Modeling mountain building and the seismic cycle in the Himalaya of Nepal

R. Cattin\(^1\) and J. P. Avouac
Département Analyse Surveillance Environnement, Laboratoire de Détectection et Géophysique, Commissariat à l’Energie Atomique, Bruyères-le-Châtel, France

Abstract. A host of information is now available regarding the geological and thermal structure as well as deformation rate across the Himalaya of central Nepal. These data are reconciled in a two dimensional mechanical model that incorporates the rheological layering of the crust which depends on the local temperature and surface processes. Over geological timescale (5 Ma) the \(\sim 20\) mm/yr estimated shortening rate across the range is accommodated by localized thrust faulting along the Main Himalayan Thrust fault (MHT). The MHT reaches the surface along the foothills, where it is called the Main Frontal Thrust fault (MFT). The MHT flattens beneath the Lesser Himalaya and forms a midcrustal ramp at the front of the Higher Himalaya, consistent with the river incision and the anticlinal structure of the Lesser Himalaya. Farther northward the MHT roots into a subhorizontal shear zone that coincides with a midcrustal seismic reflector. Aseismic slip along this shear zone is accommodated in the interseismic period by elastic straining of the upper crust, increasing the Coulomb stress beneath the front of the Higher Himalaya, where most of the microseismic activity clusters. Negligible deformation of the hanging wall requires a low apparent friction coefficient \(\mu\) less than \(\sim 0.3\) on the flat portion of the MHT. On the ramp, \(\mu\) might be as high as \(0.6\). Sensitivity tests show that a rather compliant, quartz-rich rheology and a high radioactive heat production in the upper crust of \(\sim 2.5\) \(\mu\)W/m\(^2\) is required. Erosion affects the thermal structure and interplays with crustal deformation. A dynamic equilibrium is obtained in which erosion balances tectonic uplift maintaining steady state thermal structure, topography, and deformation field. Using a linear diffusion model of erosion, we constrain the value of the mass diffusivity coefficient to \(0.5-1.6\times 10^4\) m\(^2\)/yr. This study demonstrates that the data are internally consistent and compatible with current understanding of the mechanics of crustal deformation and highlight the role of viscous flow in the lower crust and of surface erosion in orogeny processes on the long term as well as during interseismic period.

1. Introduction

The Himalayas are a particularly appropriate example to investigate mountain-building processes. Ongoing indentation of India into Eurasia has induced crustal-scale imbrication of thrust sheets on the geological timescale and crustal thickening [e.g., Le Fort, 1986] and several major thrust faulting earthquakes over the last century [e.g., Seber and Armbruster, 1981; Molnar, 1990]. Over the last few years a host of information on crustal deformation in the Himalaya of Nepal has been made available that pertains either to interseismic deformation [Bilham et al., 1997; Pandey et al., 1995; Larson et al., 1999; Jouanne et al., 1999] or to long-term deformation, that is, averaged over many seismic cycles [Lavé and Avouac, 2000]. The structure of the range is relatively well known from geological [Brunel, 1986; Schelling and Aria, 1991] and geophysical investigations in Nepal and southern Tibet [Hirn et al., 1984; Zhao et al., 1993; Nelson et al., 1996].

We show that these data can be reconciled on the basis of a single mechanical model of both long-term and interseismic processes. We focus on a N18°E section across the Himalaya of central Nepal, at the longitude of Kathmandu basin, because most of the data hold for this particular section (Plate 1). Since the thermal structure [e.g., Royden, 1993; Henry et al., 1997] and the mechanics of mountain building depends on surface processes [Molnar and England, 1990; Avouac and Barou, 1996], we have paid a particular attention to that factor.

We first summarize constraints on the model in terms of geometry or deformation. We next describe the modeling approach and assumptions and discuss the results and their sensitivity to the model parameters.
2. Constraints on Model’s Geometry, Boundary conditions, and Deformation

2.1. Geometry

The Himalaya marks the southern edge of the Tibetan Plateau. It was built as a result of thrust faulting along major faults such as the Main Central Thrust fault (MCT), Main Boundary Thrust fault (MBT) and the Main Frontal Thrust fault (MFT) [e.g., Le Fort, 1986]. Deformation has apparently prograded southward with the MFT being the most recently activated one in the sequence. The MFT is also probably the major active fault at present in the Himalaya of central Nepal [Lavé and Avouac, 2000]. This fault is thought to reach the top of the Indian basement at a depth of about 5 km to extend subhorizontally beneath the Lesser Himalaya and to root along a crustal ramp beneath the front of the Higher Himalaya [Schelling and Arita, 1991; Pandey et al., 1995; Lavé and Avouac, 2000] (Figure 1c). North of the high range may connect with the major midcrustal reflector, imaged at a depth of ~25 to ~40 km from the International Deep Profiling of Tibet and the Himalaya (INDEPTH) experiment [Zhao et al., 1993; Nelson et al., 1996]. For consistency with previous terminology, we refer to this fault as the Main Himalayan Thrust fault (MHT) [Zhao et al., 1993; Nelson et al., 1996; Bilham et al., 1997] (Figure 1e).

The seismic experiments conducted in southern Tibet all suggest a crustal thickness of the order of 70-80 km [e.g., Iliric et al., 1984; Zhao et al., 1993; Nelson et al., 1996], consistent with isostatic support of the plateau [Lyon-Caen and Molnar, 1985; Jin et al., 1996]. Gravity data are consistent with a crustal thickness of ~40 km under the Indo-Gangetic plain [Lyon-Caen and Molnar, 1985]. All those features are summarized in the synthetic section shown in Figure 1c.

2.2. Deformation and Shortening Across the Range Over the Long Term

The shortening rate across the whole range was estimated from various approaches that cover different time spans. The southward migration of proximal facies in the foreland over the last 15 to 20 Myr suggests a thrusting rate between 10 and 15 mm/yr [Lyon-Caen and Molnar, 1985]. This estimate assumes that loading of the foreland by the topography proceeds at the same pace as thrust faulting, neglecting erosion. This approach gives an average velocity over the whole Himalayan belt and thus underestimates the thrusting rate in the central Himalaya. A minimum shortening rate of 11 ±4 mm/yr was also established, for about the same time period, from balancing an upper crustal cross section [Schelling and Arita, 1991].

If E-W Quaternary extension in southern Tibet is assumed to accommodate the lateral variation of the direction of thrusting along the arcuate Himalayan front [Baranowsky et al., 1984; Armijo et al., 1986; McCaffrey and Nabelek, 1998], the estimated extension rate of about 10 mm/yr implies a Late Quaternary shortening rate of 20 ±10 mm/yr [Armijo et al., 1986]. The estimated seismic moment of the large earthquakes that struck the Himalaya over the last century yield a shortening rate of 18 ±8 mm/yr [Molnar and Deng, 1984; Molnar, 1990]. Although these estimates hold for different periods and ignore eventual lateral variation, for the purpose of the modeling we assume 20 mm/yr of horizontal shortening between India and southern Tibet along the section considered in this study. This shortening might be taken up by slip along the MHT and distributed deformation of the crustal wedge [Molnar, 1987]. In their study of deformed Holocene terraces along the Bagmati and Bakeya Rivers in the Siwaliks of central Nepal, Lavé and Avouac [2000] have shown that localized thrust faulting on the MFT, hence along the MHT, absorbs as much as 21.5 ±2 mm/yr, an amount that accounts for all shortening rate across the range. These data suggest that the Himalaya is thrust over the MHT with little internal deformation. This behavior differs significantly from that of a critical Coulomb wedge, in which internal deformation is distributed so as to insure a quasistatic balance between friction at the base of the hanging wall and lateral variation of the topography [Dahlen, 1990].

These features are summarized in Figure 1d, where we have reported 20 ±10 mm/yr between southern Tibet and India (label 1). We report 18 ±8 mm/yr between the high range and India (label 2) because the large Himalayan earthquakes are considered to have ruptured a zone that extends from the front of the Higher Himalaya to beneath the foreland [Seeber and Armbruster, 1981] and 21.5 ±2 mm/yr across the MFT where it reaches the surface in the piedmont (label 3).

2.3. Erosion

Tectonic crustal thickening due to deformation in the Himalaya is in part compensated by denudation. In the sub-

---

**Figure 1.** Summary of available information constraining the structure and both long-term and interseismic processes along a N18°E section at the longitude of Kathmandu (see Plate 1 for location) pertaining to the geometry of structures (Figure 1e) and crustal deformation over the long-term (Figures 1c and 1d) and during the interseismic period (Figures 1a and 1b). (a) Intereismic vertical velocities derived from leveling data [Jackson et al., 1992; Jackson and Bilham, 1994]. Velocities are relative to the southernmost point on the Indo-Gangetic plain. (b) Horizontal velocities with respect to India during interseismic period derived from GPS measurements [Larson et al., 1999]. (c) Incision rate derived from Holocene fluvial terraces [Lavé and Avouac, 2000] (shaded area) and from present river geometry [Lavé, 1997]. (d) Long-term shortening rate estimated by (1) Armijo et al. [1986], (2) Lyon-Caen and Molnar [1985] and (3) Lavé and Avouac [2000]. (e) The synthetic structural section was modified from Schelling and Arita [1991] and Pandey et al. [1995]. Only the seismic events (shaded circles) monitored during the period of the temporary seismic network were reported on this section because hypocentral depths determined from the permanent network alone are not accurate enough (errors are commonly in excess of 5 km). The seismc reflection profile of Zhao et al. [1993] was also reported.
Himalaya, the rate of river incision can be estimated from the elevation of abandoned Holocene fluvial terraces [Lavé and Avouac, 2000] (shaded area in Figure 1c). It appears that river incision, as well as denudation on hillslopes is at pace with tectonic uplift in the Sub Himalaya [Hurtrez et al., 1999]. Abandoned terraces along the Kali Gandaki River suggest that tectonic uplift is negligible beneath the Lesser Himalaya and increases northward beneath the Higher Himalaya [Iwata et al., 1984; Molnar, 1987]. A profile of river incision along the Narayani and Trisuli River across the
Plate 1. Seismotectonic map of central Himalaya showing major geological structures (MFT, Main Frontal Thrust; MBT, Main Boundary Thrust; MCT, Main Central Thrust), Trisuli River, Bagmati River, geodetic measurements, and seismic data. The green line indicates the leveling line used by Jackson and Bilham [1994]. GPS velocity vectors obtained by Larson et al. [1999] are represented with yellow arrows. Red circles indicate microseismicity monitored between 1994 and 1998 [Pandey et al., 1995, 1999]. Fault plane solutions are associated with earthquakes occurring on August 20, 1988, October 29, 1988, and January 31, 1991 [Chen and Kao, 1996]. Green circles show locations of well-located microseismic events monitored between July and December 1995 from three local, additional, three components short period stations. Black dashed line indicates the location of the N18°E cross section discussed in this study.
Lesser Himalaya and part of the Higher Himalaya (see Plate 1 for location) was also derived from present channel geometry using a model of river incision calibrated from a few abandoned terrace remnants (solid line in Figure 1c) [Lavé, 1997]. As argued by Burbank et al. [1996], we assume that hillslope processes are coupled to river incision so that denudation is controlled by river incision.

Erosion rates can also be constrained on average from the amount of sediments delivered by the rivers draining the range. The most recent compilation of these suggests that these rivers have transported about 0.5x10^6 m^3/yr over the last 2 Myr [Métiév, 1996], a flux that compares well with the sedimentary load in the present rivers that would amount to about 0.9x10^6 m^3/yr according to Pinet and Souriau [1988] and 1.05 ±0.3x10^6 m^3/yr according to Summerfield and Hulton [1994] and Métiév [1996]. Denudation rates probably vary laterally along the 2000 km long stretch of the Himalayan arc, but we may assume that the Himalaya of central Nepal lies in an intermediate situation. Accordingly, denudation along the section considered in this study would represent 250 to 675 m^2/yr.

2.4. Interseismic Deformation

Interseismic deformation in the Himalaya of central Nepal is now constrained by leveling data [Jackson et al., 1992; Jackson and Bilham, 1994] and Global Positioning System (GPS) measurements [Bilham et al., 1997; Larson et al., 1999; Jouanne et al., 1999]. It seems that most of the interseismic deformation is caused across a wide zone that runs along the front of the Higher Himalaya and that coincides with the belt of microseismic activity evidenced from local seismic monitoring [Pandey et al., 1995, 1999] (Plate 1 and Figure 1e). All these measurements are illustrated in Figures 1a and 1b. In order to account for the arcuate shape of the range, the GPS measurements were projected using small circles for the Himalayan arc [Bilham et al., 1997]. Given that depth determinations may be inaccurate by several kilometers in that area [Pandey et al., 1995], we show 6 months using three additional seismic stations on top of the belt of seismicity north of the Kathmandu basin. All the hypocenters located within the swath considered here are shown in Figure 1e.

Altogether, these data show a 50 km wide zone of active uplift that coincides with a steep gradient in horizontal velocities as well as with the zone of intense microseismic activity. It appears that in the interseismic period, strain and stress builds up in a relatively narrow zone around the ramp on the MHT (Figure 1). The flat portion of the MHT beneath the Higher Himalaya and southern Tibet would be creeping at about 20 mm/yr, while the ramp and the flat would remain locked [Bilham et al., 1997; Jouanne et al., 1999; Larson et al., 1999]. Modeling of this process using a subhorizontal dislocation provides a satisfactorily fit to the data, although the leveling data and the GPS measurements tend to favor solutions with slightly differing parameters [Bilham et al., 1997].

3. Modeling Approach and Assumptions

Interseismic processes can be conveniently modeled using elastic dislocations [e.g., Jackson and Bilham, 1994; Bilham et al., 1997]. Some authors have also used this approach to model long term processes and the growth of geological structures due to fault slip [King et al., 1988; Stein et al., 1988]. In these studies the geometries and slip rates of faults were prescribed a priori and deformation in the ductile layers of the crust was approximated by elastic dislocations. The interplay between surface processes and crustal deformation as well as the dependency of crustal deformation on temperature and pressure cannot be dealt with fully. We used a more comprehensive approach that accounts for the mechanical layering of the crust, the non-Newtonian rheology of rocks, and its dependency on temperature and pressure. We use a two-dimensional finite element model, ADELIC [Hasani et al., 1997], that was modified in order to incorporate surface processes and the dependency of rheology on local temperature. The limitation of this model is being two-dimensional.

3.1. Geometry and Boundary Conditions

We consider a 700 km long section (Figure 2) that approximates the N18°E section through the Himalaya of central Nepal presented in Figure 1e. The crustal thickness varies from 40 km beneath the Gangetic Plain to 75 km beneath the Tibetan Plateau. The model is initialized loaded with the present average topography along the swath considered here. Only the brittle portion of the MHT is prescribed. It includes the flat beneath the Lesser Himalaya and the midcrustal ramp at front of the Higher Himalaya. In order to test the sensitivity of the results to the assumed geometry of the MHT we have also carried on simulations in which a simple planar fault is considered (the "no-ramp model") hereafter with the same extremites as the flat-ramp fault model.

As argued by Nelson et al. [1996], we consider an Indian mantle lid that underthrusts southern Tibet. Vertical and horizontal displacements are excluded at the southern end of the model. At the northern end, vertical displacements are free, and for depths above 40 km a southward horizontal velocity of 20 mm/yr is imposed (Figure 2). The uppercrust is thus submitted to 20 mm/yr horizontal shortening. At the northern end and for depths below 40 km, we exclude horizontal displacements. The Indian mantle lid thus underthrusts Tibet with no shortening. We have arbitrarily placed this discontinuity discontinuity at a depth of 40 km in the lower crust. The model is long enough that the results in the Himalayan portion of the model are insensitive to the exact position of this discontinuity. The model is loaded with gravitational body forces (g=9.81 m/s^2) and is supported at its base by hydrostatic pressure to allow for isostatic restoring forces.

3.2. Surface Processes

Following Avouac and Burov [1996] denudation is modeled using a 1-D linear diffusion equation in the range, that is north of the MFT,
where $k$ is the mass diffusivity coefficient, expressed in units of area per time, $h$ is the elevation, and $x$ is the distance from the MFT. This linear diffusion law (equation (1)) assumes that the southward flux of sediments at the surface is proportional to the local slope. The value of $k$ was chosen to insure a denudation rate consistent with the 250-675 m$^2$/yr flow of sediments eroded from the range. In order to test the sensitivity of the model to erosion, $k$ was varied between $10^9$ and $10^{10}$ m$^2$/yr. In the foreland ($x \leq 0$), as argued by Hurtt et al. [1999], we assume that erosion balances tectonic uplift or subsidence. It implies that denudation exactly compensates uplift at the MFT and sedimentation maintains a flat foreland at a constant elevation south of the MFT. This model is only an approximation that was designed to account as simply as possible for the dependency of denudation on topography and sedimentation in the foreland (see Avouac and Burov [1996] for discussion).

3.3. Rheology of the Continental Lithosphere

A depth varying rheology has been incorporated with elastobrittle deformation in the upper part of the crust and ductile deformation in the lower part. The empirical rheological equations and laboratory-derived material properties [e.g., Brace and Kohlstedt, 1980] are used under the assumption they can be extrapolated to geological conditions.

Under low stresses, rocks deform elastically. For isotropic materials the relationship between each component of strain ($\varepsilon_{ij}$) and stress ($\sigma_{ij}$) is written

$$\varepsilon_{ij} = \frac{1 + \nu}{E} \sigma_{ij} - \frac{\nu}{E} \sigma_{kk} \delta_{ij},$$

where $E$ and $\nu$ are the Young's modulus and the Poisson's ratio, respectively.

Beyond the elastic domain, rocks deform brittlely or ductily. We use an elastoplastic pressure-dependent law with the failure criterion of Drucker-Prager [Hassani et al., 1997]. Failure occurs if

$$\frac{1}{2} (\sigma_1 - \sigma_3) = \left[ c \cot \phi + \frac{1}{2} (\sigma_1 + \sigma_3) \right] \sin \phi,$$

where $c$ is the cohesion, and $\phi$ is the internal friction angle.

Ductile flow in the lithosphere is empirically described by a law which relates the critical principal stress difference necessary to maintain a steady state strain rate to a power of the strain rate [Carter and Tessen, 1987; Tessen and Carter, 1987; Kirby and Kronenberg, 1987]. This power law creep is written

$$\dot{\varepsilon} = A_P (\sigma_1 - \sigma_3)^n \exp(-E_P/RT),$$

where $R$ is the universal gas constant, $T$ is the temperature, $E_P$ is the activation energy, and $A_P$ and $n$ are empirically determined material “constants,” assumed not to vary with stress and $(P, T)$ conditions. The ductile flow law is thus strongly dependent on rock type and temperature.

We only distinguish the crust and upper mantle. For the upper mantle, an olivine-controlled rheology is assumed (see parameters in Table 1). For the crust we have considered two end-members, by considering either a quartz-controlled or a diabase-controlled rheology (Table Table 1). The quartz like rheology implies a relatively thin elastic core in the middle crust and a thick lower crust that favors decoupling between mantle and crust, while a diabase rheology results in a stronger crust with a higher viscosity lower crust [Burov and Diament, 1995].

3.4. Fault Behavior

The MHT is assumed to follow a simple static friction law,

$$|\sigma_T| - \mu (\sigma_N - p) \leq 0 \Leftrightarrow |\sigma_T| - \mu' \sigma_N \leq 0$$


where $\sigma_T$ and $\sigma_N$ are the normal and the shear stress on the fault and $\mu$ is the friction coefficient, $\mu'$ is the effective friction coefficient, and $p$ is the pore pressure. We use a dry Coulomb-Navier criterion which underestimates the "real" friction coefficient. Numerically, we follow the approach of [Jean and Touzet, 1988] in which the fault is simulated by contact between nodes on each side of the discontinuity.

In order to simulate long-term deformation, we set a friction coefficient low enough so as to allow thrusting over the MHT with negligible internal deformation of the hanging wall. If the friction is too large, the hanging wall deforms internally and evolves toward a critical Coulomb wedge [Dahlen, 1980] inconsistent with the data. We found that the friction on the flat portion of the MFT must be smaller than 0.13. The friction on the ramp can be up to about 0.6, because it coincides with a rapid increase in the topography that inhibits internal deformation of the hanging wall. To model interseismic processes we lock the MHT. We show later that a value of $\sim 0.61$ would be sufficient to lock the MHT for several hundred years.

3.5. Thermal Structure

We compute the thermal structure and deformation field separately because the data show a simple deformation pattern with nearly steady state hanging wall and foot wall. Erosion was taken into account because it greatly affects the thermal structure [e.g., Royden, 1993]. In this modeling we assume that erosion exactly balances tectonic uplift, as computed from thrusting of the hanging wall over the flat ramp flat geometry of the MHT assuming a rigid footwall, and determine a steady state thermal structure. We have compared the 1-D analytical formulation of [Royden, 1993] that neglects conduction due to lateral gradient, with a 2-D finite elements scheme developed by Taylor (FEAP) [Zienkiewicz and Taylor, 1989] and adapted by Henry et al. [1997]. The assumed boundary conditions are a constant surface temperature of 0°C and a bottom heat flow of 15 mW/m². We have selected the set of parameters that best explains the thermobarometric data in the Himalaya, in particular the inverted geotherm [Le Fort, 1975], as discussed by Royden [1993]. We have in particular considered a relatively high heat production in the upper crust of 2.5 $\mu$W/m³. The two computations yield very similar temperature fields (Figure 3) that imply a peak of temperature and pressure in the rocks that are thrust over the MHT corresponding to about 650°C and 1-8.8 GPa. As discussed by Royden [1993], these conditions compare well with those described in the High Himalayan Crystalline sheet [Hubbard, 1989; Hubbard and Harrison, 1989]. This modeling also predicts sufficiently high temperatures some 200 km north of the Himalaya to explain partial melting in the middle crust as revealed by INDEPTH investigations [e.g., Nelson et al., 1996; Henry et al., 1997]. In order to test the sensitivity of our results to the thermal structure, we have also used the temperature field obtained by assuming a lower heat production in the upper crust (Figure 3c).

4. Experiments and Results

4.1. Description of Experiments

The "long-term" deformation is first calculated by letting the MHT slip freely until a steady state is reached. We consider that steady state is attained when the velocity field does not evolve any more ($\delta V / V < 5\%$). The interseismic phase is next computed by locking the fault. The boundary conditions remain the same during all the experiments. In the interseismic period, velocities vary slightly. Here we show results by averaging deformation rates, stress variations, and velocities computed over a 350 years long interseismic period.

Elements are sized $\Delta t \sim 5$ km throughout the upper mantle and the crust. The integration of the diffusion equation 1 requires that $\Delta t < 0.5 \Delta t^2 / k$. This condition is reached for $\Delta t < 125,000$ years for $k = 10^7$ m²/yr and for $\Delta t < 125$ years for $k = 10^5$ m²/yr. The relaxation time associated with the lowest viscosity value ($\eta = 5 \times 10^{14}$ Pa.s) is $\sim 25$ years. For all computations we chose a time step of 5 years to model long-term deformation and $\Delta t = 0.1$ year during the interseismic phase.

4.2. Discussion of the Reference Model (Model 1)

The reference model assumes a quartz-controlled rheology in the crust, an olivine-controlled rheology in the mantle, a mass diffusivity coefficient of $10^4$ m²/yr, a heat production of 2.5 $\mu$W/m³ in the upper crust, and a flat-ramp geometry of the brittle portion of the MHT. A steady state velocity field is reached after about 5x10⁵ years. A subhorizontal shear zone develops in the middle lower crust that relays to the north frictional sliding along the MHT (Plate 2a). This shear zone coincides with the middle-crustal reflector revealed by INDEPTH I experiment [Zhao et al., 1993] (Plate 2a).

In the long term the hanging wall is thrust over the MHT and its ductile continuation with little internal deformation
Figure 3. Thermal structures used in the study. (a) The temperature field was calculated from the 1-D analytical solution of Royden [1993], and a 2-D steady state finite element model [Henry et al., 1997] for a radioactive heat production of (b) 2.5 μW/m³ and (c) 1.5 μW/m³.

(Figure 4a), resulting in a nearly constant horizontal velocity in the hanging wall above the MFT (Figure 5b). Horizontal velocities in the interseismic period are slightly underestimated by 2-3 mm/yr compared to the GPS measurements, suggesting that a shortening rate slightly larger than 20 mm/yr would have given a better fit (Figure 6b).

The footwall of the MHT does not deform significantly either, and uplift essentially results from thrusting over the ramp. It yields a profile of uplift that closely mimics the observed river incision pattern (Figure 5a). As shown in Plate 2, the quartz rheology allows some ductile shear in the footwall beneath the lesser Himalaya. This deformation produces some crustal thickening that compensates the subsidence caused by sediment accumulation in the foreland. Our modeling yields a denudation pattern that nearly balances tectonic uplift (Figure 5a). This demonstrates that coupling between surface processes and crustal deformation takes place, leading to dynamic equilibrium [Avouac and Burov, 1996]. The total rate of erosion from the range in our model is ~440 m²/yr, an amount that compares well with the value of 250 to 675 m²/yr discussed above.

During the interseismic period, ductile shear in the lower crust is not significantly affected by the inhibition of frictional sliding along the MHT. This is consistent with the interpretation of interseismic processes in terms of a subhorizontal creeping dislocation [Bilham et al., 1997], except that in our modeling, elastic straining only occurs in the upper crust, while straining in the lower crust is continuously relaxed by viscous flow, and interseismic uplift includes the isostatic response to erosion.
Plate 2. Shear strain rates over the long term computed assuming (a) a quartz rheology or (b) a diabase rheology. In both cases a subhorizontal shear zone develops that closely follows the midcrustal reflector imaged by the INDEPTH experiment [Zhao et al., 1993].
Figure 4. Long-term velocity field (over many seismic cycles) and during interseismic period. (a) In the long term, frictional sliding in the upper crust and ductile shear at depth along the MHT allows overthrusting of the hanging wall with little internal deformation. (b) During the interseismic period the horizontal shortening rate is also accommodated by ductile shear in the lower crust and by elastic straining of the hanging wall around the ramp.

Figure 5. (a) Calculated erosion rates (solid line), vertical velocities (dashed line), and estimated river incision rates (shaded area). (b) Comparison between calculated (solid line) and estimated (circles with error bars) long-term horizontal velocities.
The maximum horizontal elastic straining is near the ramp, at the tip of the subhorizontal shear zone. It produces locally a steeper gradient of horizontal velocities as suggested by the GPS measurements (Figure 6b), as well as a zone of localized uplift consistent with the leveling data (Figure 6a). The model predicts different patterns of uplift over the long term and in the interseismic period. They are slightly offset with respect to each other and have different characteristic widths. These two patterns include a common isostatic response to denudation. The long-term pattern reflects passive thrusting of the hanging wall over the ramp and thus depends on the ramp's geometry, while interseismic uplift reflects elastic straining at the tip of the creeping shear zone.

As shown in Figure 7 the locking of the MHT in the interseismic period induces only small stress variations. Cumulated over 350 years they amount to only 0.3 MPa on the flat and 2 MPa on the crustal ramp. Frictional sliding on the ramp corresponds to a friction coefficient that may be as large as

---

**Figure 6.** Calculated (solid line, average; dotted lines, upper and lower bounds) and measured (open circles) interseismic (a) vertical and (b) horizontal velocities.

**Figure 7.** Variation of normal (\(\sigma_N\)) and tangential (\(\gamma\)) stresses and their ratio on the MHT. We distinguish between the flat and ramp portion of the MHT. The MHT slips freely over long term and it is locked during interseismic period.
Plate 3. (a) Comparison between the focal mechanisms and the dip of planes with maximum Coulomb stress (blue lines) due to interseismic and regional stresses. Since stress variations over the interseismic period are small, this pattern does not vary either. (b) Spatial distribution in depth of microseismicity (white circle) and Coulomb stress change on optimally oriented planes calculated after 350 years of interseismic strain accumulation.
0.6, and locking over the 350 years interseismic period requires a value only larger than 0.6 by less than 1% (Figure 7). Frictional sliding on the flat requires a much lower apparent friction. The stress field obtained from our computation corresponds to a friction of less than about 0.12 and locking for 350 years also requires only a 1% increase. The stress field in the medium is thus nearly constant during our modeling.

In order to compare the computed interseismic stress field with the seismicity we have calculated the variation of the Coulomb stress field during the interseismic period. The Coulomb criterion is often used to characterize the conditions of failure in rock mechanics [e.g., Jaeger and Cook, 1979]. It has also been shown to be able to predict fault plane orientations and the locations of triggered earthquakes and aftershocks [Oppenheimer et al., 1988; King et al., 1994; Stein et al., 1988].

According to this criterion, failure occurs on a plane oriented at \( \alpha \) to the \( \sigma_1 \) axis when the Coulomb stress \( \sigma_f \) exceeds a specific value

\[
\sigma_f = \tau_\alpha - \tan \phi (\sigma_\alpha - p) \Rightarrow \sigma_f = \tau_\alpha - \tan \phi'\sigma_\alpha,
\]

where \( \tau_\alpha \) is the shear stress on the failure plane, \( \sigma_\alpha \) is the normal stress, \( p \) is the pore pressure, \( \phi' \) is the internal friction angle, and \( \phi' \) is the effective internal friction angle. The fault plane is optimally oriented (maximum Coulomb stress) when

\[
\tan 2\alpha = -\tan \phi'.
\]

The formulations used to calculate the optimally oriented planes and the variation of Coulomb stress are presented in more detail by King et al. [1994].

The optimally oriented planes for rupture, according to the Coulomb criterion, were computed for an effective internal angle of friction of 30° (\( \tan \phi' = 0.6 \)). The regional stress field is associated with the stress calculated with our long-term approach, and the Coulomb stress change is associated with the locked fault. It appears that the stress field is consistent with focal mechanisms of the two events occurring in 1988 (Plate 3a). For the deep, magnitude \( M=6.5 \) earthquake occurring on August 20, 1988, the calculated and estimated dip angle are equal to \( -24^\circ \) and to \( 70^\circ \) respectively [Chen and Kao, 1996]. For the magnitude \( M=5.5 \) October 29, 1988, earthquake occurring near the crustal ramp, the calculated and estimated dip angle are equal to \( 36^\circ \) and to \( 30^\circ \), respectively. By contrast, the nodal planes of the \( M=5.6 \) Saroshin earthquake (B. Kumar, personal communication, 1999) do not correspond to optimally oriented planes because that event shows a strike-slip component suggesting 3-D effects not accounted for by our modeling. It can also be noticed that the dip of the ramp approximately corresponds to maximum Coulomb stress. This is consistent with the fact that frictional sliding on the ramp can occur with a friction that may be as large as 0.6.

Plate 3b shows the variation of the Coulomb stress over the interseismic period, for comparison with the hypocenters of Figure 1e. About 83% of the events fall within the area of increased Coulomb stress. This correlation suggests that stress accumulation over the interseismic period triggers microseismic activity. The magnitude distribution of these earthquakes show that they release less than 1% of the energy that accumulates due to elastic straining.

### 4.3. Sensitivity Test

The reference model thus appears to reconcile satisfactorily all the data related to interseismic and long-term deformation across the Himalaya of central Nepal. In this section we assess the sensitivity of the results to the various model's parameters.

#### 4.3.1. Sensitivity to the rheology of the crust (model 2)

The reference model assumes a relatively soft rheology (quartz type). When the crust is assumed to follow a diabase rheology, with the same thermal structure as for the reference model, it implies a higher viscosity of the lower crust and hence a stronger coupling between mantle and crustal deformation (Plate 2b). On the long term this model predicts buckling of the lithosphere [see also Burbank et al., 1993] as indicated from the predicted pattern of uplift (Figure 8a). In this model a shear zone root less deeply in the lower crust (~8 km). Compared to the reference model, it results in a pattern of uplift that is wider in the long term (Figure 8a). More importantly, contrary to the reference model, this crustal rheology does not allow any significant ductile shear in the Indian footwall (Plate 2). As a result, there is no significant crust thickening in front of the Higher Himalaya, and the Lesser Himalaya is flexed down. In that case we should observe subsidence and sedimentation in the Lesser Himalaya. This is inconsistent with the estimated pattern of river incision (Figure 8a). A rheology allowing for easy decoupling of the upper crust and mantle beneath Lesser Himalaya thus seems necessary to account for the data used in this study.

In the interseismic period the gradient of horizontal velocity is not so steep as for the reference model mainly because the diabase rheology implies a thicker elastic core [Burbank and Dietz, 1995; Avouac and Burbank, 1996]. It does not fit well the GPS measurements (Figure 8b).

#### 4.3.2. Effect of temperature (model 3)

Assuming a lower heat production in the upper crust of \( 1.5 \mu W/m^3 \), temperatures are systematically lower than for the reference model (Figure 3c). Given the use of a thermally activated rheological law (equation (4)), this change results in a thicker elastic core and higher viscosity in the lower crust. Consequently, as for the model 2 (diabase rheology), subsidence is produced in the Lesser Himalaya (Figure 8a).

#### 4.3.3. Suppression of the midcrustal ramp (model 4)

Here we consider a planar MHT that dips uniformly by \( \approx 8^\circ \) from the MFT to beneath the Higher Himalaya. Given that in the interseismic period the MHT is locked, the interseismic deformation is the same as for the reference model (Figure 8d). Horizontal displacements on the long term are not affected either, since the hanging wall is thrust along the brittle part of the MHT with negligible horizontal shortening.
as for the reference model. The major difference with the reference model lies in the predicted pattern of uplift over the long term, since it predicts significant tectonic uplift in the Lesser Himalaya due to the suppression of the flat portion of the MHT, inconsistent with the observations (Figure 8c). The suppression of the midcrustal ramp causes significant tectonic uplift in the Lesser Himalaya inconsistent with the pattern of river incision.

4.3.4. Influence of erosion. Here, we vary the mass diffusivity $k$ characterizing the intensity of erosion between $10^3$ and $10^5$ m$^2$/yr. Since erosion partly drives tectonic uplift due to isostasy, the uplift rates are affected both in the long term and during the interseismic period (Figure 9). By contrast, horizontal velocities are only slightly sensitive to the erosion rates (Figure 9). The data (erosion rate and interseismic uplift) can be fit satisfactorily for values of $k$ between $0.5 \times 10^4$ and $2 \times 10^4$ m$^2$/yr. We have also computed the total amount of material eroded from the range as a function of $k$ (Figure 10). It appears that values of $k$ between $0.5 \times 10^4$ and $1.6 \times 10^4$ m$^2$/yr are required to fit the estimated range of 250 to $675 \times 10^4$ m$^2$/yr (Figure 10). We found that linear diffusion model of erosion provides a good fit to the data including the estimated pattern of river incision and the total amount of sediments eroded from the section of $k$ to $0.5-1.6 \times 10^4$ m$^2$/yr.

4.3.5. Influence of friction along the MHT. Negligible deformation of the crustal wedge above the MHT requires a low apparent friction on the flat portion of the MHT. The model described here corresponds to a friction of less than $\sim 0.13$. We have also computed the maximum friction so as to prevent internal deformation of the hanging wall, using the analytical formulation of Dahlen [1990]. We find that the basal friction must be smaller than 0.2 for an internal pore pressure fluid ratio of less than 0.4, or smaller than
0.3 if a pore fluid pressure ratio as large as 0.9 is assumed (Figure 11). The same deformation pattern as that obtained in the reference model could therefore be obtained for a friction of up to 0.3. The major difference would lie in larger deviatoric stresses in the upper crust.

The relatively good agreement of the planes with maximum Coulomb shear stress with the estimated dip angle of the ramp and the few fault plane solutions in the area suggest that the stress field computed in the reference model is probably realistic. We thus conclude that the data suggest a low apparent friction on the flat portion of the MHT necessarily less than 0.3 and most probably of the order of 0.1. We have similarly estimated the maximum possible friction on the ramp (Figure 11). The result is independent of the pore pressure ratio, and we find a maximum friction of about 0.6 close to the internal friction of the upper crust. This is simply because the ramp is oriented nearly optimally with respect to the stress field.

5. Conclusion and Implications

Our modeling appears to reconcile all data on recent deformation in the Himalaya of central Nepal and the geometry of the thrust system as constrained from available geological, geomorphic, and geophysical data. First of all it demonstrates the internal consistency of the data set and its compatibility with current understanding of the mechanics of crustal deformation.

According to our modeling, ~20 mm/yr of slip along the MHT must be accommodated by either coseismic, preset-
Figure 10. Total amount of material eroded from the range as a function of mass diffusivity and estimated values derived from the total volume of the sediment in the Bengal and Indus fans and in Ganges plain [Summerfield and Hutton, 1994; Metivier, 1996], from river incision [Lave and Avouac, 2000] and from sedimentary load in the present rivers [Pinet and Souriau, 1988]. The model predicts a flux of sediments consistent with estimations for k between 0.5x10^4 and 1.6x10^4 m^2/yr.

mic, or postseismic slip. The large Himalayan earthquakes, with magnitude larger than 8, probably account for most of the deficit of interseismic deformation compared to long-term deformation. The coseismic displacement is computed for 6 m of slip along the brittle portion of the MHT [Oka da, 1992], assuming that the length of the rupture is 200 km [Pandey and Molnar, 1988]. We show that this average slip on the locked fault does actually compare well with the difference between the long-term and interseismic deformation accumulated during 300 years (Figure 12).

The flat-ramp geometry of the MHT is essentially constrained from the pattern of river incision [Lavé, 1997; Molnar, 1987], but it should be recalled that it was also inferred from the anticlinal structure of the Lesser Himalaya [e.g., Schelling and Arita, 1991; Brunel, 1986] and from gravity modeling [Lyon-Caen and Molnar, 1985]. It moreover explains the position of the front of the Higher Himalaya well to the north of the MFT. Indeed, in our modeling, interseismic uplift beneath the Higher Himalaya is due to elastic straining of the upper crust so that it does not contribute to the building of the topography. This view differs from that of Bilham et al. [1997], who pointed out the coincidence of the maximum rate of uplift indicated by the leveling data with the front of the high range. In our modeling the topographic front is rather maintained by thrusting over the ramp over the long term.

Frictional sliding along the flat portion of the MHT over the long term requires a low apparent friction of less than 0.3. Such a low value might be taken to suggest the presence of a gouge layer [e.g., Scholz, 1987; Marone et al., 1990] and of fluids trapped along the decollement [e.g., Lachenbruch and Sass, 1991]. The fluids could be released by the dehydration reactions due to temperature and pressure increase during underthrusting of the Indian crust beneath the Lesser Himalaya. Another explanation is that the flat decollement is only activated during large earthquakes and that the rupture propagates dynamically even though static shear stress would be small compared to those required for quasistatic frictional sliding.

In our modeling the flat portion and the midcrustal ramp along the MHT are locked during the interseismic period. Aseismic slip at deeper levels along this shear zone is accommodated by elastic straining of the upper crust. This process induces stress buildup increasing Coulomb stresses and hence triggering microseismic activity beneath the Higher

Figure 11. Diagram showing the relationship between angle of internal friction (φ), internal pore pressure (λ), and the critical friction coefficient at the base of a Coulomb wedge according to Dahlén’s [1990] analytical formulation. If the friction along the decollement at the base of the wedge exceeds the critical value, the wedge deforms internally. In our model we assume an internal friction angle of 30°, a dip angle of 4° along the flat decollement and of 16° along the ramp. The mean slope of the topography is about 1° above the decollement and 4° above the ramp. Whatever the pore ratio, deformation of the hanging wall above the ramp requires a basal friction in excess of ~0.6. This is consistent with the fact that in our modeling the normal and the tangential stresses along the ramp correspond to an apparent friction of 0.6. The hanging wall does not significantly deform internally over the ramp. On the flat decollement the friction must be lower than ~0.3 if a large pore pressure ratio is assumed. The observations of little shortening in the Lesser Himalaya on the long term thus requires a friction of less than ~0.3 or less than ~0.2 if a low pore pressure is assumed.
Himalaya. If, as seems to be the case, the recurrence interval between two major earthquakes at a given point along the arc is of the order of a few centuries [e.g., Molnar, 1987] then the stress variations must be relatively small (~1 MPa). Only small variations, of less than 1%, of the friction coefficient on the MHT is required to allow stress accumulation and release over seismic cycle.

The modeling shows that a rather soft rheology of the crust is required to allow decoupling between the crust and mantle. We found that a quartz rheology for the whole crust together with rather high temperatures obtained by assuming a large heat production in the upper crust provides the best fit to the observations. Other authors have already pointed out to the necessity of a low viscosity lower crust beneath Tibet [e.g., Jin et al., 1996] and to its role in various tectonic context [Lobkovsky, 1988; Kruse et al., 1991; Gregory and Chase, 1994].

Finally, we also show that erosion is a major factor in the mechanics of the Himalayan orogeny. It leads to a steady state thermal structure with rather hot temperatures in the upper crust due to rapid vertical advection. It moreover balances the zone of localized uplift above the midcrustal ramp and prevents the range broadening southward. This leads to the development of a steady state deformation field in which crustal deformation and surface processes are coupled as described also by Avouac and Burov [1996] on the basis of a different kind of modeling. Unexpectedly, we find that interseismic deformation is also sensitive to erosion. The fact that a fraction of the uplift pattern in the interseismic period is due to isostatic response to erosion helps reconciling the leveling data with the GPS measurements of horizontal displacements. Conversely, it implies that some constraints on denudation rates can be derived from the measurement of interseismic uplift.

Acknowledgments. This modeling was built on previous results acquired in a collaborative program that has benefited from the contributions of J. Lavé, G. Burov, and of our Nepalese colleagues at the Department of Mines and Geology, in particular M. R. Pandey and R. P. Tandukar. Constructive reviews by S. H. Lambr, B. Parson, and R. Stein contributed to improving the manuscript. We are most grateful to P. Henry for computing the thermal structure used in this study and for fruitful discussions of our results, and we thank J. Chéry for providing the finite element code ADELI.

References


