Viscous heating allows thrusting to overcome crustal-scale buckling: Numerical investigation with application to the Himalayan syntaxes

J.-P. Burg *, S.M. Schmalholz

Geologisches Institut, ETH-Zentrum, Leonhardstrasse 19 LEB, CH 8092 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 11 December 2007
Received in revised form 11 July 2008
Accepted 15 July 2008
Available online 25 July 2008

Editor: C.P. Jaupart

Keywords:
folding
thrusting
viscous heating
Himalaya syntaxes
modelling

ABSTRACT

The eastern and western Himalayan syntaxes are crustal-scale antiforms whose metamorphic evolution is coeval in the India-Asia collisional history. Finite element modelling of lithospheric shortening and the resulting crustal-scale folds have previously been used to debate the structural conditions for the development of such structures. However, numerical models could not account for the metamorphic evolution associated with folding. In continuation of previous work, we present two-dimensional finite element modelling of lithospheric shortening in which thermal effects are implemented. The new models are consistent with earlier interpretations that crustal folding is the plausible mechanism that shaped the two Himalayan syntaxes. They further lend support to the concept that lithospheric buckling is a basic response to large-scale continental shortening and an efficient mountain building process. Introducing viscous shear heating shows, however, that the buckling mode shifts to a thrusting mode within few % shortening, depending on the initial thermal/rheological structure of the deforming lithosphere. The change to the thrusting mode prohibits further fold amplification and lateral fold propagation. This may explain why, in continental lithosphere, crustal-scale folds are isolated whereas regular, periodic crustal fold trains are rare. Focusing deformation on through-limbs thrust zones accompanies the establishment of inverted metamorphic gradients. These results offer new working hypotheses on how large thrusts like the Himalaya Main Central Thrust nucleate. Results further show that the thermal structure of the lithosphere strongly controls three fundamental deformation and metamorphic modes: (1) a cold lithosphere mainly deforms by thermally activated thrusting and exhibits large areas with significant tectonic overpressures (twice lithostatic); (2) a warm lithosphere is essentially buckled and significant tectonic overpressure builds up in the upper crust and (3) a hot lithosphere tends to thicken homogenously and mainly records lithostatic pressures below the upper crust.

1. Introduction

Surface processes are attracting much attention because they control active and relict landscapes and interact with regional climatic conditions (Molnar and England, 1990). Yet, the topographic evolution of any region results from the complex interplay between internal (i.e. processes below the Earth’s surface such as crustal shortening) and external (i.e. processes on the Earth’s surface or within the atmosphere such as erosion) geodynamic processes. Internal processes, i.e., rock mass movements, control three main surface behaviours on a regional scale: (1) homogeneous surface uplift that leads to plateaus is due to continuous crustal thickening (by magmatic underplating, (e.g. Maclellan and Lovell, 2002) and/or diffuse deformation, (e.g. Dewey and Bird, 1970)) or denudation isostatic rebound (e.g. Emery and Aubrey, 1985); (2) homoclinal tilt is due to vast isostatic flexures such as rift shoulders (e.g. Van der Beek et al., 1994) or to the steady-state response to tectonic mass flux in accretionary wedges (Davis et al., 1983); and (3) localised surface uplift or subsidence atop plume heads (e.g. Saunders et al., 2007) and small intrusions (e.g. Delcaillau et al., 2006; Schmalholz et al., 2002). Low-temperature thermochronology (e.g. Gallagher et al., 1998) and cosmogenic isotope analysis (e.g. Gosse and Phillips, 2001) have been successful in specifying the time and space scales characterizing rock cooling/exhumation and absolute surface dating during any type (plateau, tilt, localised) of surface behaviour. Exhumation and surface dating are keys only to the vertical component of rock mass transfer. However, this component is only a partial picture of the total rock flux, which may also involve a horizontal tectonic component. Computer-based models are certainly the most promising tool to provide an integrated and coherent visualization of combined rock movements and surface processes. Following this line of thought, this work focuses on the tectonic behaviour of crustal rocks in large-scale anticlines because the growth of such structures has a strong impact on
geomorphology and exhumation processes. Our specific case study is the Namche Barwa Syntaxis, at the eastern end of the Himalaya (e.g. Zheng and Chang, 1979; Burg et al., 1997, 1998; Zeitler et al., 2001b).

We first address some methodological issues and pay attention to the influence of viscous heating, a physical process inherent to rock deformation. Indeed, the length- and time-scales of crustal and lithospheric folds place them in a structural field where viscous heating predominates during deformation (Burg and Gerya, 2005; Hobbs et al., 2007), and this thermal effect should have major consequences on both the rheological evolution of the folded domain (e.g. Ranalli and Murphy, 1987; Ranalli, 1995; Kaus and Podladchikov, 2006) and the subsequent metamorphic regional gradient. The metamorphic consequences are not trivial since core migmatites and symmetrally distributed and outward increasing metamorphic ages are one of the signatures of large crustal folds (Burg et al., 1997, 1998; Zeitler et al., 2001b). In addition, a better understanding of the thermal history across crustal domes should influence the exhumation rates deduced from theo-geochronological measurements (e.g. Whittington, 1996). The thermal consideration is therefore an essential addition to the previous mechanical models (Burg and Podladchikov, 1999, 2000) that set the starting stage of the modelling presented here. We restrict our discussion to the same two-dimensional model lithospheres with the same parameters as in the referenced works in order to focus on the rheological variations that thermo-mechanical simulations reveal in course of buckling, and discuss their geological implications in terms of deep-seated influence of crustal and mantle rock movements and metamorphic history. These new models show that the buckling mode turns to a thrusting mode within few % shortening, depending on the initial thermal/rheological structure of the deforming lithosphere, because syn-folding viscous heating strongly localises deformation on through-limbs thrust zones. Movement of such crustal-scale thrusts accompanies the establishment of inverted metamorphic gradients. This result offers new working hypotheses on how the crustal-scale Main Central Thrust has nucleated to form the Himalaya and addresses the question as to whether the two Himalayan syntaxes are “fault-propagation folds” at both extremities of the Main Central Thrust.

![Lithosphere model used for numerical experiments. Additional information in Table 1.](image)

The models also show that lithospheric shortening takes place according to three fundamental deformation modes: (1) homogeneous thickening, (2) folding and (3) thrusting. The absolute value of the vertical strain rate equals the horizontal shortening strain rate during homogeneous thickening, whilst the vertical strain rate can be significantly larger than the horizontal strain rate and the internal deformation of mechanically strong layers is minimal (e.g., a strong upper mantle maintains its initial thickness) during buckling. Generally, thickening and folding are coeval but, depending on the physical parameters, one of the two modes dominates (Schmalholz et al., 2002). The shortening rates and the initial rheological and thermal structure of the lithosphere, which also influences tectonic

**Table 1**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_1 ) (°C)</td>
<td>1200</td>
</tr>
<tr>
<td>( T_2 ) (°C)</td>
<td>1230</td>
</tr>
<tr>
<td>Erosion</td>
<td>No</td>
</tr>
<tr>
<td>NAME</td>
<td>WARM</td>
</tr>
<tr>
<td>Model name</td>
<td>WARM</td>
</tr>
<tr>
<td>Model name</td>
<td>WARMN</td>
</tr>
<tr>
<td>Model name</td>
<td>COLD</td>
</tr>
<tr>
<td>Model name</td>
<td>HOT</td>
</tr>
<tr>
<td>Model name</td>
<td>INTERMEDIATE</td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>Model name</th>
<th>( T_1 ) (°C)</th>
<th>( T_2 ) (°C)</th>
<th>Erosion</th>
</tr>
</thead>
<tbody>
<tr>
<td>WARM</td>
<td>1200</td>
<td>1230</td>
<td>No</td>
</tr>
<tr>
<td>WARMN</td>
<td>1200</td>
<td>1230</td>
<td>No</td>
</tr>
<tr>
<td>COLD</td>
<td>900</td>
<td>930</td>
<td>No</td>
</tr>
<tr>
<td>HOT</td>
<td>1300</td>
<td>1310</td>
<td>No</td>
</tr>
<tr>
<td>INTERMEDIATE</td>
<td>1000</td>
<td>1030</td>
<td>No</td>
</tr>
</tbody>
</table>

---

overpressure, rule the deformation mode. Homogeneous thickening occurs in hot lithosphere and significant tectonic overpressure due to the brittle strength of rocks is only present in the upper crust. Folding occurs in warm lithosphere and significant tectonic overpressure sets up in both the upper crust and the upper mantle. Thrusting arises in cold lithosphere and significant tectonic overpressure dominates in both the crust and the upper mantle.

2. Modelling

2.1. Method

The applied mathematical model is described in the Appendix A; equations were solved numerically. We employed a two-dimensional finite element code that couples plane strain mechanical and thermal calculations. We use a deforming Lagrangian mesh and apply a mixed finite element model using 9-node biquadratic quadrilateral elements with a discontinuous, linear pressure approximation (so-called Q2-P1 element; e.g. Bathe, 1996). The pressure variables are statically compensated on the element level to reduce the number of unknowns. The temperature is approximated with the same 9-node elements as the velocities and is solved simultaneously with the velocities. The material is considered incompressible and Uzawa-type iterations are applied to achieve incompressibility. The robustness of the model has been verified for buckle-folding (Frehner and Schmalholz, 2006) and necking in Newtonian and power-law materials. The effects of gravity, thermal and compositional buoyancy and viscosity dependencies are included. Plasticity is implemented by iteratively modifying the stresses so that stresses remain either on the Mohr–Coulomb or von Mises yield function (Burg and Podladchikov, 1999). The dilatancy angle is zero. This implementation provides correct plastic shear bands and fulfils both the Mohr–Coulomb and von Mises failure criteria. Plastically deforming areas are continuously filled with distributed plastic shear bands whose individual thickness is controlled by spacing of the numerical grid if no regularization technique is applied. Here, we do not resolve the distributed plastic shear bands numerically because we want to study localized shear deformation. Localized large strain on a single plastic shear band can usually be achieved only by employing some kind of a priori defined strain softening where a material parameter, such as cohesion, is reduced with increasing strain. We do not employ strain softening because we want to study localized shear deformation caused by viscous heating, which is a direct consequence of the conservation of energy and the temperature dependence of the rock viscosities. The resolution of finite-element grids was 301 (horizontal) × 151 (vertical). Consistency of the results has been checked with higher and smaller resolutions. 600 to 1200 time steps of 10 to 30 thousand years each were applied, corresponding to durations of 6 to 36 million years, respectively.

2.2. Model set-up and boundary conditions

The set-up (Fig. 1), its rationale and the material parameters (Table 1) are based on those presented in (Burg and Podladchikov, 1999, 2000). A stratified box has three compositional layers, 1) upper granitic crust (25 km), 2) lower diabase crust (10 km) and 3) subcrustal olivine lithosphere (85 km). The mechanical boundary conditions consist of a bottom boundary, fixed in the vertical direction and free to slip in the horizontal direction; lateral boundaries are free to slip in the vertical direction; and the upper boundary is a free surface. At variance to the setting employed by Burg and Podladchikov (1999, 2000), the new models used wet instead of a dry olivine to avoid excessively high stresses in the mantle and a constant shortening strain rate instead of constant shortening velocity. This does not affect the results because the applied amount of shortening is small (<30%) and, therefore, the shortening strain rate does not vary significantly by using a constant shortening velocity. Furthermore, the boundaries with depth-constant shortening strain rate are far away from the...
model center where the deformation of interest takes place and where strain rates can develop freely, not remaining constant with depth. Experience has shown that the models with constant shortening strain rate are numerically more stable during the initial stages of thermal equilibration and viscoelastic stress build-up. The so-called kinematic boundary conditions applied here are frequently applied in thin-sheet models of lithospheric deformation.

The starting configuration has a relaxed stress state with gravitational load and a non-linear, steady state temperature distribution. Thermal boundary conditions are 0 °C at the surface and fixed basal temperatures, which are (i) $T_1$ beneath the left “half” of the model and (ii) $T_2$ beneath the right “half” of the model (Fig. 1, Table 2). These simulate the orogenic lithosphere to the north against the Indian lithosphere, to the south, in the Himalayan case study of orogenic syntaxes (Burg and Podladchikov, 1999, 2000). In the middle of the model domain, where $T_1$ and $T_2$ meet, the Moho is elevated by 250 m, at the apex of a 30 km wide, Gaussian-shaped indentation. This geometrical perturbation is preferred to the thermal perturbation used in (Burg and Podladchikov, 1999) to nucleate folding within a smaller amount of shortening. The type of this small initial perturbation does not affect the general results (see section Sensitivity studies). The temperature differences at the bottom of the models imply two initial stress profiles, one for each temperature zone. Respective values of $T_1$ and $T_2$ designate the five numerical models presented here (Table 2). There is no lateral heat flux through the sides.

At each time step the vertical position of the entire model box is adjusted (i.e. shifted downwards by constant displacement) to keep the far-field surface elevation at zero level. The technique is justified because the lithosphere is not deforming at distance from the Himalayas and, therefore, stands for the relative topographic elevation at isostatic equilibrium. This approach allows the adoption of a kinematic (vertically fixed) basal boundary condition as opposed to a more complex, though strictly valid, isostatic condition of vanishing differential stresses at some compensation level. For comparison, simulations have been performed with a 160 km thick lithosphere in which the total vertical stresses did hardly vary laterally at its base and isostatic equilibrium was achieved within the model domain. The first-order results were the same (see section Sensitivity studies). Our experience has shown that using the isostatic lower boundary conditions requires smaller time steps but produces similar results.

**Fig. 3.** Buckled COLD and HOT lithospheres without versus with viscous heating for similar amounts of shortening. The amplitudes of the folds within the lithosphere are comparable in both models, but the HOT lithosphere had to be shortened for 10% more than the COLD lithosphere to yield about the same fold amplitudes. At 20% shortening the folds within the HOT lithosphere are hardly visible. The displayed grid is not the numerical grid, which is much finer. The convex upward deflection of the Moho at about $x = 100$ km of the HOT model is an effect of weakening due to viscous heating (see Fig. 6).
2.3. Rheological laws

Experimental rheological laws for dry granite, diabase and wet olivine were extrapolated to tectonic time scales (Ranalli, 1995; Carter and Tsen, 1987; Afonso et al., 2007) and the parameter values (Table 1) are typical of those adopted for lithospheric scale modelling. A visco–elasto–plastic rheology is applied. The Maxwell model is used for the viscoelastic rheology and the upper convected time derivative is used to warrant rheological objectivity (see Appendix A, Schmalholz et al., 2001). For the ductile deformation we only apply a power-law creep rheology, because this is the rheology that was used in (Burg and Podladchikov, 1999, 2000). A high stress rheology including the Peierls mechanism is not applied.

A critical question in understanding large-scale deformation of the continental lithosphere is where strength resides at any deformation stage. For this purpose, rheological profiles are generated continuously, throughout the modelling. The position of the brittle–ductile transition for each compositional layer is free to evolve both in depth and magnitude following crustal thickening, the corresponding temperature changes at the base of the crust and the viscoelastic stress build up (Fig. 2). The stress profiles and the stress build up are considerably different for initial shortening of the COLD and HOT models (Fig. 2). The overall deformation during these initial stages is close to pure shear and vertical gradients in horizontal strain rates are small in the two models. The applied Mohr–Coulomb failure criterion provides yield stresses that are smaller than the maximal strength values predicted by Byerlee’s rule (e.g. Kohlstedt et al., 1995). This application is a conservative approach that avoids using the stress magnitudes that are at the maximal limit of experimentally and theoretically defined yield stresses. The reduced maximal strength values could be considered to be the effect of pore fluid pressures on Byerlee’s rules. Indeed, pore pressure can strongly reduce the slope of the yield strength versus depth curve (e.g. Kohlstedt et al., 1995). For sensitivity studies, the von Mises failure criterion was used for the mantle. This criterion can be modelled with the Mohr–Coulomb criterion by setting the friction angle to zero and the cohesion to the von Mises yield strength.

2.4. Viscous heating

Viscous heating was implemented as described in Burg and Gerya (2005). Only the dissipative part of the deformation (i.e. viscous and plastic) can contribute to the viscous heating while the elastic energy is stored and not available for heat generation (e.g. Kaus and Podladchikov, 2006). In most simulations, only the viscous strain rate is used to calculate viscous heating and the plastic strain rate is neglected (see Appendix A). This is a conservative approach because adding the plastic strain rate can only increase the amount of viscous heating.

Fig. 4. Vertical displacement of topography and Moho surfaces of buckled COLD and HOT lithospheres (model result in Fig. 3). The vertical coordinates of the topography and Moho surface have been adjusted to be at zero at the left of both graphs. The fold wavelengths within the COLD lithosphere are about twice larger than the fold wavelengths within the HOT lithosphere.
Fig. 5. Buckled WARM and INTERMEDIATE lithospheres with viscous heating displaying initiation of viscous heating shear bands and buckled WARM lithosphere with erosion (WARMe in Table 2). Erosion triggers thrusting at an earlier stage than without erosion. The applied erosion coefficient $E_c$ keeps the free surface nearly flat, representing an end-member scenario with very strong erosion. As in Fig. 3, the displayed grid is much larger than the numerical grid.

Fig. 6. Viscous heating in the four lithospheres under consideration (Tables 1 and 2).
Furthermore, the plastic strain rates in shear bands are most sensitive to the applied numerical technique and resolution because they depend on the width of the plastic shear band. The impact of plastic strain rates on the effect of viscous heating is discussed in the section on sensitivity studies. The viscous heating term was implemented explicitly and solved iteratively within the same iteration loop as that applied to solve for the strain rate dependent power–law viscosity.

2.5. Erosion

Erosion and redeposition of surface sediments are numerically simulated with a diffusion equation applied to the top, free surface of the model (e.g. Burg and Podladchikov, 2000; Podladchikov et al., 1993). The erosion coefficient $E_c$ would correspond to the thermal diffusivity if diffusion is applied to a thermal problem. The diffusion equation is solved at every time step for the free model surface. We did not apply more sophisticated and realistic erosion models because we want to show the fundamental impact of the largest-scale erosion.

3. Numerical results

We first checked the consistency of the new thermo-mechanical code, having constant strain rate boundary conditions, with constant shortening velocity models obtained earlier where thermal coupling was only done through temperature dependent viscosities. For this purpose we compared the two extreme, thermo-mechanical cases without viscous heating: the cold lithosphere (i.e. stronger mantle) and the hot lithosphere (i.e. weaker mantle) of Burg and Podladchikov (1999) with results from the new code using similar cold and hot lithospheres without viscous heating and without erosion. Comparison of these two situations allows further discussion of the dynamics of lithospheric shortening and buckling.

3.1. Dynamics of folding without viscous heating

The COLD and HOT models in their new versions (Fig. 3), like in their older one (Burg and Podladchikov, 1999), demonstrate that employing constant strain rate instead of constant lateral velocity does not affect the primary response of the plates to applied shortening. In both cases, ultimately unstable homogeneous thickening of the lithospheres (up to 10 to 15%) leads to buckling. Old and new versions further emphasise the effects of initial conditions. 1) Similar wavelength and amplitudes are obtained for the same amounts of bulk shortening, and the buckle wavelength (ca. 300–400 km versus ca. 100–200 km) and amplitude (ca. 25 vs. ca. 5 km) are larger on the cold lithosphere (Figs. 3 and 4). 2) The differential topography grows fast and exponentially, as has been demonstrated in all previous buckling experiments (e.g. Ramberg, 1964; Ramsay, 1974; Johnson and Fletcher, 1994) and more recently (Schmalholz et al., 2002) and the final topography of the cold type models is unacceptably high (Fig. 4). 3) Lateral propagation of buckling is more manifest in cold than in hot lithosphere (Fig. 4). Folds that develop at the surface and at the Moho have the same wavelength and are in phase in

Fig. 7. Viscous heating established at an early stage of the shortened/buckled COLD lithosphere. The three graphs focus around the model center, where the main fold develops. The dashed white line is the boundary between upper and lower crust. Vertical axes are exaggerated. A): Viscous heating, $H$ (W/m$^3$), is strongest in the compressive part of the mantle fold hinge. B): Variation by more than one order of magnitude of the second invariant of the strain rate, $\varepsilon_{II}$ (1/s) within the displayed lithospheric region. C): The second invariant of the deviatoric stresses, $\tau_{II}$ (MPa), varies mainly in the vertical direction. The distribution of the product of $\varepsilon_{II}$ times $\tau_{II}$ explains well the distribution of $H$, although $H$ does not exactly correspond to this product.
both COLD and HOT lithospheres (Fig. 4). 4) Synclinal basins between anticlines on the HOT lithosphere subside relatively less than on COLD lithosphere (Fig. 4).

The evolutions and shapes of the rheological profiles for the COLD and HOT lithospheres are considerably different (Fig. 2). The effective thickness of the crustal and mantle layers is also considerably larger for the COLD than for the HOT lithosphere. The COLD yield strength profile has small ductile indentations that remain minor during stress build-up. This indicates almost no decoupling between rheological layers of such lithospheres. Conversely, the upper crust, lower crust and upper mantle are clearly separated and can thus be decoupled in the HOT lithosphere.

3.2. Dynamics of folding with viscous heating

Unless stated otherwise, only viscous strain rates have been used to calculate the amount of viscous heating in the simulations (see Appendix A). Experiments with viscous heating applied to HOT lithospheres (Fig. 3) produce only small amplitude buckle folds (Fig. 4). Also, folding of the crust is decoupled from mantle folding so that the topography wavelength is smaller than the wavelength of the mantle folds (Fig. 4). The viscous heating localized the deformation generating one large, synclinal buckle-fold in the mantle. The resulting crustal thickening is isostatically compensated with an elevated topography showing the mechanical decoupling of the crust and mantle lithosphere (Fig. 4). Results with viscous heating in COLD lithospheres spectacularly and fundamentally differ from results without viscous heating (Fig. 3).

Viscous heating and the corresponding decrease in effective viscosity trigger within 2% shortening and quickly amplify the development of one main shear zone, to the point that the code would not permit more than 16% shortening (Fig. 3). The four models HOT, WARM, INTER-MEDIATE and COLD show a transition from thickening through folding to lithospheric scale shearing/thrusting, with the INTERMEDIATE lithospheres displaying even two conjugate shear zones before the code failed at 15%, due to extreme deformation of some elements (Fig. 5). In all cases, the shear zones localize deformation and remain at the same position, at inflexion points of buckle limbs. From a modelling point of view, it is worth mentioning that localized and geometrically stable shear zones are usually obtained with some kind of strain softening model (e.g., Huisman et al., 2005) because standard plasticity tends to fill a homogenous model domain with many equally spaced shear zones. At the scale of the models presented here, small-scale plastic shear bands do not have to be resolved numerically and the plasticity model using the Mohr–Coulomb failure criterion is employed to limit the maximal differential stresses. From a geological point of view, these results indicate that viscous heating is essential in transferring and focusing deformation from a distributed buckling mode to a localized shear zone mode and that the buckle wavelength controls the spacing between major thrusts.

Shorter wavelength, i.e. more localized structures, develop with growing bulk strain and are controlled by the instantaneous rheological configuration of the lithosphere. The asymmetry reflects an inherent physical process triggered by differences in basal temperature of the shortened lithosphere.

3.3. Pressure–temperature evolution

Folding controls the distribution of viscous heating and high temperatures are generated in the compressive parts of the fold hinges (Figs. 6 and 7) where viscous heating shear zones nucleate. Viscous heating actually emphasizes and maintains major shear zones at the main fold hinges (Fig. 6), which is strongly seen for the COLD
The spatial distribution of viscous heating can be explained by the distributions of both stresses and strain rates (Fig. 7). Whereas stresses vary mainly in the vertical direction, the strain rates vary in both vertical and horizontal direction by more than one order of magnitude. The significant strain rate variations are caused by folding even if fold amplitudes remain very small (Fig. 7). Within a few % bulk shortening, the temperature field shows inverted isotherms nearly parallel to the shear zones, with temperatures over 800 °C overlying zones at <400 °C in the COLD lithosphere (Fig. 8).

The dynamic pressure deviates considerably from the lithostatic pressure around these shear zones (Fig. 9) where the differential stresses are high, which is mainly in the upper parts of the lithospheric mantle (Fig. 2). In the mechanically strong parts of the lithosphere, the dynamic pressure becomes about twice the lithostatic pressure (Fig. 9), as expected from mechanical force equilibrium (Petrini and Podladchikov, 2000). Close to the surface, the dynamic pressure can exceed the lithostatic pressure by a factor of three or more. Yet, it is questionable whether close to the surface the employed rheologies are applicable. We conclude, however, that inverted metamorphic gradients affected by tectonic overpressure should exist in buckling continental lithospheres, in particular if they belong to cold lithospheres. The tectonic overpressure becomes significantly smaller for a hotter thermal structure of the lithosphere (Fig. 9). The HOT lithosphere shows significant tectonic overpressure only in the upper crust and in a thin layer within the lithospheric mantle, while the major part of the lithosphere exhibits lithostatic pressure values (Fig. 9). The dynamic pressure field is relatively homogenous for the HOT lithosphere but becomes increasingly heterogeneous towards the COLD lithosphere (Fig. 10).

3.4. Effects of erosion

The effects of erosion were tested on the WARM lithosphere with viscous heating (Fig. 5), which was the most satisfying result in terms of regional relevance. One can readily see that erosion favours the development of thrusts cutting limbs at an earlier stage than without erosion. Erosion is therefore an important agent in triggering a thrust-dominated, more brittle-like mode of lithospheric shortening at the expenses of a distributed thickening or buckling mode. The erosion particularly accelerates the amplification of folding within the WARM lithosphere locally and consequently increases local strain rates and viscous heating. The trigger for the localization of viscous heating is always the heterogeneous stress and non-elastic strain rate distribution due to folding. However, the erosion modifies the folding behaviour of the lithosphere and therefore indirectly affects the timing and localization of viscous heating shear bands. This result is in line with the idea that erosion impacts the structural behaviour of deforming zones (e.g. Beaumont et al., 1992; Summerfield, 1999; Zeitler et al., 2001a). Yet, tectonics remains the main actor.

3.5. Sensitivity studies

The material parameters, rheological laws and boundary conditions applied in this study were chosen to closely match the simulations.
presented in (Burg and Podladchikov, 1999, 2000) in order to assess the impact of viscous heating on lithospheric shortening and buckling. Several sensitivity studies with different rheologies and boundary conditions have been performed to evaluate the significance of the presented simulations. The temperature distribution (Figs. 11 and 12) introduces results of the sensitivity studies. The initial geometrical perturbation has no effect on results (Fig. 11). On the other hand, using both the viscous and plastic strain rates strongly affects results for the COLD lithosphere (Fig. 11). The plastic strain rates cause the development of two conjugate thrusts for the Mohr–Coulomb failure criterion. However, if a von Mises failure criterion with a yield strength of 1250 MPa is used, only one thrust develops with the same orientation as in the simulation with only ductile strain rates but Mohr–Coulomb failure criterion. The reason is presumably the stronger localization in materials with a Mohr–Coulomb rheology than in materials with a von Mises failure criterion. The high topography in the COLD simulations is partly supported by the rigid base. Simulations with an initially 160 km thick WARM lithosphere essentially generate the same result than the COLD lithosphere with 120 km initial thickness (Fig. 11), but the topographic support of the rigid base of the 160 km thick lithosphere is negligible because the maximal differential stresses along this base are everywhere smaller than 1.5 MPa. In contrast, maximal differential stresses for the COLD lithosphere with 120 km initial thickness locally reach 12 MPa along the rigid base. These stress magnitudes impel a deviation from the vertical total stress on the rigid base at isostatic equilibrium. This deviation is about 14% for the COLD lithosphere and only 2% for the WARM lithosphere. The latter figure is quite acceptable.

Using both viscous and plastic strain rates to calculate viscous heating or using a von Mises failure criterion has no impact on the main results for the HOT lithosphere (Fig. 12). However, a ten times larger boundary shortening rate has a major impact because then large-scale shear zones also develop in the HOT lithosphere (Fig. 12). As for the COLD lithosphere, the orientation of the shear zones depends on the failure criterion and the applied strain rates used to calculate viscous heating (Fig. 12).

The isotherms in Figs. 8, 11 and 12 indicate that all shear zones have a similar thickness of about 10 km. This exact thickness is difficult to measure because there is a continuous transition of strain rates, stresses and temperatures across the shear zones. The declared 10 km correspond to the characteristic thermal length which is proportional to \( \sqrt{\kappa/\varepsilon} \) (e.g., Burg and Gerya, 2005; Hobbs et al., 2007) where \( \kappa \) is the thermal diffusivity of about \( 10^{-6} \) m²/s in all simulations and \( \varepsilon \) is the characteristic strain rate within the shear zones, between \( 10^{-13} \) and \( 10^{-14} \) 1/s for most simulations. These values provide characteristic thermal lengths between 3 and 10 km.

4. Relevance to the Himalayan “syntaxes”

4.1. Outline

The Himalaya is an active mountain belt that terminates at both ends in syntaxes (Wadia, 1931), i.e. areas where orogenic structures turn sharply about a vertical axis. These syntaxes are crustal antiforms that fold the Tethyan suture zone between India and Eurasia around half-windows of Indian crust (Burg et al., 1997; Zeitler et al., 2001b; Wadia, 1957; Gansser, 1966; Treloar et al., 1991; Ding et al., 2001). Both syntaxes record remarkably similar thermo-mechanical evolution of basement rocks overprinted by Himalayan metamorphism and
Pliocene–Pleistocene high-grade metamorphism and anatexis. Both syntaxes straddle the same Neogene time span and have undergone rapid denudation during fold amplification. Important faults and shear zones bound both antiforms; neotectonics of the western syntaxis provide evidence for reverse faulting (e.g. Butler and Prior, 1988), which is less well documented for the eastern syntaxis (Chang et al., 1992; Booth et al., 2004).

A rise of the Moho level is recorded under the western syntaxis (Farah et al., 1984). No information on the depth of the Moho is available beneath the eastern one.

Geological interpretation indicates crustal-scale folding as the mechanism that produced these orogenic structures. Kinematic studies indicate that both the western and eastern antiforms result from compression nearly orthogonal to their axial trace (e.g. Butler and Prior, 1988), which is less well documented for the eastern syntaxis (Chang et al., 1992; Booth et al., 2004).

Fig. 11. Sensitivity study for the COLD lithosphere. Reference models are those of Fig. 8. A) Model without initial geometrical perturbation at the Moho, demonstrating that the initial perturbation has no significant impact. B) Model using both viscous and plastic strain rates to calculate viscous heating. The Mohr–Coulomb failure criterion causes two conjugate plastic shear bands that overprint the single shear band of A). C) Same model as B) but with a von Mises failure criterion (yield strength=1250 MPa). The von Mises failure criterion does not yield the same intensity of plastic localization as the Mohr–Coulomb criterion and a single shear band as in A) dominates the final result. However, lateral temperature variations indicate the presence of several other shear bands. D) WARM lithosphere with an initial lithospheric thickness of 160 km. Due to the larger thickness than in Fig. 8, temperatures in the crust and upper mantle are similar to those in the COLD lithosphere. One shear zone forms with the same orientation as in A) but the topography is not supported by the rigid model base, which indicates that the model base does not affect first order results.

A WARM lithosphere (Table 2) was the mechanical-only model most compatible with the amplitude and wavelength of both Himalayan syntaxes (Burg and Podladchikov, 1995, 2000). Implementing shear heating yields results that are as pertinent as with the mechanical-only solutions and thus emphasise the primary response of crustal buckle folding. Buckling is observable after some homogeneous, distributed shortening of the lithospheric plate, which strongly depends on the magnitude of any perturbation (e.g. thermal, geometrical, material, compositional) within the lithosphere. Buckling was observable in all presented models after 8% (80 km) to 20% (200 km) shortening, with colder models buckling earlier.

Thermo-mechanical models would agree with a WARM to INTERMEDIATE lithosphere (which are both consistent with the structure and properties of the Indian lithosphere (Henry et al., 1997; Pandey and Agrawal, 1999)) and provide alternative interpretations to several points: 1) The most prominent is the establishment of viscous-heating shear zones across limb inflexions. These shear zones prevent deformation becoming dominated by new adjacent buckle hinges once the major fold reaches the locking-up stage. Although the amplitude of the buckle fold pattern is achieved in the bulk strain range of 10–25%, the intervening shear zones will absorb further shortening. 2) Asymmetric folding and thrusting is facing towards the cold parts of the lithosphere instead of the hot side in mechanical models. This is more consistent with the classical in-sequence attitude of major thrusts in orogenic systems, in particular in the Himalayas (Gansser, 1964; Hodges, 2000; Yin and Harrison, 2000). However, the orientation of thrusting strongly depends on the plastic failure criterion (Figs. 11 and 12). 3) We do not face the problem of unacceptably high topography because the crust of
WARM and INTERMEDIATE lithospheres is more decoupled from the mantle than in COLD lithosphere. Therefore, the lithospheric mantle is less able to support strong topographic gradients.

Thermo-mechanical models shed new light on the understanding of the two Himalayan syntaxes. Both are known for high thermal gradients and abundant hot springs (Chang et al., 1992; Craw et al., 1997). This information points to active metamorphism in the buckled crust, as geophysical survey of the western syntax indicates (Park and Mackie, 2000; Meltzer et al., 2001), and has led to the concept of rapid advection of isotherms while hot rocks of the deep crust flow towards regions of deep incision (Zeitler et al., 2001b). Numerical modelling presented here indicates that the thermal evolution inherent to buckling and subsequent, viscous heating shear zones produce sharply localised metamorphic anomalies. Those would account for low-pressure melting and strong geothermal gradients, with the advantage of being a simpler process than diverted crustal flow and subsequent heat advection towards rapidly eroding areas (e.g. Zeitler et al., 2001b). Folding is also more consistent with seismic tomography showing that, below the western syntax, crust flows upward from depth rather than laterally, along a shallow detachment (Meltzer et al., 2001). Following the lex parsimonyae (Occam’s razor principle), the alternative solution we propose, consistent with conjugate, somewhat symmetrical shear zones in fold limbs, may be preferable.

The simultaneous subsidence of synformal basins was an important result of mechanical-only models, with application to the synclinal Peshawar and Kashmir basins on both sides of, and directly related to, the western syntaxis. This interpretation explained a first order feature, namely their location, yet had difficulties in identifying such basins aside the eastern syntaxis. Thermo-mechanical calculation suggests that the warmer the lithosphere, the less such synforms subside, and a difference in lithospheric properties between the western and eastern extremities of the Himalayas is a plausible explanation. During folding of the lithosphere the surface develops synclines and anticlines. However, these synclines do not necessarily subside with respect to a constant sea-level and generate large depo-centers for sediments. This is due to the simultaneous thickening and folding of the crust and the thickening related isostatic uplift. Therefore, the absence of synclinal sedimentary basins next to crustal-scale anticlines does not contradict a folding origin of these anticlines.

An eye-catching feature is the spontaneous large-scale asymmetry developing once initially symmetric buckling evolves into a “thrusting” mode. The dip direction is controlled by the small difference in bottom temperatures of the colliding plates and by the applied plastic failure criterion. The spacing between major thrusts is controlled by the buckling mode. Pressure (Fig. 10) and temperature (Fig. 8) fields show strong anomalies next to the shear zones. Inverted isotherms nearly parallel to the shear zones characterize the temperature field, with temperatures over 800 °C overlying zones at <400 °C in the COLD lithosphere. The result images a metamorphic field gradient similar in size and intensity to that linked to the Main Central Thrust of the Himalayas (e.g. Le Fort, 1975; Jaupart and Provost, 1985; Harrison et al., 1998; Pêcher, 1989) and major thrusts in other orogenic systems (e.g. Graham and England, 1976; Burg et al., 1989; Burg et al., 1984; Patrick

![Fig. 12. Sensitivity study for the HOT lithosphere. Reference model in Fig. 8. A) Model using both viscous and plastic strain rates to calculate viscous heating. B) Model using both viscous and plastic strain rates to calculate viscous heating and using a von Mises failure criterion with yield strength of 1250 MPa. Both A) and B) are comparable to the reference simulation of Fig. 10 C) Model with a shortening rate of 3*10^{-15}, i.e. one order of magnitude larger than in the reference simulation. Shear zones at those in the COLD lithosphere develop. D) Model using both viscous and plastic strain rates to calculate viscous heating, using a von Mises failure criterion with yield strength=1250 MPa and with a shortening rate of 3*10^{-15}. Shear zones develop in the centre of the model with a different orientation than in C).](image-url)
reviews. The ETH-Zurich supports this work. We appreciate H. Schmeling’s and an anonymous reviewer’s in depth discussion with Y. Podladchikov and B. Kaus impact analyses to stabilize and localize deformation at crustal scales. New modelling yields exhumation rates similar to those recognised in the eastern and western Himalayan syntaxes, and evolves from a buckling to a thrusting mode, as reported from large-scale analogue modelling (Cobbold et al., 1993; Cloetingh et al., 1999; Martínd and Davy, 1994; Burg et al., 1994, 2002). Numerical modelling additionally demonstrates rheological and thermal changes in course of deformation. The results of the coupled thermo-mechanical model show that viscous heating is essential in the development of crustal-scale shear zones and that these shear zones are stable in space and time. To achieve such temporal and spatial stability of major shear zones, some kind of strain softening is usually implemented in numerical algorithms (e.g. Huisman et al., 2005). Viscous heating is intrinsically a natural process that suffices to stabilize and localize deformation at the crustal scale.

5. Discussion

Thermo-mechanical modelling presented in this paper ensures force balance in addition to the heat and mass balance, which are satisfied in thermo-kinematic models. Results support the plausibility of lithospheric folding as a primary response to tectonic shortening, as argued from analogue (Dave and Cobbold, 1991; Cobbold et al., 1993), analytical (e.g. Schmalholz et al., 2002) and numerical (Burg et al., 1997; Burg and Podladchikov, 2000; Cloetingh et al., 1999) modelling. New modelling yields exhumation rates similar to those recognised in the eastern and western Himalayan syntaxes, and evolves from a buckling to a thrusting mode, as reported from large-scale analogue modelling (Cobbold et al., 1993; Cloetingh et al., 1999; Martínd and Davy, 1994; Burg et al., 1994, 2002). Numerical modelling additionally demonstrates rheological and thermal changes in course of deformation. The results of the coupled thermo-mechanical model show that viscous heating is essential in the development of crustal-scale shear zones and that these shear zones are stable in space and time. To achieve such temporal and spatial stability of major shear zones, some kind of strain softening is usually implemented in numerical algorithms (e.g. Huisman et al., 2005). Viscous heating is intrinsically a natural process that suffices to stabilize and localize deformation at the crustal scale.

6. Conclusion

New 2D-finite element numerical experiments based on a thermo-mechanical approach are consistent with earlier interpretations that crustal folding is the plausible mechanism that shaped the two Himalayan syntaxes. They further lend support to the concept that lithospheric buckling is a basic response to large-scale continental shortening and an efficient mountain building process. Introducing shear heating shows, however, that the buckling mode shifts to a thrusting mode within a few % shortening, depending on the initial thermal/rheological structure of the deforming lithosphere. The change to the thrusting mode prohibits further fold growth and lateral fold propagation. This would explain why crustal folds like the Himalayan syntaxes are localized structures whereas periodic crustal fold trains are rare in continental lithospheres. Focusing deformation on through-limbs thrust zones accompanies the establishment of inverted metamorphic gradients. Folding of both crustal and sub-crustal levels indicates coupling of all lithospheric layers during this deformation mode. This result offers new working hypotheses on how crustal thrusts like the Himalaya Main Central Thrust nucleate and evolve. A refined answer concerning the regional application obviously needs three-dimensional modelling.

Acknowledgements

These results have been obtained thanks to the support of the ETH-Zurich. Passionate discussion with Y. Podladchikov and B. Kaus impacted our work. Paul Tackley helped in further formulating our ideas. We appreciate H. Schmeling’s and an anonymous reviewer’s in depth reviews. The ETH-Zurich supports this work.

Appendix A

We apply the concepts of continuum mechanics to mathematically describe lithospheric deformation. The equations describing conservation of linear momentum for slow flow under the effects of gravity are (symbols explained in Table 1):

\[ \frac{\partial \sigma_{ij}}{\partial x_j} = 0 \]  

(A1)

The indexes i and j stand from 1 to 2 and represent the two spatial directions (1 = horizontal and 2 = vertical direction). Repeated indexes are summed. The gravity vector g includes the gravitational acceleration (9.81 m/s²) for the vertical direction and zero for the horizontal direction. We assume that there are no couple stresses acting internally in the continuum and, therefore, conservation of angular momentum provides a symmetric stress tensor:

\[ \tau_{ij} = \sigma_{ij} \]  

(A2)

We further assume incompressible flow because bulk deformation is significantly smaller than shear deformation of rocks during lithospheric deformation. We apply the Boussinesq approximation and neglect all density variations except for densities that are multiplied by the gravitational acceleration in Eq. (A1). Conservation of mass is then described by:

\[ \frac{\partial \rho}{\partial t} = 0 \]  

(A3)

The conservation of energy provides the equation for the temperature:

\[ \rho C v(T) \left( \frac{\partial T}{\partial x_j} \right) = \frac{\partial}{\partial x_j} \left( k \frac{\partial T}{\partial x_j} \right) + H + \tau_{ij} \left( \varepsilon_{ij} + \varepsilon_{ij}^p \right) \]  

(A4)

The dot above a symbol represents the partial time derivative. The temperature dependence of density is described by:

\[ \rho = \rho_0 [1 - \alpha(T - T_0)] \]  

where T_0 is the temperature at which the reference density \rho_0 is determined. We apply the small deformation theory and assume that individual strain rate components can be summed. The total, deviatoric strain rate is then the sum of the viscous, elastic and plastic deviatoric strain rates:

\[ \dot{\varepsilon}_{ij} = \dot{\varepsilon}_{ij}^{p} + \dot{\varepsilon}_{ij}^{v} + \dot{\varepsilon}_{ij}^{p} = \frac{1}{2} \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) + \frac{1}{2} \frac{\partial \sigma_{ij}}{\partial x_k} + \lambda \frac{\partial \sigma_{ij}}{\partial x_k} \]  

(A6)

The total, deviatoric strain rate for incompressible materials is:

\[ \dot{\varepsilon}_{ij} = \frac{1}{2} \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) \]  

(A7)

and the total stress is:

\[ \sigma_{ij} = -p \delta_{ij} + \tau_{ij} \]  

(A8)

The time derivative, D/ Dt, of the deviatoric stresses in Eq. (A6) is the objective time derivative, which is necessary to make the rheological formulation independent of the coordinate system (i.e. objective). Various correct formulations of the objective time derivatives exist and we use the upper convected time derivative (Schmalholz et al., 2001). Incompressibility described in Eq. (A3) is achieved numerically using Uzawa type iterations during which the pressure is also calculated. Our strategy for implementing the viscoelastoplastic rheology is to first calculate the viscoelastic trial stresses and if these stresses violate the plastic yield criterion they are reduced by plastic strain increments bringing stresses back to the yield function. This implementation of plasticity generates correct plastic shear bands identical to shear bands generated with other plasticity algorithms (Kaus, 2005). We solve Eq. (A6) for the deviatoric stress, neglecting the plastic strain rate, and approximate the partial time derivative of the deviatoric
stresses inside the objective time derivative with a finite difference operator, which yields:

\[ \tau_{ij} = D_{ij} + B_{ij} \sigma_{old}^{prld} \]  \hspace{1cm} (A9)

with

\[ D = \frac{1}{\eta} \begin{vmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{vmatrix} B = \frac{1}{1 + \tau} \begin{vmatrix} 1 & 2\Delta \frac{\partial \mathbf{v}_1}{\partial \mathbf{x}_1} & 0 \\ 2\Delta \frac{\partial \mathbf{v}_1}{\partial \mathbf{x}_1} & 1 & 2\Delta \frac{\partial \mathbf{v}_1}{\partial \mathbf{x}_2} \\ 0 & 0 & 1 \end{vmatrix} \]  \hspace{1cm} (A10)

\[ \Delta t \] is the time interval (or time step) and \( \sigma_{old}^{prld} \) are old stresses resulting from the finite difference approximation of the partial time derivative of stresses. The spatial velocity gradients appear in matrix \( B \) because the upper convected time derivative is used (Schmalholz et al., 2001). The effective viscosity depending on strain rates and temperature is:

\[ \eta = A^{\frac{1}{2}}_{ijkl} \exp \left( \frac{E}{nRT} \right) \]  \hspace{1cm} (A11)

with

\[ \hat{e}_{ij} = \frac{1}{4} \left( \hat{e}_{11} - \hat{e}_{22} \right)^2 + \hat{e}_{12}^2 \]  \hspace{1cm} (A12)

After the viscoelastic stresses are calculated, a stress correction has to be made if the yield criterion is violated, i.e. the yield function is larger than zero. The yield function, \( F \), and the potential function, \( Q \), for a Mohr–Coulomb material are:

\[ F = \tau_{ii} + \sigma_{n} \sin \psi \cos \theta - C \]  \hspace{1cm} (A13)

\[ Q = \tau_{ii} + \sigma_{n} \sin \psi + \text{constant} \]

with

\[ \tau_{ii} = \frac{1}{4} (\sigma_{11}^2 - \sigma_{22}^2)^2 + \tau_{12}^2 \]  \hspace{1cm} (A14)

\[ \sigma_{n} = \frac{\sigma_{11}^2 - \sigma_{22}^2}{2} \]

We assume the dilatancy angle, \( \psi \), to be zero in the potential function which means that the plastic flow is incompressible. The derivative of the plastic potential function with respect to stress provides the plastic strain rates Eq. (A6). These derivatives are:

\[ \frac{\partial Q}{\partial \sigma_{11}} = \sigma_{11} - \sigma_{22} \]  \hspace{1cm} (A15)

\[ \frac{\partial Q}{\partial \sigma_{22}} = -\sigma_{11} + \sigma_{22} \]

\[ \frac{\partial Q}{\partial \sigma_{12}} = 2\sigma_{12} \]

Based on these derivatives, the stress increments bringing stresses back to the yield function can be calculated. Stress increments that must be subtracted from the viscoelastic trial stresses are (Kaus, 2005):

\[ \Delta \sigma_{11} = (1-f) \frac{\sigma_{11}^2 - \sigma_{22}^2}{2} \]  \hspace{1cm} (A16)

\[ \Delta \sigma_{22} = -(1-f) \frac{\sigma_{11}^2 - \sigma_{22}^2}{2} \]

\[ \Delta \sigma_{12} = (1-f) \sigma_{12} \]

with \( f = \frac{\sigma_{11}^2 \sin \psi + \cos \theta}{\tau_{11}} \)

Pressure remains constant during stress correction because plastic flow is assumed to be incompressible. The above formulation for a Mohr–Coulomb failure criterion can also be used for a von Mises failure criterion by setting the angle of internal friction, \( \theta \), to zero and cohesion to the value of the von Mises yield strength.

The stress–strain rate relationship is nonlinear due to the (i) strain rate and temperature dependences of the effective viscosity, (ii) velocity–gradient dependence of the objective time derivative and (iii) stress dependence of the plastic stress correction. All non-linearities are solved within the same Picard iteration loop.

Erosion and redeposition of sediments is described with a diffusion type equation for the surface topography, \( S \):

\[ \dot{S} = \frac{\partial S}{\partial t} = -E \frac{\partial S}{\partial Q} \]  \hspace{1cm} (A17)

The surface topography is continuously deformed due to the Lagrangian finite element mesh and the free surface boundary condition. Surface topography is modified after each time step by applying the diffusion equation.

The two velocities and the temperature are the three unknowns in the final system of equations. The pressure is discontinuous across elements, is eliminated on the element level and is calculated during Uzawa iterations that are applied to fulfill the incompressibility equation.

References


