation of subduction, e.g. due to tearing, may alter the efficiency of the anchoring effect, disconnecting the deep slab portion from the shallow one, and facilitating backward trench migration associated with an advancing upper plate.

Timing of Slab penetration in the lower mantle

Another question that arises from our reconstruction is why the penetration occurred only at \( \sim 50 \) Ma, considering that subduction was at work for a much longer time. We cannot exclude earlier Mesozoic episodes of subduction penetration into the lower mantle. However, a global survey of deep high velocity slab anomalies does not detect any large scale feature under the South Atlantic, but only a high velocity feature in the Central Atlantic and beneath the Georgia Island covering the core mantle boundary (van der Meer et al., 2010). Considering its location, this anomaly is likely related to a pre-Jurassic Farallon subduction or a Paleo-Tethyan subduction episode (van der Meer et al., 2010).

Some tomography models (Fig. S1) show an elliptical high velocity anomaly beneath Brazil from 1600 to 1900 km (e.g., Simmons et al., 2010). From the correspondence between trench position and the depth of the anomaly, this penetration episode cannot be older than 60 Ma (Fig. S10). The shape of the anomaly and its disconnection with the long Cordillera-Farallon anomaly provides an indication, if confirmed by other tomography models, of a localized subduction episode that occurred in the early Tertiary in correspondence to the present-day Bolivian orocline.

The lack of a large, deep, fast velocity anomaly related to Mesozoic episode of slab penetration into the lower mantle, such as the one detected beneath North America, beneath the South Atlantic can be explained by the young seafloor age of the oceanic lithosphere at the trench and by long-lasting episodes of ridge subduction (Seton et al., 2012), which may have segmented the subducting structure. In other words, it could be possible that only in the Tertiary the subducting oceanic floor was old enough, at least in its northern portion, to produce enough negative buoyancy to penetrate through the upper/lower mantle discontinuity and to produce a detectable velocity anomaly. The average age of subducting seafloor during the 100–60 Ma interval can be inferred to be 55, 21 and 11 Ma based on Seton et al. (2012), moving southward from 15°S to 30°S and to 45°S, respectively. In the Tertiary, the age of seafloor was almost double that, 65, 53 and 35 Ma moving southward from 15°S to 30°S and to 45°S, respectively.

An alternative explanation is related to the critical size or aspect ratio of the convection cell, to favor large-scale, whole mantle convection (Husson et al., 2012). The South Atlantic cell widened since the breakup of the South Atlantic from an aspect ratio of \( \sim 1 \) (width of the South American continent over mantle thickness), to the actual one, which is 2–3 (Husson et al., 2012). This aspect ratio could have prevailed for the last \( \sim 50 \) Ma, at the time that the slab remained anchored into the mantle (Fig. 7).

Implication for the build up of the Altiplano

High plateaux are the extreme consequence of crustal thickening. But their origin remains enigmatic. Some models propose that plateau formation is intimately related to the dripping of thickened mantle lithosphere. However, the potential to thicken the mantle lithosphere by pure shear may be inhibited by its negative buoyancy and by the presence of a weak layer in the lower crust, which may instead produce delamination. Recent models have indeed suggested that crustal thickening may be favored by the delamina-

---

**Fig. 7.** (a) Interpreted tomographic section and (b, c, d, e, f) reconstruction of subduction system along a 20°S section as in Fig. 5. Kinematics parameters for trench migration and plate motion are shown in Figs. 2 and 3.
tion or removal of mantle lithosphere (Sobolev and Babeyko, 2005; De Celles et al., 2009). The correspondence between the present-day extent of the Altipiano and the geometry of the Salta rift structure may alternatively suggest that mantle lithosphere was thinned by the previous extensional phase (Oncken et al., 2006; Armijo et al., 2015).

Our model does not consider lithospheric rheology or chemical buoyancy in detail, including the way deformation may be localized on the upper plate. The role of pre-existing structural inheritance of the crust and lithosphere and the role of surface process are of critical importance on the style of orogeny. Even if those elements are not included, our simplified model sets up the stress boundary conditions necessary to build up a plateau. In our scenario, plateaus represent an end member of a different style of crustal thickening involving subduction on one side. Subduction may be accompanied by delamination on the opposite side (Sobolev and Babeyko, 2005; Armijo et al., 2015). This double sided downwelling (subduction and delamination) may be favored by the entry of the slab into the lower mantle, generating a protracted, large-scale flow dragging the upper plate and subducting plate one against the other stabilizing the trench. Following the nomenclature of Conrad and Lithgow-Bertelloni (2002), we define this process as “slab suction orogeny” (Fig. 8; Faccenna et al., 2013), implying symmetric forcing on the overriding plate. This mechanism contrasts with “slab pull orogeny”, where upper mantle subduction may favor accretion of crustal slices forming a crustal wedge, due to asymmetric force transmission. The transition between those two styles of orogeny is thus ultimately controlled by the capability of the slab to penetrate into the lower mantle. A similar mechanism, i.e. slab penetration into the lower mantle, has been recently proposed for SE Asia to explain a recent surge of compression (Yang et al., 2016).

In the Cordillera, this transition occurred in the late Cretaceous in North America and in the Eocene in South America due to penetration of slab into the lower mantle. Differences in the rheology and structure of the upper and subducting plate are not included here and may further modulate the tectonic evolution of the area. More sophisticated models are therefore required to couple the lithospheric and crustal structure to mantle evolution.

6. Conclusion

We analyze the kinematics of the Nazca subduction system and compare it with tomography to understand why Andean Cordillera main orogenic phase started only at ~50 Ma, after a long phase of westward drift of South America. From the reconstruction, it emerges that the entry of the slab into the lower mantle corresponds to the onset of shortening. Numerical models show that the entry of the slab into the lower mantle may generate a surge of compression in the upper plate because trench migration slows down after anchoring, and because the upper plate is dragged against the subduction zone by a large-scale return flow. This model is able to solve the apparent “paradox” of Cordillera subduction orogeny, and mantle dynamics may represent the controlling factor for the deformation of the back-arc region and mountain building.

Acknowledgements

We thank Jean Pierre Brun, Barbara Carrapa, Peter De Celles, and Sean Willett for constructive discussions and Laurent Husson, who first explored the connection between deep slab dynamics and the formation of the Andes. We thank the editor and the reviewers for constructive criticisms. We thank Trond Torsvik for sharing his latest plate velocity model. This work has been written during sabbatical leaves at GFZ Potsdam to CF and TWB supported by Humboldt Foundation Fellowships.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.epsl.2017.01.041.

References


