Testing the effects of topography, geometry and kinematics on modeled thermochronometer cooling ages in the eastern Bhutan Himalaya

Michelle E. Gilmore[1], Nadine McQuarrie [1], Paul Eizenhöfer [1], Todd A. Ehlers [2]

[1] Department of Geology and Environmental Science, University of Pittsburgh, Pittsburgh, 15260, USA
5   [2] Department of Geoscience, University of Tübingen, Tübingen, D-72074, Germany

Correspondence to: Nadine McQuarrie (nmcq@pitt.edu)

Abstract. The temporal and kinematic evolution of fold-thrust belts is a critical component for evaluating the viability of proposed plate tectonic, geodynamic and even climatic processes in regions of convergence. Thermochronometer data have the potential to provide temporal constraints, but interpretations of these data are sensitive to both exhumational and

10   deformational processes. In this study, reconstructions of a balanced geologic cross section in the Himalayan fold-thrust belt of eastern Bhutan are used in a flexural and thermal-kinematic model to understand the sensitivity of predicted cooling ages to changes in fault kinematics, geometry and topography. We sequentially deform the cross section with ~10-km deformation steps and apply flexural loading and erosional unloading at each step to develop a high-resolution evolution of deformation, erosion, and burial over time.

15   Comparison of model-predicted cooling ages to published thermochronometer data reveals that cooling ages are most sensitive to (1) location and magnitude of fault ramps, (2) variable shortening rates between 68-6.4 mm/yr, and (3) timing and magnitude of out-of-sequence faulting. The predicted ages are less sensitive to (4) radiogenic heat production, and (5) estimates of topographic evolution. We propose a revised cross section geometry that separates one large ramp previously proposed for the modern decollement into two smaller ramps. The revised cross section results in an improved fit to

20   observed ages, particularly young AFT ages (2-6 Ma) located north of the Main Central Thrust.

1 Introduction

Cooling ages recorded by thermochronometers are a direct function of the timing, magnitude, and rate of exhumation in fold thrust belts (e.g., Ehlers and Farley 2003; Shi and Wang 1987; Huerta and Rodgers, 2006; Rahn and Grassemann, 1999; McQuarrie and Ehlers, 2017). However, the rate and magnitude of exhumation may be strongly controlled by the geometry

25   and rate of deformation (Lock and Willett 2008, McQuarrie and Ehlers, 2015). Previous studies have shown that thermochronometers are most sensitive to the vertical motion of material, such as fault motion over a fault ramp, which focuses exhumation at that location (Whipp et al., 2007; Herman et al., 2010; Robert et al., 2011; Coutand et al., 2014; McQuarrie and Ehlers, 2015). Because of this, several hundred kilometers of horizontal shortening such as motion along a flat decollement, a phenomenon commonly observed in fold-thrust belts, may occur without a significant thermal cooling

signal (e.g., Batt and Brandon, 2002; Huntington et al., 2007; Whipp et al., 2007; Coutand et al., 2014). Thus potential variations in cross section geometry such as the spatial distribution of ramps, the order of faulting, and how fault and ramp positions change with time are predicted to have a significant impact on the exhumation history of fold-thrust belts.

The shape of subsurface isotherms and the cooling history of minerals are also controlled by the evolution of topography,
5   something that is largely unknown and often modeled either in steady state (e.g. Coutand et al., 2014; Herman et al., 2010; Whipp et al., 2007) or as a muted topography that increases relief with time (e.g. Erdos et al., 2014). The spatial and temporal changes in cooling rate due to topographic relief depend on topographic wavelength and amplitude, exhumation rate and duration, and the thermochronometer system recording the change (Ehlers and Farley, 2003; Braun et al., 2002; Mancktelow and Grasemann, 1997; Stuwe et al., 1994). Therefore the sensitivity of modeled cooling ages due to a
10 prescribed topographic evolution becomes a critical component in evaluating fold-thrust belt exhumation. In this study, we evaluate the effect of cross section geometry and kinematics as well as the effect of topographic assumptions on modeled cooling ages using a balanced geologic cross section and associated thermochronometer data from the Bhutan Himalaya.

2 Geologic background

The Himalaya orogen initiated with collision of the Indian Plate with the Asian Plate c. 50-55 Ma (e.g., Patriat and Achache,
15 1984; Yin and Harrision, 2000; Hodges, 2000). Estimates for the initiation of motion on the Main Central Thurst (MCT) in the fold-thrust belt range from ~25 to 20 Ma (e.g., Hodges et al., 1996; Daniel et al., 2003; Tobgay et al., 2012). The Himalayan fold-thrust belt is composed of south-verging structures and extends from suture zone in southern Tibet to the modern foreland in India. It comprises igneous, sedimentary, and metasedimentary rocks of Paleoproterozoic to Quaternary age (Gansser, 1964; Powell and Conaghan, 1973; LeFort, 1975; Mattauer, 1986; Hodges, 2000; DeCelles et al., 2002; Yin,
20 2006).

2.1 Tectonostratigraphy

The Himalayan fold-thrust belt is divided into four tectonostratigraphic zones that span much of the east-west extent of the orogen. From south to north, these are the Subhimalaya, Lesser Himalaya, Greater Himalaya, and Tethyan Himalaya (Fig. 1). All of these units were derived from rocks originally deposited on the Indian Plate (Heim and Gansser, 1939; Gansser,
25 1964). In the following section, we present the tectonostratigraphy and surrounding structures expressed along the Trashigang cross section of the Bhutan Himalaya (Fig. 1 and 2) (Long et al., 2011a; Long et al., 2011b).

The Subhimalayan zone is located north of the Main Frontal Thrust (MFT) and composed of synorogenic sedimentary deposits from the Himalayan foreland basin. In Bhutan the MFT emplaces a single thrust sheet of Miocene-Pliocene Subhimalayan units referred to as the Siwalik Group over modern foreland basin deposits (Gansser, 1983; Long et al.,
30 2011b).

The Lesser Himalayan zone consists of a package of Neoproterozoic to Permian strata, collectively grouped as the Upper Lesser Himalaya, and a suite of Paleoproterozoic strata comprising the Lower Lesser Himalaya (Long et al., 2011a).

The youngest unit of the Upper Lesser Himalaya, the Permian Gondwana succession, is exposed north of the Subhimalaya zone in the hanging wall of the Main Boundary Thrust (MBT) and in the immediate footwall of the thrust sheet carrying the stratigraphically older Permian Diuri Formation. North of these units, multiple fault-bound packages of the Neoproterozoic-Cambrian Baxa Group are repeated in the Upper Lesser Himalayan duplex. The Shumar Thrust (ST) exposed immediately to

5   the north is interpreted as the roof thrust of the system (McQuarrie et al., 2008; Long et al., 2011b).

In the hanging wall of the Shumar Thrust, the Paleoproterozoic Daling-Shumar Group overlain by the stratigraphically unconformable Neoproterozoic-Ordovician Jashidanda Formation. These strata are repeated multiple times to form the Lower Lesser Himalayan duplex with the Main Central Thrust (MCT) as the roof thrust (McQuarrie et al., 2008; Long et al., 2011b).

10   The MCT separates the southern Lesser Himalayan zone from Greater Himalayan zone located north of the MCT (Heim and Gansser, 1939; Gansser, 1964). The Greater Himalaya is divided into two structural levels: the lower unit is above the MCT but below the out-of-sequence Kakhtang Thrust (KT), while the higher unit is in the hanging wall of the KT (Grujic et al., 2002). Regional-scale antiforms and synforms mapped throughout the GH are interpreted to be a result of underlying Lesser Himalayan duplex formation (Long et al., 2011b).

15   **2.2 Thermochronologic Data**

The ^{40}Ar/^{39}Ar (MAr) and apatite fission track (AFT) data used in this study are from previous studies) and are presented with 2s analytical error (Supplementary Table 1) (Stüwe and Foster, 2001; Grujic et al., 2006; Long et al., 2012; Coutand et al., 2014). Previously published zircon (U-Th)/He (ZHe) data are determined from the mean age of replicates (typically 3 grains) and are presented in this study with a 2s error that encompasses the range in measured ages (Long et al., 2012). Cooling ages

20   are shown in map view in figure 1 and plotted versus distance along the Trashigang cross section in figure 2. We limit the data we evaluate to those within a 15 km distance from the cross section line. The AFT and ZHe data published by Coutand et al. (2014) north of the KT were projected 27 km 129° SE, maintaining the dataset's original distance between the KT and STD to evaluate modeled age predictions north of the KT (Fig. 1).

MAr data are published only for Greater Himalayan rocks in the hanging wall of the MCT and range from 14.1±0.4 to

25   11.0±0.4 Ma (Stüwe and Foster, 2001). The spatial extent of this dataset is limited to 9 km along the line of section from 52 to 61 km north of the MFT, including two cooling ages of 14.1 and 11.1 Ma from samples less than 0.5 km apart. The age range of these four MAr cooling ages is interpreted to represent cooling through the effective closure temperature of white mica at ~14-11 Ma.

The eight ZHe samples from Lesser Himalayan rocks that we use have cooling ages ranging from 11.6±0.1 to 7.3±0.8 Ma

30   along a 35-km across-strike distance (~17-52 km N of the MFT). These ages are interpreted to indicate rapid cooling through the zircon (U-Th)/He closure temperature at ~9.5-11.5 Ma (Long et al., 2012). North of the MCT, ZHe cooling ages are limited to two samples from the structurally higher Greater Himalayan and Tethyan rocks. These samples were reset at 7.4±1.6 Ma and 7.1±0.3 Ma respectively (Coutand et al., 2014, Long et al., 2012).

AFT cooling ages from the Lesser Himalaya are limited to four samples that range between 6.3±2.3 Ma to 4.2±1.0 Ma (Long et al., 2012; Grujic et al., 2006). In the structurally lower Greater Himalaya, AFT cooling ages progressively decrease from south to north from 7.8±2.8 Ma to 3.7±0.6 Ma (Grujic et al., 2006), with one young AFT age of 3.1±1.2 Ma directly north of the MCT (Stüwe and Foster, 2001)., The range in ages of six AFT samples from the structurally higher Greater Himalaya is

5   2.5±0.4 Ma to 4.2±0.8 Ma (Coutand et al., 2014). Due to the close spatial proximity of these cooling ages, we represent the AFT data from this region as one collective sample point that includes the spatial and temporal variability of the entire cluster when comparing the data to model results in the following sections. We apply the same approach for the cluster of three closely spaced AFT data from the immediate hanging wall of the MCT, 53 km north of the MFT (Fig. 2) (Stuwe and Foster, 2001; Grujic et al., 2006).

10   **3 Methods**

3.1 Flexural and Kinematic Model

Long et al. (2011b) published a balanced cross section in the Trashigang region of Bhutan (Fig. 2). We used the structural modeling software *Move* (Midland Valley) to sequentially deform the Trashigang section using slip amounts determined from the cross section. The cross section was deformed in ~10 km increments with accounting for isostatic loading due to

15   fault displacement and unloading due to erosion. The magnitude of isostatic load was determined from the difference between each increment of deformed topography and the topography of the previous step (McQuarrie and Ehlers, 2015). Erosional offloading was based the difference between the deformed, isostatically loaded profile and a new topographic profile generated at each deformation step (McQuarrie and Ehlers, 2015). Including isostatic response in the model produces a record of syn-deformational exhumation and deposition, facilitate the steepening of the decollement over time, and

20   develops a foreland basin.

3.1.1 Model Parameters

Effective elastic thickness (EET), crustal density, and initial decollement dip were adjusted so that the final modeled cross section was the best fit possible to the observed geology at the surface, foreland basin thickness (6 km), and decollement dip (4°) (Long et al., 2011b). We placed highest priority on matching surface geology. Young's modulus and mantle density

25   were held constant at 70 GPa and 3.3 g/cm^3 respectively. Best-fitting flexural models in this study used values of 65-70 km EET, correlating well with regional estimates (Jordan and Watts, 2005) and values of 40-70 km in Nepal (Berthet et al., 2013). Flexural model parameters are presented in Table 1 along with the kinematic and topographic variations used in each flexural model.

A two-dimensional grid of points spaced 0.5 km apart was distributed across the section and sequentially deformed with the

30   cross section to generate high-resolution displacement vectors describing how the kinematics of the system evolve in ~10 km

increments. The displacement field is converted into a velocity field that can be used in the thermal and age prediction model Pecube by assigning an age to each step.

3.1.2 Kinematic Variations Considered

Out-of-sequence thrusting along the KT occurred sometime between 14 Ma and 8 Ma, significantly more recently than

5 motion on the MCT (Davidson et al., 1997; Grujic et al., 2002; Hollister and Grujic, 2006). However, uncertainty remains regarding the magnitude and age of slip along the KT. Long et al., (2011a) argued for 31-53 km of minimum KT displacement. We tested three kinematic scenarios in the flexural models by varying the relative timing of KT motion, called the Early KT, Split KT, and Late KT models (Fig. 3). Early KT is modeled with 45 km of motion along the KT immediately following motion on the Shumar Thrust (Fig. 3c.1). In Split KT, out-of-sequence thrusting is modeled in two separate stages

10 with 25 km of motion applied after deformation along the Shumar Thrust, followed by 20 km of motion after Upper Lesser Himalayan duplexing (Fig. 3c.2). Late KT is modeled with 45 km of out-of-sequence thrusting after development of the Upper Lesser Himalayan duplex, similar to the proposed model of sequential deformation by Long et al. (2012) (Fig. 3c.3).

3.1.3 Topographic History Estimation

To model the isostatic response to deformation and erosion, we tested three different methods of estimating the topographic

15 evolution in Move. Each method was variable in profile resolution and in its ability to account for common factors of fold-thrust belt development such as deformation front migration, localized topographic uplift, and structural subsidence. The three topographic models were evaluated in the thermal-kinematic model to determine the sensitivity of the predicted thermochronometric data to each topographic scenario.

The "No Topography" scenario is the simplest of the three estimations with a topographic profile that remained at sea level

20 throughout the entire section reconstruction. We also tested a "Template Topography" scenario with a topographic profile similar to the modern topographic gradient of Bhutan (Duncan et al., 2003) that maintains a steep gradient in the first 25 km behind the active deformation front, followed by shallower gradient with elevation increasing along a two-degree slope to a maximum of 5 km. This shape of the Template Topography profile remains the same north of the deformation front, and the profile is spatially translated as the location of the deformation front is adjusted progressively southward throughout the

25 sequential development of the fold-thrust belt. A critical caveat to the Template Topography is that topographic elevations are not perturbed by isostatic loading. Thus the grid points in the model subside due to deformation-induced loading, but the topography does not. The third topographic model, "Python Topography", estimates a topographic profile for each flexurally loaded ~10-km deformational step using a Python-based computer script (Python Topography) (McQuarrie and Ehlers, 2015). New topography is generated using a user-defined slope (e.g., two degrees) where active uplift occurs, while in areas

30 without active uplift, the program follows existing topography. This approach allows topography to respond to deformational loading and erosional unloading. For models using the Python and Template topographies, the initial topography assigned to the restored section simulates a pre-existing fold-thrust belt in the Tethyan sequence before the initiation of the MCT

(Ratschbacher et al., 1994; Murphy and Yin, 2003; Webb et al., 2011). This topography maintains 0 km elevation from the southern end of the restored cross section to the Lower Lesser Himalaya. Across the Lesser Himalaya, topographic elevation increases to 5 km across a distance of 140 km and reaches a maximum elevation of 5 km above the lower Greater Himalaya, which at its southernmost extent is buried at a depth of 16 km below sea level.

3.2 Thermal and Cooling Age Prediction Model

The velocity field and topography for each increment of deformation, after displacement, isostasy, and erosion have been applied, is input into a University of Tübingen modified version of the thermal-kinematic predictive model Pecube (Braun, 2003; Whipp et al., 2009; McQuarrie and Ehlers, 2015). The thermal-kinematic model functions as: (1) a kinematic model that uses fault geometries and high-resolution point tracking inputs from the Move software to calculate rock transport (advection) velocities; (2) a transient thermal model that calculates the thermal field using fault motion, erosion above the topographic surface, rock thermophysical properties, and thermal boundary conditions; and (3) a set of age prediction algorithms (Ehlers et al., 2005) that calculate a suite of thermochronometer ages for material at the topographic surface for each deformation step using the thermal histories of particles as they are exhumed and cooled from depth to the model surface (e.g. Coutand et al., 2014; McQuarrie and Ehlers, 2015). The model predicts the cooling ages of rock samples that could hypothetically be collected from the surface at each deformation step, up to the modern deformed cross section geometry and topography.

3.2.1 Radiogenic Heat Production

We varied radiogenic heat production (A_O) to test the sensitivity of predicted cooling ages to variations in rock thermophysical properties. Calculated values of radiogenic heat production in the Himalaya are highly variable. A low heat production estimate of 0.8 $\mu W/m^3$ for the entire Indian Shield was calculated based on observed low in heat flow by Ray and Rao (2000), but other measurements have been estimated as high as 1.5-5.5 $\mu W/m^3$ due to the abundance of potassium, uranium, and thorium in granitic and gneissic rocks (Menon et al., 2003). Similar ranges of heat production values from 1.5 to 6.0 $\mu W/m^3$, with clustering around 4 $\mu W/m^3$, have also been found for Greater Himalayan rocks (e.g., England et al., 1992; Whipp et al., 2007). Herman et al. (2010) concluded a best-fitting heat production value of 2.2 $\mu W/m^3$ in their own thermal-kinematic model using a constant basal temperature of 750°C. In this study we tested models using radiogenic heat production values ranging from 4.0 to 1.0 $\mu W/m^3$. All thermal rock property parameters input in Pecube simulations are listed in Table 2.

3.2.2 Variable Deformation, Age And Rate

To compare the effects of differing age and rate of fault motion on predicted cooling ages, several deformation ages and velocities were tested. Details for combinations of velocities, heat production values, and flexural models are in Table 3.

A constant velocity of 17.3 mm/yr using the MCT initiation age of 23 Ma concluded by Long et al. (2012) was tested to determine if a generalized long-term rate of shortening can adequately reproduce published cooling ages. This rate is comparable to the ~15-25 mm/yr estimates of modern convergence for the Himalaya (Bilham et al., 1997; Larson et al., 1999; Banerjee and Burgmann, 2002; Zhang et al., 2004; Bettinelli et al., 2006; Banerjee et al., 2008) and long-term rates of

5   shortening through the Himalaya (DeCelles et al., 2001; Lave and Avouac, 2000; Long et al., 2011b).

However, in Bhutan, variable rates of shortening have been proposed based on the integration of shortening estimates from balanced cross sections with thermochronometer data. These rates range from as low as 4 mm/yr (Long et al., 2012) to as high as 75 mm/yr (McQuarrie and Ehlers, 2015). We evaluate two published variable deformation rate scenarios. Velocity model A is based on rates proposed by Long et al. (2012) along the Trashigang section with pulses of rapid deformation

10   during MCT motion (32 mm/yr) and the formation of the Upper Lesser Himalayan duplex (37-41 mm/yr), separated by slower periods of deformation during Lower Lesser Himalayan duplexing (15 mm/yr) and motion along the MBT and MFT (4-6 mm/yr). In velocity model B, MCT motion initiates at 20 Ma and at slower velocity (21 mm/yr). This younger and slower deformation requires faster motion for the subsequent duplexing of the Lower Lesser Himalaya (22-25 mm/yr) and Upper Lesser Himalaya (69-75 mm/yr). Other rates of motion in this scenario are comparable to velocity model A.

15   McQuarrie and Ehlers (2015) propose similar younger fold-thrust belt initiation and fast rates of shortening during Upper Lesser Himalayan duplexing (75 mm/yr) based on thermal-kinematic models they conducted of the Kuru Chu section in Bhutan.

4 Results

4.1 Flexural Model

20   Summaries of the final output of all seven flexural models to the published Trashigang cross section (Long et al., 2011b) are presented in Table 1; supplemental figure 1 contains images of the results of each model. Because the flexural models control locations and magnitudes of erosion and burial that are input into the thermal model, in this section we describe the sensitivity of the vertical path of points and the resulting geology that is exposed at the surface to the estimated topographic evolution, different proposed kinematics, and total magnitude of subsidence, which is illustrated by the final shape of the

25   decollement.

The difference between the modeling results are subtle but show local variations in total erosion of 0.5-4 km that is reflected in the final geology exposed at the surface of the model and depth to stratigraphic markers within the model. All models produced foreland basin depths within 2 km of the estimated 6 km thickness. Average decollement dips varied from 3.75° to 5.4°. Six out of seven modeled decollement angles are steeper than the 4°N dip along the footwall flats of the published

30   section, based on INDEPTH reflection seismology (Hauck et al., 1998). However six out of seven models are still within the 4-6° angle estimated for the Main Himalayan Thrust (Ni and Barazangi, 1984; Hauck et al., 1998; Pandey et al., 1999; Mitra et al., 2005; Schulte-Pelkum et al., 2005).

Each of the different kinematic scenarios produced different flexural responses (Table 1). Models using Late KT deformation produced the deepest foreland basins and steepest decollement dips, along with under-eroded geology at the surface compared to the published section. These results are a function of out-of-sequence thrusting producing a greater magnitude of late-stage loading in the hinterland of the models. Early KT and Split KT scenarios have decollement dips

5 shallower than Late KT models and result in a better match the surface geology data, except when using No Topography. Differences among Early KT and Split KT decollement dips and surface geology are not systematic, indicating these differences are less driven by kinematics and appear to be more sensitive to slight variations in flexural isostasy parameters and the profile of the topographic load. The poorest fit to surface geology was produced by the model combining Split KT with No Topography (Supplemental Fig. 1). In all other model combinations, exposed geology is within 1 km of the modern

10 geology observed at the surface, with particularly good fits combining Early KT deformation with the Template Topography, and Split KT with Python Topography.

Topographic profiles from the final deformation step of each model vary in fit and misfit to observed topography along the Trashigang line of section (Fig. 4). The sea-level No Topography profile is obviously the worst fit of the three estimations. Template Topography fits the steep topographic rise from the MFT to the southern trace of the Shumar thrust; however to

15 the north of the Shumar Thrust, estimated elevations are ~1 km greater than observed. Python Topography provides a better fit for the northern half of the section, including a local drop in elevation from 77-90 km along section north of the MFT. However, the average 2° slope assigned to the Python Topography profile resulted in under-predicted elevations from 13-55 km, where the average observed topographic slope of the range is steeper (4.5°). Overall, Python Topography best reproduces the observed topography along the cross section.

20 For models using Python and Template topography, we attribute the differences in decollement dip to the lower topographic relief produced using a 2° angle with Python Topography, compared to the steeper topographic angle near the deformation front and overall higher elevations with Template Topography (Fig. 4). The shallower topography from the Python Topography requires a steeper decollement to accommodate the same amount of material within the tapered wedge while matching the surface geology.The most significant result of the flexural model was identifying the relationship between

25 uplift or subsidence of rock (as represented by the two-dimensional grid of points) and the uplift, subsidence or static position of topography. The static profile used when modeling with Template Topography or No Topography can result in regions of non-erosion and burial (with respect to the topographic surface). When the deformation front shifts toward the foreland, higher topography is translated southward with no direct relationship for where structural uplift is occurring. Additionally, material (as represented by the grid) will subside in areas responding to flexural loading while topography does

30 not. This latter example is especially relevant south of the Kakhtang Thrust during out of sequence thrusting. While using Python Topography, both points and topography subside in front of the Kakhtang Thrust, which allows for minor amounts erosion to occur across the entire section during fault motion. Using the Template Topography, points subside due to the imposed load but topography does not which simulates burial in this region. Thus the Template Topography disconnects the

topographic evolution from the kinematic and flexural evolution by not accounting for structural uplift and subsidence. The thermal consequences of the different flexural models are explored in Section 4.2.

4.2 Predicted Cooling Ages Across the Cross-section

4.2.1 Effect of Heat Production and Constant Shortening Velocity on Predicted Ages

5   By holding velocity constant and testing multiple values of radiogenic heat production in Pecube, we can discern the effect that adjusting radiogenic heat production may have on the output of predicted cooling ages as well as the viability of a constant rate of shortening with time. We compare predicted cooling ages for AFT, ZHe and MAr systems to published ages using a range of surface radiogenic heat production (Ao) values (Fig. 5). The kinematic input is from the flexural model combining Split KT and Python Topography, coupled with the constant velocity of 17.3 mm/yr from 23 Ma to the present.

10   The most apparent trend among all three thermochronometer systems is that predicted cooling ages become younger as the surface heat production increases from 1.0 to 3.0 $\mu W/m^3$. For example, modeled MAr ages remain unreset south of the Kakhtang Thrust (Fig. 5a) when Ao is 1.0 $\mu W/m^3$, while output with higher Ao of 3.0 $\mu W/m^3$ yields MAr ages as young as 5.8 Ma in the Upper Lesser Himalayan (Fig. 5a). The difference in predicted reset MAr ages in structurally higher Greater Himalaya rocks north of the Kakhtang Thrust is 3-7 Myr between simulations with Ao values of 1.0 and 3.0 $\mu W/m^3$. Similar

15   differences in cooling ages are observed in ZHe and AFT plots as well (Fig. 5b,c). An important result is that although the predicted ages change values, the shapes of the cooling curves remain broadly the same. This is readily seen in the pattern of predicted ZHe and AFT ages across the structurally lower Greater Himalaya between ~60 and 80 km north of the MFT. Cooling ages predicted in this region highlight the transition from rocks recently transported over an active ramp to rocks that have not been transported over the ramp; predicted cooling ages south of the ramp are notably younger than rocks to the

20   north. The trend of older predicted cooling ages 65-85 km north of the MFT forms an upside-down U shape in the Greater Himalaya section between the Main Central and Kakhtang thrusts regardless of surface radiogenic heat production value (Fig. 5c).

Evaluating the fit between measured and modeled ages predicted by the different thermal models indicates that the best match to published AFT ages is with an Ao of 1.0 $\mu W/m^3$, Even with this cool thermal parameter, we find that predicted ages

25   are still to young to fit published AFT and ZHe ages but significantly too old to match published MAr data. These simultaneous over- and under-estimations of published ages requires that rates of deformation and exhumation are more complex.

4.2.2 Effect of Shortening Velocity Variations on Predicted Ages

A constant rate of deformation described in the previous section does not produce cooling ages that match all three

30   thermochronometer systems (Fig. 5). In this section we present modeled cooling ages from two variable velocity schemes that are compared to published cooling ages: velocity model A (Long et al., 2012), and velocity model B (McQuarrie and

Ehlers, 2015) (Table 3). All variable velocity models presented in this section used a heat production value of 2.5 μW/m³, with the flexural model combining Split KT and Python Topography as input.

Velocity model A produces a noticeably better fit than a constant deformation rate. Modeled MAr ages are 2 Myr older that the 2 oldest published ages (14.1 Ma), and predicted cooling ages are within error of 16 out of 28 published AFT and ZHe

5   cooling ages (57% fit) (Fig. 6a). Predicted AFT ages fit 7 out of 15 published samples (47% fit), with no fit to AFT data from 70 km to 120 km. Predicted ZHe ages match 9 out of 13 samples (69% fit). Predicted MAr ages are reset during Lower Lesser Himalayan duplex formation and motion on the Shumar Thrust, which ceases activity at 15 Ma for velocity model A. The MAr system records cooling during Lower Lesser Himalaya duplexing, regardless of its age or rate.

A faster rate of deformation (37.3 mm/yr) than in the constant velocity model during the formation of the Upper Lesser

10   Himalayan duplex produces older ages (~10 Ma) and a better match to published ZHe data (Fig. 6a). As the final two horses of the duplex are stacked along a large decollement ramp (Fig. 3c), the increased magnitude of uplift and erosion rapidly cools rocks through the ZHe closure isotherm (Fig. 6a). Younger ZHe ages predicted 50-55 km north of the MFT appear to be a function of heightened local erosion as the rest of the Upper Lesser Himalayan duplex and the overlying hanging wall of the Shumar Thrust are carried over this same decollement ramp during MFT motion.

15   Samples from the Upper Lesser Himalaya cool through the AFT closure isotherm after out-of-sequence motion on the Kakhtang Thrust and rapid deceleration in deformation rate from 37 mm/yr to 6 mm/yr at 10 Ma. Both the out-of-sequence thrusting and the slower deformation rate create a prolonged timeframe for AFT ages to cool, causing scatter in predicted ages that range from 10 Ma to 4 Ma (Fig. 6a). These ages are a result of slow, uplift-induced exhumation as subsequent motion along the MBT moved these rocks up and over the MHT ramp located 65 km from the MFT. Predicted AFT ages in

20   the Greater Himalaya systematically increase north of the ramp again, similar to the pattern observed with the constant velocity output. The older predicted AFT ages located 65-85 km from the MFT cool much earlier in the deformation history when the fold and thrust belt was lifted over a ramp in the Lower Lesser Himalayan during early stages of Upper Lesser Himalayan duplexing (3c.2).

Velocity model B uses an earlier MCT initiation at 20 Ma and a rate of Upper Lesser Himalaya duplexing that is twice the

25   rate used in velocity model A (Table 3). Despite this difference, fits to published data are remarkably similar to velocity model A, with a marginally improved fit (Fig. 6b). Predicted MAr ages produce a better match to published data due to a younger age for the growth of the Lower Lesser Himalaya duplexing: 17-13.5 Ma with velocity B versus 20-15 Ma with velocity A. Using velocity B also predicted older and better-fitting modeled ZHe data across the Upper Lesser Himalaya (11-10 Ma) due to faster and earlier duplexing, which ends 11 Ma in this scenario versus 10 Ma in velocity A. Ten out of 13 ZHe

30   ages (76%) and to 8 out of 15 AFT ages (53%) are reproduced within error. Similar to velocity model A, predicted AFT ages remain too old in the GH zone from 70 km to 100 km north of the MFT.

Overall the adjustments to the timing and rates of deformation used in velocity model B result in the best fit to published thermochronometer data out of the tested velocity models for this cross section. Velocity model B also encompasses the

greatest variation in shortening rate over the time of fold-thrust belt development, from 75 to 5.5 mm/yr (Table 3). However, there is still a large discrepancy between predicted and measured AFT ages across Greater Himalaya.

4.2.3 Effect of Topographic Development on Cooling Ages

We evaluate the sensitivity of the predicted thermochronometer ages to different topographic development scenarios using the Split KT kinematic scenario and velocity model B. The resulting predicted ages for different thermochronometer systems are shown in figure 7. Different methods of estimating the evolution of topography have a much smaller effect on the predicted cooling ages than adjustments to deformation velocity, heat production or geometry (Section 4.2.5).

The No Topography model generated predicted ages that are remarkably similar to the Python Topography. In detail, No Topography yields identical or slightly older (0.5 to 3 Myr) predicted ages than the Python Topography, with greatest difference in the predicted AFT ages. This is in contrast to initial expectations that the over-eroded No Topography model would produce younger cooling ages than the other topographies (Table 1, Supplementary Fig. 1).

Results from the Template Topography versus Python Topography models show greater differences in predicted cooling age trends. In ZHe and MAr plots, the largest difference between the models is the spatial width of the reset cooling ages. For example, the Python Topography model fully reset MAr ages start at 33 km from the MFT while the Template Topography model predicted ages are reset at 36-37 km from the MFT. For predicted ZHe ages, reset ages from the Python Topography model start at 10 km north of the MFT while reset ages from the Template topography do not appear in results until 10 km farther north.. The wide discrepancy in predicted AFT ages is a function of a high degree of scatter in the predicted ages from the Template Topography model. Between 35 and 90 km, these ages range from 13-3 Ma without any pattern, except for directly over the ramp at 55-65 km from the MFT. This highly irregular cooling history is a function of the topography not accounting for localized uplift, erosion, or structural subsidence with time. Static topography inaccurately models the burial of material where particle points are subsiding and topography is not subsiding, and produces over-erosion of material where particle points experience rock uplift but topography remains static. These results highlight the need for estimates of topographic evolution that account for areas of structural uplift and subsidence during fold-thrust belt evolution.

4.2.4 Influence of Kinematic Variation on Cooling Ages

Changes to the prescribed kinematic order used in forward modeling the cross section were tested using flexural models with Python Topography, coupled with velocity model B and a surface heat production of 2.5 μW/m^3 in the thermal-kinematic model. Because different thrust structures have different slip magnitudes, it is not possible to have precisely the same velocities with different kinematics. To most closely evaluate the effect of kinematic variations in out-of-sequence thrusting, we kept the age at which velocities change the same whenever possible. Predicted cooling age output for Early KT and Late KT kinematic scenarios are plotted in figure 8 and compared with results from the same Split KT scenario used in Section 4.2.3.

Modeled fits to published MAr data appear to be unaffected by changes to the timing of out-of-sequence thrusting, with all flexural models predicting cooling ages of ~14 Ma in the hanging wall of the MCT. This is expected because all changes to out-of-sequence thrusting occur after the formation of the Lower Lesser Himalaya duplex, when most of these particles have cooled through the MAr system isotherm.

5   Each of the different kinematic scenarios predicted significantly different ZHe ages across the Upper Lesser Himalaya duplex, implying that there is a particular kinematic order of deformation required in the flexural model to generate the measured cooling ages. The pattern of predicted ZHe ages between 10 and 65 km from the MFT is controlled by age and rate of displacement of the Upper Lesser Himalaya duplex, the final step of which places duplexed Baxa units over younger rocks on a ramp in the MHT (Fig. 3d). This last step structurally elevates the entire duplex and increases local exhumation.

10 Continued motion of the duplex over this ramp cools the rocks through the AFT system. In the Split KT kinematic model, displacement over this ramp occurs at 11 Ma just before the second stage of motion on the Kakhtang Thrust. In the Early KT and Late KT models, Upper Lesser Himalaya duplexing is immediately followed by motion of the duplex over this ramp between 10 and 7 Ma, after a marked decrease in velocity at 10-11 Ma (Table 3). The altered timing of this displacement results in young (7-10 Ma) ZHe ages and AFT ages. Compared to Split KT model results, the younger ZHe ages predicted

15 are a poorer fit to published data at 10-35 km from the MFT.

The 4-5 Myr gap between published ZHe and AFT data in this area of the cross section is only reproduced using the Split KT kinematic model (Fig. 6). The second stage of out-of-sequence thrusting in this model postdates the development of the Upper Lesser Himalaya duplex but predates motion of the duplex over the ramp of younger rocks, causing a 4-5 Myr delay between these two processes that focus exhumation in the Lesser Himalaya.

20 Though the magnitude and timing of out-of-sequence thrusting impacts the timing of cooling across this area, none of these three kinematic scenarios adequately reproduced any AFT data from the GH north of ~70 km from MCT (Figs. 6 and 8). The fit of predicted ages to published AFT data across the Upper Lesser Himalaya 10-30 km north of the MFT is similar across all three models. This is due to similar ages and rates of fault motion along the MBT and MFT from ~7.3 Ma to the present, when Upper Lesser Himalayan rocks cool through the AFT closure isotherm in the models. Predicted ages in the Greater

25 Himalaya north of the MCT vary primarily with regard to the amplitude of the upside-down U shape from ~65 to 85 km north of the MFT (Figs. 6 and 8). In this area the predicted AFT ages were set during Upper Lesser Himalaya duplexing, when GH material is carried over a ramp in the MHT. Cooling ages were subsequently modified by burial in the immediate footwall of the Kakhtang Thrust. Ages predicted using Late KT and Split KT kinematic scenarios are older than Early KT model predictions because subsidence postdates AFT closure in Late KT and Split KT models. The strong gradient from

30 recently reset AFT ages predicted 50-70 km north of the MFT to earlier AFT ages north of 70 km is a function of the prescribed geometry of the Main Himalaya Thrust, which cannot reproduce the AFT signal in the Greater Himalaya north of 70 km.

4.3 Influence of Cross-Section Geometric Variations on Ages

Because geometry of the cross section has a first-order control on ages (e.g., McQuarrie and Ehlers, 2015), we predict that the solution to the difference between observed and modeled thermochronometer ages is in the geometry of the cross section. In results from the best fitting thermal-kinematic model using current cross section geometry (Fig. 6), Greater Himalaya

5   rocks are carried over two active decollement ramps: (1) ~125-145 km north of the MFT, the footwall cutoff of the Daling formation during Upper Lesser Himalaya duplexing; and (2) ~60-70 km north of the MFT, the footwall cutoff of Baxa and Diuri units during MFT motion. Motion over the footwall cutoff of the Daling formation facilitated AFT cooling in the model from 13-11 Ma, too early to produce the measured ages of 3-6 Ma, while more recent motion over the footwall cutoff of Baxa and Diuri units set the young AFT ages modeled at 50-65 km along section. We hypothesized that modifying the

10   locations of decollement ramps in the original geometry would result in an improved model of the young observed AFT ages across the Greater Himalaya.

In a new, modified version of the Trashigang cross section, the decollement has been adjusted to partition the large ramp cutting through the Diuri and Baxa units of the Upper Lesser Himalaya into two separate ramps (Fig. 9). The footwall cutoff of the Diuri has remained in its same position along the decollement, but the footwall cutoff of the Baxa unit has been shifted

15   ~35 km north to the northern end of the Lower Lesser Himalayan duplex, ~95 km north of the MFT.

Two flexural models with this new decollement geometry were tested in Pecube, one created using Python Topography and another using No Topography to highlight the sensitivity of the cooling signal to the prescribed topographic and thus erosional history (Table 1, Supplementary Fig. 1). Both models use the Split KT kinematic scenario, modified to accommodate updated magnitudes of displacement in the flexural model. Multiple velocity and heat production

20   combinations were coupled with these models in Pecube. Predicted ages from these two flexural models are presented and compared in the following section using an updated velocity model called velocity model C (Table 3).

4.3.1 Fit of Predicted Cooling Ages Using Modified Geometry with Python Topography

The modified geometry model using the Python Topography resulted in a noticeably different and better-fitting predicted age trend in the region north of the MCT (Fig. 9a). Unlike previous models that used a uniform surface heat production (Ao)

25   value, the best fit was achieved using a higher surface heat production of 4.0 $\mu W/m^3$ in the region of exposed Greater Himalaya rocks and a lower 2.0 $\mu W/m^3$ for Lesser Himalaya rocks, divided by the MCT. Using different values for radiogenic heat production is consistent with previous studies that have noted the higher heat production capacity of Greater Himalaya rocks which cluster around 4.0 $\mu W/m^3$, while Lesser Himalayan rocks have a lower average heat production value of 2.5 $\mu W/m^3$ (Roy and Rao, 2000; Menon et al., 2003; England et al., 1992; Whipp et al., 2007; Herman et al., 2010).

30   Using this higher heat production from the trace of the MCT northward produced predicted ages that matched measured ages for both the MAr and ZHe systems. Perhaps most critically, the modified geometry provides an improved fit and matches 10 out of 12 AFT data from 53 to 120 km north of the MFT (83% fit). The notable difference with the new geometry is that the

"U" shape in the immediate footwall of the KT is narrower (15 km across) and lower amplitude (3 Ma). However as observed with the original geometry, although the values of the predicted ages are older with lower Ao values, the first-order shape of the predicted ages remains the same (Supplementary Fig. 2). Thus using the same heat production value as the Lesser Himalaya rocks (2.0 µW/m^3) through the entire section, the predicted trend of AFT cooling ages that young

5   northward between the MCT and KT is broadly 3-4 Ma older but retains a similar shape. This indicates that the trend of cooling ages in this area is a function of the subsurface geometry of the decollement ramp.

South of the MCT, predicted thermochronometer ages do not have as strong of a match to the measured ages as with the previous best fit model (Python topography, Split KT, velocity B; Fig. 6b). The revised geometry matches half of the measured ZHe ages in the Lesser Himalaya using the lower heat production of 2.0 µW/m^3 and fits all three published AFT

10   ages within error. Similar to changing the kinematics, changing the cross section geometry alters the prescribed rates because the magnitude of shortening has changed (Table 3). In the best fit model of the original geometry, the Upper Lesser Himalaya duplex deforms quickly from 13 to 11 Ma. However in the revised geometry model, this duplex initially deforms quickly from 13-11 Ma, but the rate of deformation during the last 53 km of duplex formation is slower from 11-7.4 Ma. This slower rate and younger age limits the southern extent of the reset ZHe ages at the surface and results in slightly

15   younger (8.5 to 10 Ma) predicted ZHe ages between 20 and 30 km from the MFT.

4.3.2 Fit of Predicted Cooling Ages Using Modified Geometry with No Topography

The output of modeling with No Topography is similar to that with the Python Topography, with the largest differences in predicted thermochronometer ages in lower temperature systems (Fig. 9b), as observed in section 4.2.3. Using the higher heat production of 4.0 µW/m^3, predicted ZHe ages are ~1 Myr older than the Python Topography predicted ages

20   for the region north of the MFT indicating earlier erosion of this region in the No Topography model. ZHe ages predicted through the Lesser Himalaya with 2.0 µW/m^3 heat production are remarkably similar to the predicted Python Topography, as identified in previous models (Fig. 8b) and again emphasize the control of rates and geometry on the cooling ages. In addition, the revised geometry using No Topography does not preserve the 4-5 Myr gap between ZHe and AFT data from the Upper Lesser Himalaya, similar to previous No Topography models (Fig 8c). The higher total erosion in the front of the

25   system (Supplementary Fig. 1, F and G) lead to earlier erosion and older predicted AFT ages for the No Topography model. While the final flexural output and the predicted cooling ages for these two models are similar (Fig. 9, Supplemental Fig. 1), the intermediate timing and magnitude of exhumation and burial is again affected by the prescribed topographic scenario and ultimately affects the timing of mineral cooling.

5 Discussion

5.1 Evaluating the sensitivity of predicted cooling ages

Each component in the kinematics of a fold thrust belt system imparts a characteristic cooling pattern to the predicted ages at the surface. Emplacement of a large thrust sheet as with motion of the MCT imparts a pattern of reset cooling ages that is the oldest at the thrust tip and decreases towards the active ramp (Lock and Willett, 2008; McQuarrie and Ehlers, 2015). A southward growing duplex will produce a pattern of cooling ages that young toward the south (Lock and Willett, 2008; McQuarrie and Ehlers, 2017). While rocks record cooling associated with every stage of structural evolution, the events that are recorded by any given thermochronometer system are dependent upon the magnitude of exhumation associated with each component of deformation and the thermal history of the rocks: length and magnitude of burial, speed of exhumation, and heat production. If the magnitude of exhumation is particularly close to that necessary to reset a thermochronometer system, the predicted pattern of cooling ages can be significantly altered by small changes in modeled topography or heat production. For example, minor changes to the prescribed topography or thermal parameters can shift the signal of preserved AFT ages to record the southward propagation of a duplex versus the displacement of material over a decollement ramp (which would have a northward propagating signal) when the magnitude of exhumation associated with the decollement ramp is small (Supplementary Fig 2). Below we discuss the effects of different topographic models, topographic evolutions, and thermal parameters on cooling ages predicted in Pecube.

5.1.1 Sensitivity of predicted cooling ages to the prescribed topographic evolution

Although our evaluation of different topographic models indicates a modest sensitivity in predicted cooling ages to how topography is estimated, modeling an evolving topography such as a topographic slope that either increases or decreases with time can significantly change the predicted pattern of cooling ages by controlling the magnitude of erosion that occurs at a given time and the exhumation event during which a thermochronometer is reset. This marked change in the pattern of cooling, such as recording an older southward propagating duplex rather than a northward migrating ramp, occurs if the magnitude of exhumation associated with an exhumation event is small (~2 km or less) and close to the amount necessary to reset a given thermochronometer. In the case of the young AFT ages across the Greater Himalayan (Fig. 9), the northward younging trend is imparted by recent motion over a decollement ramp that must be north of the youngest age. However, this ramp spans a vertical distance of 2.5 km, half the height of other decollement ramps farther hinterland such as the ramp through the lower Lesser Himalayan. The smaller magnitude of vertical uplift and exhumation associated with this ramp makes cooling ages associated with it more sensitive to changes in other parameters.

The most basic requirement to reproduce observed cooling ages is to match the timing of exhumation with the structures that are producing the across strike exhumation pattern (McQuarrie and Ehlers, 2015). In our best fitting flexural and thermal model combination (Figs. 9a and 10a), ~2.0-3.5 km of exhumation, from 6 Ma to the present, was required to match the AFT ages that decrease in age from 6-3.5 Ma 65-90 km north of the MFT. To simulate this magnitude of exhumation

following the decrease in topographic elevations south of the KT due to KT loading, the prescribed topographic taper angle was reduced from 2.0 to 1.5 degrees during MBT and MFT motion, with a maximum elevation of 3 km modeled in the final cross section. The magnitude and timing of this exhumation was critical to generate cooling ages across the Greater Himalayan that recorded the signal of recent motion over the decollement ramp and fit the published data (Section 5.4).

5 Other flexural models evaluated in this study, did not predict cooling ages across the lower Greater Himalayan that matched published thermochronometer ages despite using the same geometry, kinematics and thermal properties. The difference in predicted and observed ages were a function of both slightly different topographic evolution scenarios that control magnitudes and timing of erosion and slightly different elastic thickness parameters that control the amount of subsidence. For example, the flexural model results that was used to generate predicted ages shown in figure 10b used an initial taper

10 angle of 2° and an EET of 75 km compared to the best fit model that used an initial taper angle of 2° and an EET that increased from 60 km early in the deformation history to 85 km for the second pulse of motion on the KT and displacement on the MBT and MFT. Higher EET values early in the modeled deformation steps facilitated more erosion (0.5 to 1.5 km) between 17 and 8 Ma, thus resetting AFT ages at this time (Fig. 10b versus 10a). In addition, the model displayed in figure 10b used a steeper topographic angle in the immediate foreland (3°) and a 2° angle in the hinterland to more closely match

15 modern topography and the exposed geology (Fig. 10b, Supplementary Fig. 1h). This steeper topography resulted in less erosion from 8 Ma to present (~0.3-1.5) than the best fit model (Fig 10a, Supplementary Fig. 1f), however the predicted surface geology of both models is almost identical. The change in the exhumation history between the two models, although minor, produced a markedly different pattern of cooling ages at the surface between 55 and 85 km from the MFT. The model with 0.5 to 1.5 km of additional exhumation early in the model history produced cooling ages that record the signal of

20 an older duplexing with southward younging of cooling ages (Fig. 10b).

Although matching the geology exposed at the surface is a critical test to evaluate the accuracy of the flexural model, we were able to match the measured AFT data, with a predicted AFT age pattern using a flexural model that is under eroded in the hinterland between 55 and 85 km from the MFT (Supplementary Fig. 1i). Structurally lower Greater Himalayan material was under-eroded by ~2-3 km in the final step of the flexural model with Tethyan material preserved at the surface that

25 presently should not exist along the Trashigang section (Figs. 1 and 2). Similar to the best fitting model, topography maintained a 2° taper until ~8 Ma, however the EET was 65 km. The lower EET allowed for more subsidence in the hinterland and thus less total erosion. From 6 Ma to present EET was increased to 70 km and the topographic angle was reduced to 1.75°. The stronger EET facilitated less subsidence particularly in the Greater Himalayan region. From 6 Ma to present the hanging wall of the MCT underwent 2.0- 2.5 km of exhumation (similar to our best fit model) and thus produced

30 similar predicted AFT ages with the same pattern as observed data (Fig. 10). In summary, the AFT ages are the most sensitive to the evolution of topography and small changes (0.5-1.5 km) in exhumation magnitude as expected. Although the change in topographic taper angle from 2° to 1.5° may account for up to 0.5 km of exhumation, small, 5 km changes in EET which control the amount of subsidence had a larger effect (~ 1 km) on the age and magnitude of exhumation.

5.1.2 Synthesis of the Effect of Thermophysical Properties on Cooling Ages

Altering the thermal history of the model by imparting a hotter or colder thermal field can also result in different cooling signals preserved at the topographic surface if the exhumation amount is close to a particular closure temperature for a thermochronometer system. For instance, the best fitting model run with exclusively a 2.0 µW/m^3 surface heat production

5   (Ao) value (Supplementary Fig. 2a), predicted AFT ages are reset at ~7-11 Ma from 65-85 km north of the MFT (Supplementary Fig. 2a), with a slight trend of output becoming younger toward the north across the structurally lower Greater Himalayan. The pattern of these cooling ages, particularly between 75 and 90 km north of the MFT, is recording a signal of older structural uplift instead of recent motion over the decollement ramp. The difference in AFT cooling ages between surface heat production (Ao) values of 4.0 and 2.0 highlight that magnitude of exhumation in this model is just at

10   the amount necessary to reset AFT ages. Thus, rocks at the surface are at a critical thermal threshold where, when exhumed through a low thermal gradient , the particles will preserve a different age pattern, than when exhumation through a higher thermal gradient.

5.2 Using Thermochronology to Evaluate Structural Geometry

Using traditional geologic and geophysical constraints to create balanced cross sections can often result in multiple

15   interpretations of the subsurface geology with significant variations in proposed subsurface structures, decollement ramp locations, and total shortening estimates. While kinematic reconstructions of balanced cross sections can help in determining the viability and kinematic sequence of a cross section, thermochronometer data can offer additional insights into predicting subsurface geometry. The geometry of the subsurface and location of ramps in the decollement impart a first order control on the thermochronologic trends present at the surface (Robert et al., 2011; McQuarrie et al., 2014; Coutand et al., 2014;

20   McQuarrie and Ehlers, 2015, 2017). In this study, Pecube output from two decollement geometries of the Trashigang cross section were compared, and an additional ramp in the MHT resulted in a noticeable change in cooling ages modeled. This finding is particularly evident in modeled AFT ages across the Greater Himalaya. Modeled AFT ages across the Greater Himalayan using the original cross section geometry reflected a cooling signal imparted by a larger ramp through the Lower Lesser Himalayan that did not fit the trend of published data (Fig. 6b). Even with a higher heat production assigned to the

25   models, the location and magnitude of this cooling signal did not change (Fig. 5). Modeling the modified geometry with an additional decollement ramp, resulted in a different pattern of predicted ages that better matched the trend of published data (Fig. 9). Applying a higher radiogenic heat production facilitated an improved fit to the data. By adjusting the decollement and creating vertical uplift and additional erosion in this area, we were able to successfully model the previously unmatched data across the lower Greater Himalayan.

5.3 Estimates of Timing and Rates of Deformation

The thermal and flexural model combination that best fit published thermochronometer data used rates of deformation that ranged from 68.4 to 6.7 mm/yr (Table 3). Our results provide estimates on the timing of motion along discrete structures present in the Himalayan fold-thrust belt of eastern Bhutan.

5 Our results are insensitive to MCT motion as early as 23 Ma or as late as 20 Ma due to the limited location of MAr data available along the Trashigang section. However, the initiation and rate of displacement of the MCT control the predicted MAr ages between 60 to 90 km from the MFT (Fig. 6) in the location of the Saktang Klippe (Fig. 1and 2), and thus provide direction for future research that could confirm predicted ages and rates. In our best fitting model, MCT motion occurred from 20 Ma until 18 Ma, at a rate of 29 mm/yr.

10 The transition from lower LH to upper LH duplexing in our best fit velocity model of the Trashigang section occurs at 13.5-13.0 Ma suggesting formation of the lower LH duplex from ~18-13 Ma. The 11.1-14.1 Ma ages of MAr samples directly above the lower LH duplex provide the age constraints (Fig. 9). Our predicted MAr ages are closer to the younger end of the range of data present along the Trashigang section (12-14 Ma), suggesting that lower LH duplexing can not continue longer than 13 Ma. The 9-11 Ma ZHe signal present in the upper LH, requires an end to rapid shortening by 11.0 Ma to fit the

15 measured cooling ages. As shown with our revised geometry, extending shortening of the LH duplex till 7.5 Ma predicts ZHe ages that are slightly younger than the measured ages. The limited window of time (13-9 Ma) and high magnitude of shortening (146 km) requires a fast shortening rate (68.4 mm/yr) while the upper LH duplex forms. The younger predicted ages provide a better match to the ZHe ages measured directly to the west along the Kuri Chu valley (Long et al., 2012; McQuarrie and Ehlers, 2015). The higher degree of scatter when both datasets are plotted together (Fig. 9) may reflect actual

20 variation in the uncertainty of the data versus differences in cooling between the two sections.

Similar to previous studies (Long et al., 2012; McQuarrie and Ehlers 2015) our results also suggest a slowing in the shortening rate to 14 mm/yr or less sometime between 11 and 9 Ma the shortening rate maybe as low as 5.4 mm/yr during MBT and MFT motion from ~6 Ma to the present to fit AFT data.

The age and rate of deformation from the best-fit model of the Trashigang region has noted similarities and differences to

25 other thermokinematic models of the area. While we found pronounced variation in shortening rates and magnitude of rates that are very similar to those presented by McQuarrie and Ehlers (2015), the window of time of rapid shortening (13-11 Ma; this study) does not overlap with the period proposed for the Kuri Chu section (11-8.5 Ma) immediately to the west (McQuarrie and Ehlers, 2015). The difference in the windows of rapid shortening is a result of the difference in ZHe ages between the two regions. In the Kuri Chu region 25 km to the east, the ZHe ages cluster between 8.5 and 10 Ma. Along the

30 Trashigang section (this study), the ages range from 9.5 to 11.5 Ma (Figs. 1, 9). If the lower LH duplex in the Trashigang region continued to ~12 Ma with a timing and rate of deformation more similar to those proposed by McQuarrie and Ehlers (2015), including a younger age in which the MHT slows (7.4 Ma versus 11 Ma) then the predicted ZHe ages are younger than the observed ages. The impact of velocity variations plotted in figure 6 highlight the change from an 11 Ma decrease in

shortening rate versus a 10 Ma decrease. The difference in age of the fast rates between the Kuri Chu and Trashigang sections suggests that 8.5-11.5 Ma is the acceptable range in ZHe ages for both sections (Fig. 9). Thus revised thermokinematic models using this larger range of cooling ages for both regions would allow for matching velocities between the cross-sections. Even though the details of the rates would change in both sections, general trends would remain

5 the similar such as: slow velocities between ~18-12 Ma, fast (60-70 mm/yr) velocities between ~12-9 Ma, and slow velocities from ~9 Ma to present with perhaps a more significant decrease in the last 6 Myr. This post-6 Ma decrease in convergence is consistent with the significant decrease in erosion rate at 6 Ma in eastern Bhutan proposed by Coutand et al. (2014).

6 Conclusions

10 This study presents a successful approach for using thermochronometer data to test the viability of a proposed cross section geometry based on forward models of the kinematic, exhumational, and thermal history of an area. The cross section geometry imparts a model of the horizontal and vertical component of displacement. We found that the location and magnitude of vertical displacement has the most significant control on the resulting cooling signal recorded by a suite of themochronometers. Mismatches between modeled and published thermochronometer ages provide insight into how cross

15 sections can be modified and reevaluated in order to create a more accurate solution to known geologic and thermochronologic constraints. We found that the addition of a ramp under the Greater Himalaya in our flexural model resulted in more accurately modeled cooling ages across this region while also preserving the modeled accuracy of other geologic and geophysical parameters.

Timing and rates of deformation in compressional settings can be quantified by coupling a high-resolution flexural model of

20 a balanced cross section with the thermal-kinematic model Pecube. Adjusting the timing of motion along structures also changes the timing of corresponding exhumation and affects mineral cooling ages. We applied a variable rate of deformation to obtain a best fitting model of the Trashigang cross section in Bhutan with velocities ranging from 68.4 to 6.7 mm/yr.

Topographic evolution also affects the exhumational history and thus can influence cooling ages, though generally this factor is less influential than other factors such as geometry, deformation velocity, and prescribed kinematics. We found that the

25 timing and magnitude of erosion controls which component of deformation and associated exhumation is recorded by a given thermochronometer system. Structural signals such as duplex formation and ramp propagation that are recorded in the cooling ages of each thermochronometer system may be altered by modifying the topographic evolution, most noticeable with lowest temperature thermochronometers. Thus small topographic changes can produce markedly different results in cooling age patterns for the same kinematic model, particularly when particles are at a depth and temperature close to the

30 closure temperature of a given mineral system. While changes in topographic gradients over multi-million year time scales are often uncertain, we can use thermal-kinematic modeling coupled with flexural models that estimate topographic evolution to better understand what is driving large- and small-scale changes in the pattern of exhumation over time and space.

This work promotes a new perspective on the spatial nature of thermochronometers, particularly the importance of considering the aerial distribution of cooling ages in the direction of transport and their relationship to the structural evolution of a landscape. Due to the predominant lateral transport of material that occurs in fold and thrust belts, the across-strike pattern of cooling ages from thermochronometers spanning a wide range of temperature and spatial coverage provide

5   the most robust constraints to the structural geometry and rate of deformation. Forward modeling cross sections and cooling ages using high-resolution spatial and temporal scales reveals which structures are responsible for a given cooling pattern, their geometry, and the rate at which they move, insights that are unavailable with other modeling workflows.

Acknowledgements

McQuarrie acknowledges research support from an Alexander von Humboldt Foundation fellowship, NSF (EAR 0738522)

10   and the University of Pittsburgh. Ehlers acknowledges supported by European Research Council (ERC). Midland Valley provided software support and use of the program Move. We thank Willi Kappler (University of Tübingen) for his time and programming assistance, and reviewers for their edits and suggestions. The data for this paper are available through contacting N. McQuarrie, while the modified version of Pecube that can be coupled to 2-D Move restoration files is available through T.A. Ehlers.

15   **References**

Adlakha, V.A., Lang, K.A., Patel, R.C., Lal, N., and Huntington, K.W.: Rapid long-term erosion in the rain shadow of the Shillong Plateau, Eastern Himalaya, Tectonophysics. 582, 76–83, 2013.

Banerjee, P. and Burgmann, R.: Convergence across the northwest Himalaya from GPS measurement, Geophys. Res. Lett., 29(13), 1652, 2002.

20   Banerjee, P., Burgmann, R., Nagarajan, B., and Apel, E.: Intraplate deformation of the Indian subcontinent, Geophys. Res. Lett., 35, L18301, 2008.

Batt, G. E. and Brandon, M.T.: Lateral thinking: 2-D interpretation of thermochronology in convergent orogenic settings, Tectonophys., 349, 185 – 201, 2002.

Berthet, T., Hetényi, G., Cattin, R., Sapkota, S.N., Champollion, C., Kandel, T., Doerflinger, E., Drukpa, D., Lechmann, S.,
25   and Bonnin, M.: Lateral uniformity of India Plate strength over central and eastern Nepal, Geophys J. Int, 195(3), 1481-1493, 2013.

Bettinelli, P., Avouac, J. P., Flouzat, M., Jouanne, F., Bollinger, L., Willis, P., and Chitrakar, G. R.: Plate motion of India and interseismic strain in the Nepal Himalaya from GPS and DORIS measurements, J. Geod., 80, 567–589, 2006.

Bhargava, O. N.: The Bhutan Himalaya: A Geological Account, Spec. Publ. Ser. Geol. Surv. India, 39, 245 pp, 1995.

30   Bilham, R., Larson, K., Freymueller, J., and Project Idylhim Members: GPS measurements of present-day convergence across the Nepal Himalaya, Nature, 386, 61–64, 1997.

Braun, J.: Pecube: a new finite-element code to solve the 3D heat transport equation including the effects of a time-varying, finite amplitude surface topography, Computers & Geosci., 29, 787-794, 2003.

Célérier, J., Harrison, T.M., Beyssac, O., Herman, F., Dunlap, W.J., and Webb, A.G.: The Kumaun and Garwhal Lesser Himalaya, India. Part 2: Thermal and deformation histories, Geol Soc. Am. Bull., 121, 1281-1297, 2009.

5   Coutand, I., Whipp Jr., D.M., Grujic, D., Bernet, M., Fellin, M.G., Bookhagen, B., Landry, K.R., Ghalley, S.K., and Duncan, C.: Geometry and kinematics of the Main Himalayan Thrust and Neogene crustal exhumation in the Bhutanese Himalaya derived from inversion of multithermochronologic data, J. Geophys. Res, Solid Earth, 119, 2014.

Davidson, C., Grujic, D., Hollister, L. S., and Schmid, S. M.: Metamorphic reactions related to decompression and synkinematic intrusion of leucogranite, High Himalayan crystallines, Bhutan, J. Metamorph. Geol., 15, 593–612, 1997.

10   DeCelles, P.G., Robinson, D.M., Quade, J., Ojha, T.P., Garzione, C.N., Copeland, P., and Upreti, B.N.: Stratigraphy, structure, and tectonic evolution of the Himalayan fold–thrust belt in western Nepal, Tectonics, 20, 487–509, 2001.

DeCelles, P. G., Robinson, D. M., and Zandt, G.: Implications of shortening in the Himalayan fold-thrust belt for uplift of the Tibetan Plateau, Tectonics, 21(6), 1062, 2002.

Duncan, C., Masket, J., and Fielding, E.: How steep are the Himalaya? Characteristics and implications along-strike
15   topographic variations, Geology, 31, 75–78, 2003.

Ehlers, T.A., et al: Computational tools for low-temperature thermochronometer interpretation, Rev in Min. and Geochem. 58, 589-622. DOI: 10.2138/rmg.2005.58.22, 2005.

Ehlers, T.A. and Farley, K.A.: Apatite (U–Th)/He thermochronometry; methods and applications to problems in tectonic and
20   surface processes, Earth Planet. Sci. Lett., 206, 1–14, 2003.

England, P., Le Fort, P., Molnar, P., and Pêcher, A: .Heat sources for Tertiary metamorphism and anatexis in the Annapurna-Manaslu region (central Nepal), J. Geophys. Res., 97, 2107–2128, 1992.

Gansser, A.: Geology of the Himalayas, 289 pp., Wiley-Interscience, New York, 1964.

Gansser, A.: Geology of the Bhutan Himalaya, 181 pp., Birkäuser, Basel, 1983.

25   Grujic, D., Hollister, L. S., and Parrish, R. R.: Himalayan metamorphic sequence as an orogenic channel: Insight from Bhutan, Earth Planet. Sci. Lett., 198, 177–191, 2002.

Grujic, D., Coutand, I., Bookhagen, B., Bonnet, S., Blythe, A., and Duncan, C.: Climatic forcing of erosion, landscape, and tectonics in the Bhutan Himalayas, Geology, 34, 801–804, 2006.

Hammer, P., Berthet, T., Hetényi, G., Cattin, R., DRupka, D., Chophel, J., Lechmann, S., Le Moigne, N., Champollion, C.,
30   and Doerflinger, E.: Flexure of the India plate underneath the Bhutan Himalaya, Geophys. Res. Lett., 40(16), 4225–4230, 2013.

Heim, A. and Gansser, A.: Central Himalayas—Geological observations of Swiss expedition, 1936, Mem. Soc. Helv. Sci. Nat., 73, 1–245, 1939.

Herman, F., Copeland, P., Avouac, J.P., Bollinger, L.. Mahéo, G.. Le Fort, P., Rai, S., Foster, D., Pêcher, A., Stüwe, K., and Henry, P.: Exhumation, crustal deformation, and thermal structure of the Nepal Himalaya derived from the inversion of thermo-chronological and thermobarometric data and modeling of the topography, J. Geophys. Res., 115, B06407, 2010.

Hodges, K.V., Parrish, R.R. and Searle, M.P.: Tectonic evolution of the central Annapurna Range, Nepalese Himalayas: Tectonics, v. 15, p. 1264–1291, 1996.

Hodges, K. V.: Tectonics of the Himalaya and southern Tibet from two perspectives, Geol. Soc. Am. Bull., 112, 324–350, 2000.

Hollister, L. S. and Grujic, D.: Pulsed channel flow in Bhutan, Geol. Soc. Spec. Publ., 268, 415–423, 2006.

Huntington, K.W., Ehlers, T.A., Hodges, K.V. and Whipp Jr., D.M.: Topography, exhumation pathway, age uncertainties, and the interpretation of thermochronometer ages, Tectonics, 26, TC4012, 2007.

Jain, A.K., Kumar, D., Singh, S., Kumar, A., and Lal, N.: Timing, quantification and tectonic modelling of Pliocene–Quaternary movements in the NW Himalaya: Evidence from fission track dating. Earth Planet. Sci. Lett., 179, 437–451, 2000.

Larson, K. M., Burgmann, R., Bilham, R. and Freymueller, J.: Kinematics of the India-Eurasia collision zone from GPS measurements, J. Geophys. Res., 104, 1077–1093, 1999.

Lavé, J. and Avouac, J. P.: Active folding of fluvial terraces across the Siwaliks Hills, Himalayas of central Nepal, J. Geophys. Res., 105, 5735–5770, 2000.

LeFort, P.: Himalayas: The collided range, present knowledge of the continental arc, Am. J. Sci., 275-A, 1–44, 1975.

Long, S., N. McQuarrie, N., Tobgay, T., Rose, C., Gehrels, G., and Grujic, D.: Tectonostratigraphy of the Lesser Himalaya of Bhutan: Implica- tions for the along-strike stratigraphic continuity of the northern Indian margin, Geol. Soc. Am. Bull., 123, 1406–1426, 2011a.

Long, S., McQuarrie, N., Tobgay, T., and Grujic, D.: Geometry and crustal shortening of the Himalayan fold-thrust belt, eastern and central Bhutan, Geol. Soc. Am. Bull., 123, 1427–1447, 2011b.

Long, S.P., McQuarrie, N., Tobgay, T., Coutand, I., Cooper, F., Reiners, P., Wartho, J., and Hodges, K.V.: Variable shortening rates in the eastern Himalayan thrust belt, Bhutan: insights from multiple thermochronologic and geochrono-logic datasets tied to kinematic reconstructions, Tectonics, 31, TC5004, 2012.

Mattauer, M.: Intracontinental subduction, crust-mantle décollemont and crustal-stacking wedge in the Himalayas and other collision belts, in Collision Tectonics, edited by M. P. Coward and A. C. Ries, Geol. Soc. Spec. Publ., 19, 37–50, 1986.

McQuarrie, N., Robinson, D., Long, S., Tobgay, T., Grujic, D., Gehrels, G., and Ducea, M.: Preliminary stratigraphic and structural architecture of Bhutan: Implications for the along-strike architecture of the Himalayan system, Earth Planet. Sci. Lett., 272, 105–117, 2008.

McQuarrie, N., Tobgay, T., Long, S.P., Reiners, P.W., and Cosca, M.A. Variable exhumation rates and variable displacement rates: documenting a recent slowing of Himalayan shortening in western Bhutan, Earth Planet. Sci. Lett., 386, 161-174, 2014.

McQuarrie, N. and Ehlers, T.A: Influence of thrust belt geometry and shortening rate on thermochronometer cooling ages: Insights from the Bhutan Himalaya: Tectonics, 34, doi:10.1002/ 2014TC003783, 2015.

5   McQuarrie, N. and Ehlers T.A.: Techniques for understanding fold-thrust belt kinematics and thermal evolution: in Law, R.D., Thigpen, J.R., Merschat, A.J., Stowell, H.H., (eds) Linkages and Feedbacks in Orogenic Systems. Geol. Soc. Am. Mem. 213. doi:10.1130/2017.1213(02), 2017.

Menon, R., Kumar, P., Reddy, G. and Srinivasan, R.: Radiogenic heat production of late Archaean Bundelkhand granite and some Protero- zoic gneisses and granitoids of central India, Curr. Sci., 85, 634–638, 2003.

Mitra, S., Priestley, K., Bhattacharyya, A. and Gaur, V.K.: Crustal structure and earthquake focal depths beneath
10   northeastern India and southern Tibet, Geophys. J. Int., 160, 227–248, 2005.

Patriat, P. and Achache, J.: Indian-Eurasian collision chronology has implications for crustal shortening and driving mechanisms of plates, Nature, 311, 615– 621, 1984.

Powell, C.M., and Conaghan, P.J.: Plate tectonics and the Himalayas, Earth Planet. Sci. Lett. 20, 1–12, 1973.

Ray, L., Bhattacharya and A. Roy, S.: Thermal conductivity of Higher Himalayan Crystallines from Garhwal Himalaya,
15   India. Tectonophys. 434, 71–79, 2007.

Robert, X., van der Beek, P., Braun, J., Perry, C., Dubille, M. and Mugnier, J.-L.: Assessing Quaternary reactivation of the Main Central Thrust zone (central Nepal Himalaya): New thermochronologic data and numerical modeling, Geology, 37(8), 731–734, 2009.

Stüwe, K. and Foster, D.: 40Ar/39Ar, pressure, temperature and fission track constraints on the age and nature of
20   metamorphism around the Main Central Thrust in the eastern Bhutan Himalaya, J. Asian Earth Sci., 19(1–2), 85–95, 2001.

Thiede, R., Ehlers, T.A., Bookhagen, B., and Strecker, M.R.: Erosional variability along the northwest Himalayan Front, J. Geophys. Res., Earth Surface, 114, F01015, 2009.

Thiede, R. C. and Ehlers, T. A.: Large spatial and temporal variations in Himalayan denudation, Earth Planet. Sci. Lett., 371, 278-293, 2013.

25   Whipp, D.M. Jr., Ehlers, T.A., Blythe, A.E., Huntington, K.W., Hodges, K.V., and Burbank, D.W.: Plio-quaternary exhumation history of the central Nepalese Himalaya: 2. Thermokinematic and thermochronometer age prediction model, Tectonics, 26, TC3003, 2007.

Whipp, D., Ehlers, T.E., Braun, and J., Spath, C.: Effects of exhumation kinematics and topographic evolution on detrital thermochronometer data, J. Geophys. Res. 114, F04021, 2009.

30   Yin, A.: Cenozoic tectonic evolution of the Himalayan orogen as constrained by along-strike variation of structural geometry, exhumation history, and foreland sedimentation, Earth Sci. Rev., 76, 1–131, 2006.

Yin, A. and Harrison, T.M.: Geologic evolution of the Himalayan-Tibetan orogeny, Ann. Rev. Earth Planet. Sci., 28, 211– 280, 2000.

Zhang, P. Z., et al.: Continuous deformation of the Tibetan Plateau from global positioning system data, Geology, 32, 809–
35   812, 2004.
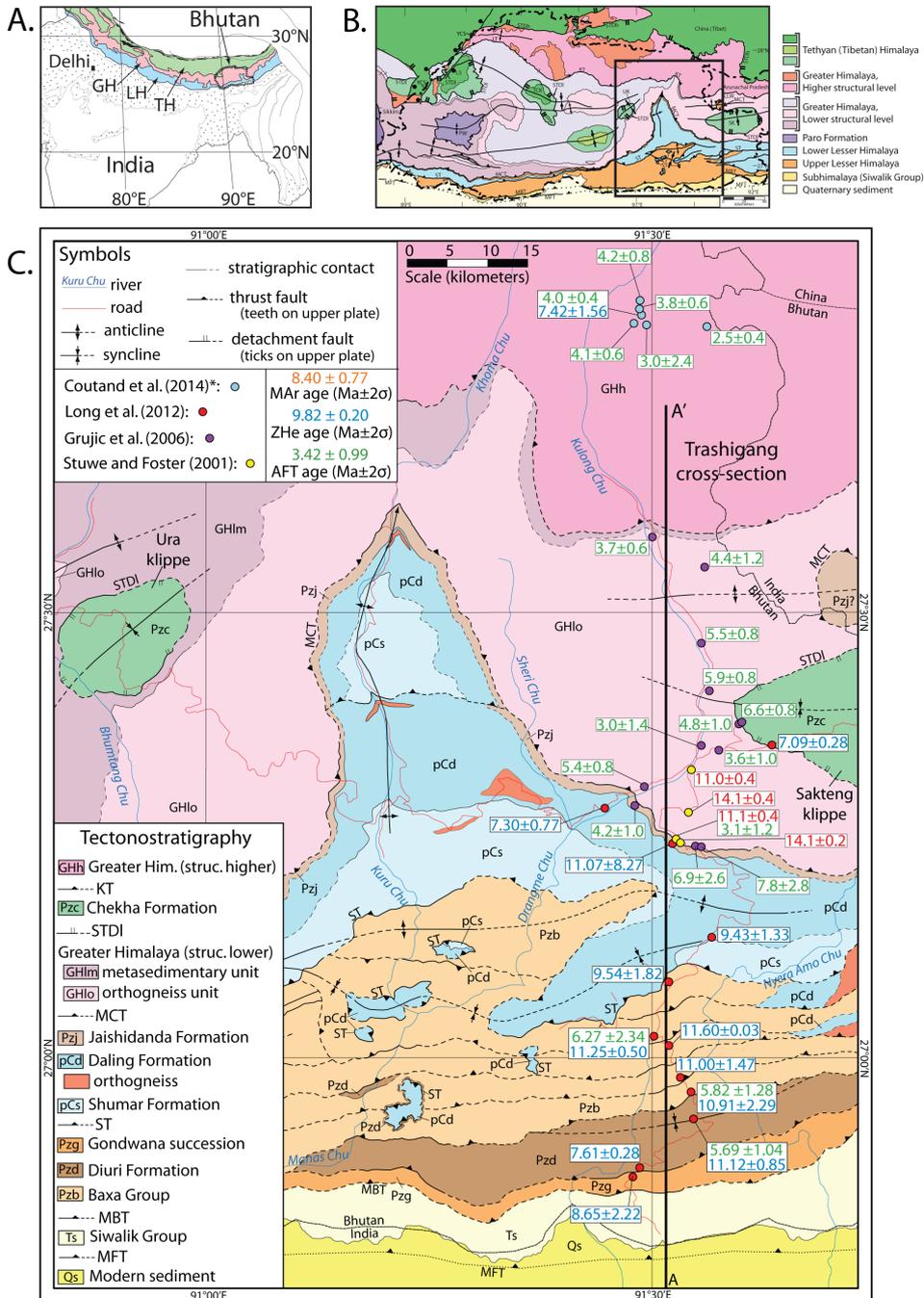
Figure 1: A. Generalized geologic map of central and eastern Himalayan orogen, modified from Gansser (1983). Abbreviations are GH: Greater Himalaya, LH: Lesser Himalaya, TH: Tethyan Himalaya

B. Simplified tectonostratigraphic map of Bhutan, modified from Long et al. (2012). Abbreviations of structures from north to south: YCS: Yadong Cross-Structure, STDh: structurally higher South Tibetan Detachment, LT: Laya thrust, KT: Kakhtang

5 **Thrust, STDl: structurally lower South Tibetan Detachment, MCT: Main Central Thrust, ST: Shumar Thrust MBT: Main Boundary Thrust, MFT: Main Frontal Thrust. Abbreviations of windows and klippen from east to west: LLW: Lum La window, SK: Sakteng klippe; UK: Ura klippe, TCK: Tang Chu klippe, PW: Paro window LS: Lingshi syncline. The border of Bhutan is marked as a dashed and bolded line, and the area of Fig 1C outlined as a solid black rectangle.**

C. Geologic map of eastern Bhutan with Trashigang section line A-A' (Fig 2) shown, modified from Long et al. (2012) with
10 **reported thermochronologic data. Blue circles mark AFT and ZHe data from Coutand et al. (2014) that has been projected 27 km SE at 129 degrees, preserving distance of data between KT and STDh structures. Cooling ages are reported in Myr. Abbreviations of units in inset are GHh: Greater Himalaya, structurally higher; Pza: Chekha Formation; GHlm: Greater Himalaya, structurally lower, metasedimentary unit; GHlo: Greater Himalaya, structurally lower, orthogneiss unit; Pzj: Jaishidanda Formation; pCd Daling Formation; pCs: Shumar Formation; Pzg: Gondwana succession; Pzd: Diuri Formation; Pzb: Baxa Group Ts: Siwalik**
15 **Group; Qs: Modern sediment.**
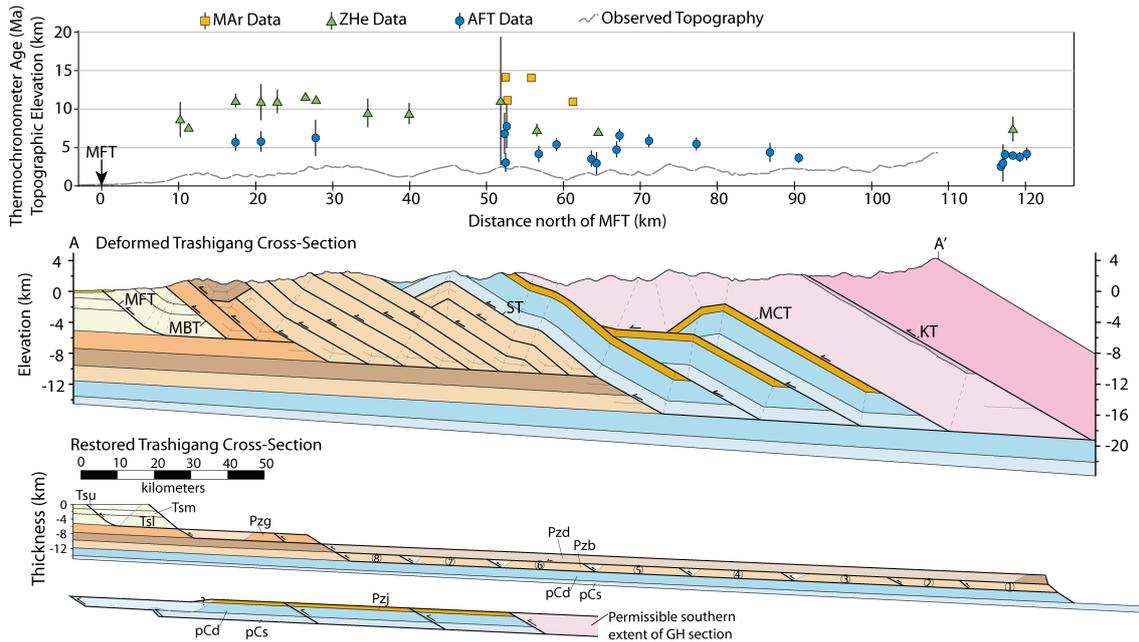
Figure 2: A. MAr, ZHe, and AFT thermochronometer data and elevation along the Trashigang section plotted in the direction of transport. B. Simplified balanced geologic cross section of the Trashigang region of Bhutan, modified from Long et al. (2011a). Scale of the deformed section is represented on the above graph. Unit abbreviations are shown in the stratigraphic column of
20 **figure 1. Abbreviations same as figure 1.**
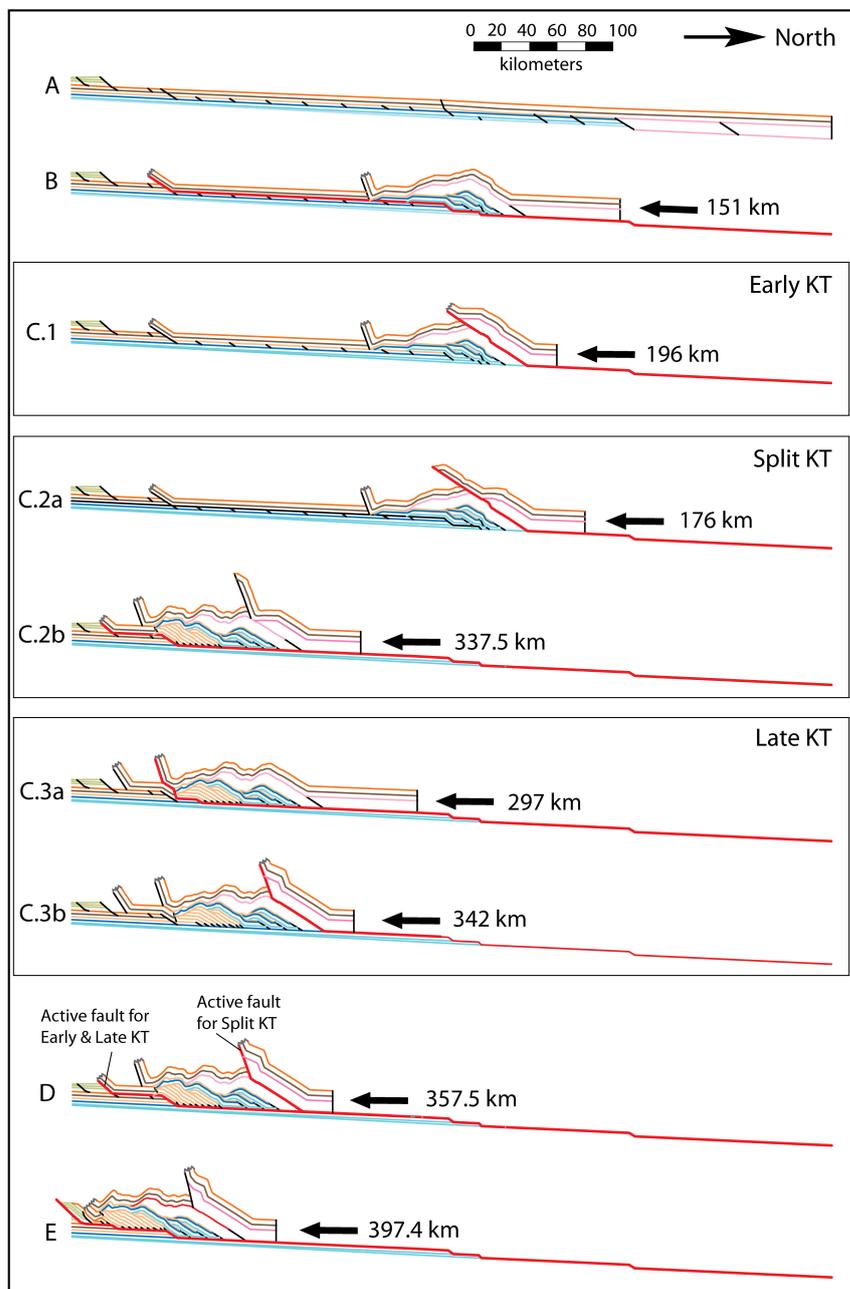
Figure 3: Sequential kinematic reconstruction of the Trashigang cross section depicting three kinematic scenarios of out-of-sequence thrusting tested in this study. Net slip amounts are shown for each subfigure A: the restored section used in the kinematic model; B: deformation along the MCT and ST, including duplexing of the lower LH; C.1: KT motion prior to upper LH duplexing (Early KT); C.2: KT motion before and after upper LH duplexing (Split KT); C.3: KT motion after seven out of eight

5   **horses of upper LH duplex have been deformed (Late KT); D: completion of out-of-sequence thrusting and Upper LH duplexing. Note that the most recent active fault in this step for Split KT varies; E: deformation along MBT and MFT.**
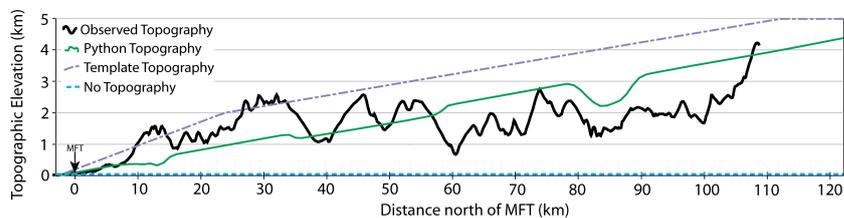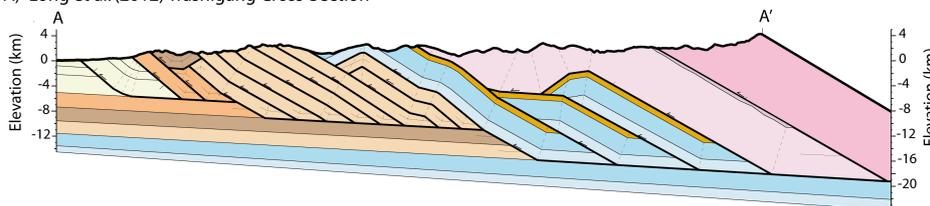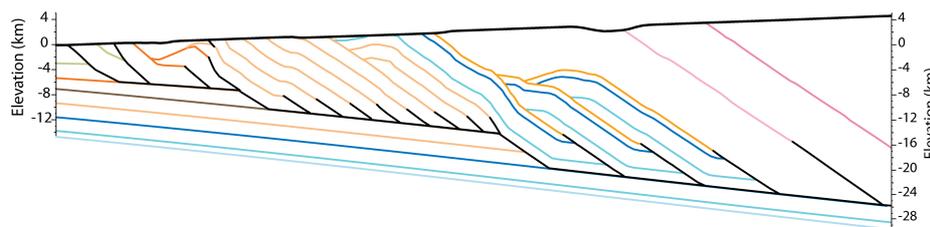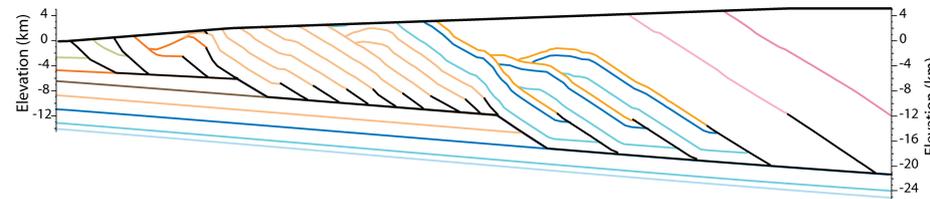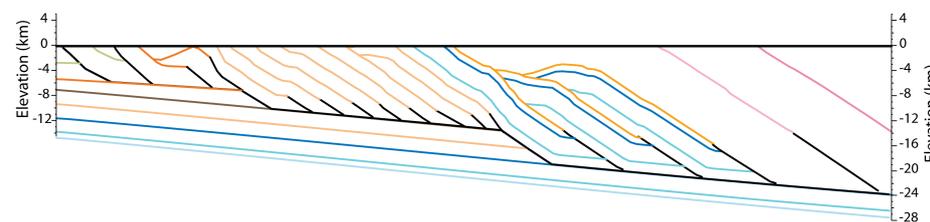
Figure 4: A) Vertically exaggerated topographic model elevation compared to observed topography of the Trashigang section Flexural model topographic elevations of the deformed cross sections using Split KT and Python, Template, and No Topography. B) Trashigang cross section simplified from Long et al. (2011a) compared to flexural model outputs using Split KT and Python, Template, and No Topography.
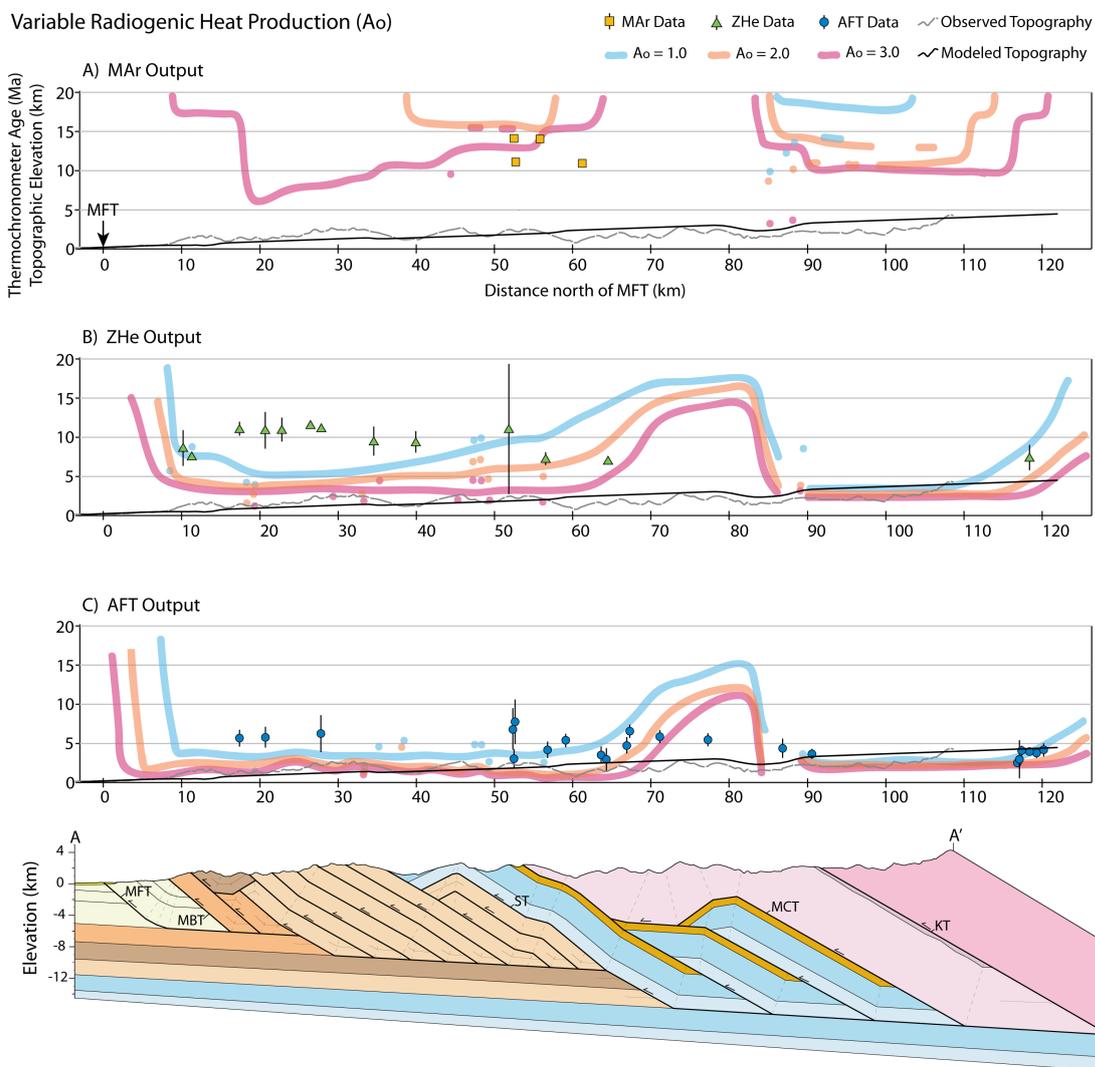
Figure 5: Predicted MAr (a), ZHe (b), and AFT (c) cooling ages from Pecube using variable heat production (Ao) values of 1.0, 2.0, and 3.0 μW/m³ compared to published cooling data. Other model variables are set as constant velocity, Split KT, Python Topography.
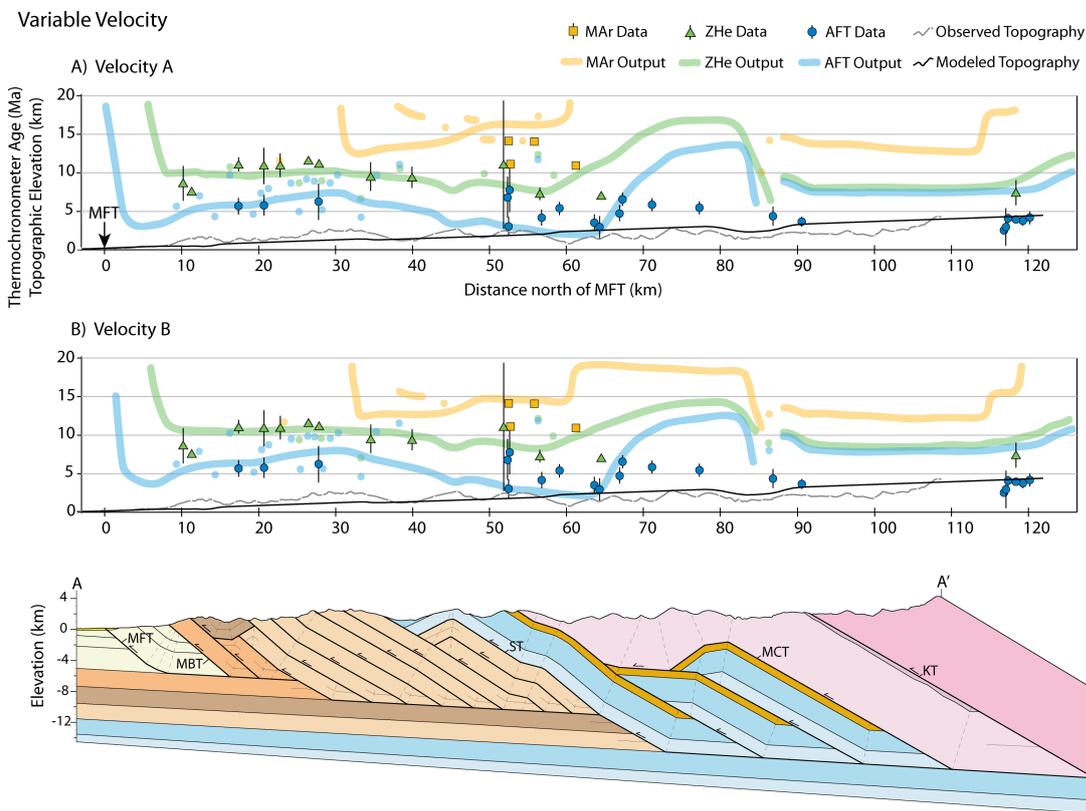
Figure 6: Predicted MAr (yellow), ZHe (green), and AFT (blue) cooling ages using variable velocities A (A) and B (B) compared to published thermochronometer data. Other model variables are set as Split KT, Python Topography, and Ao = 2.5 μW/m³.
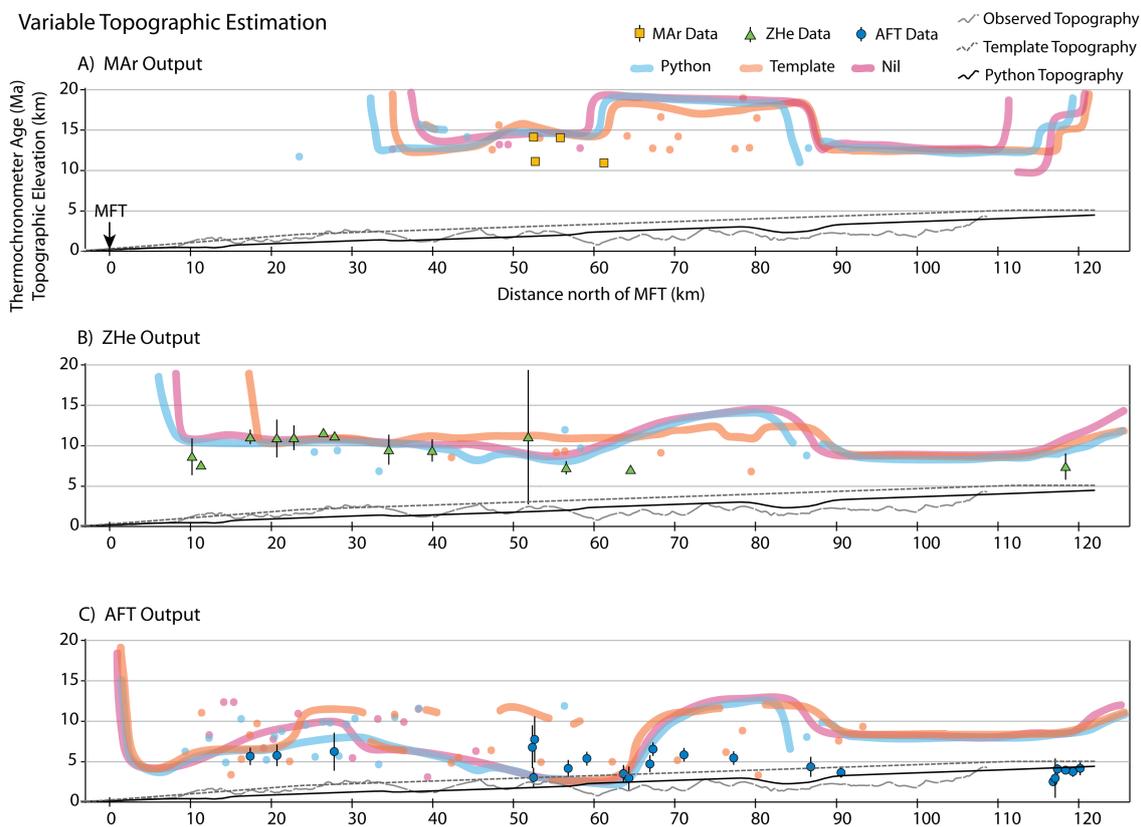
Figure 7: Predicted MAr (a), ZHe (b), and AFT (c) cooling ages using Python, Template, and No Topography models compared to published thermochronometer data. Other model variables are set as Split KT, Velocity B, and Ao = 2.5 μW/m³.
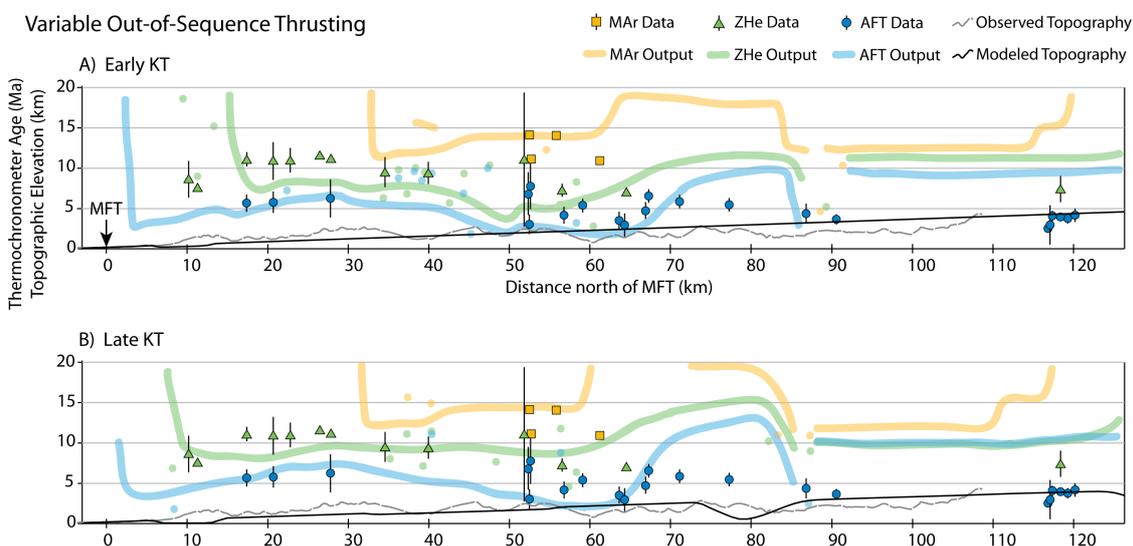
Figure 8: Predicted MAr, ZHe, and AFT cooling ages using flexural models using Early KT (a) and Late KT (b) kinematic scenarios compared to published thermochronometer data. Other model variables are set as Python Topography, Velocity B, and Ao = 2.5 μW/m³.
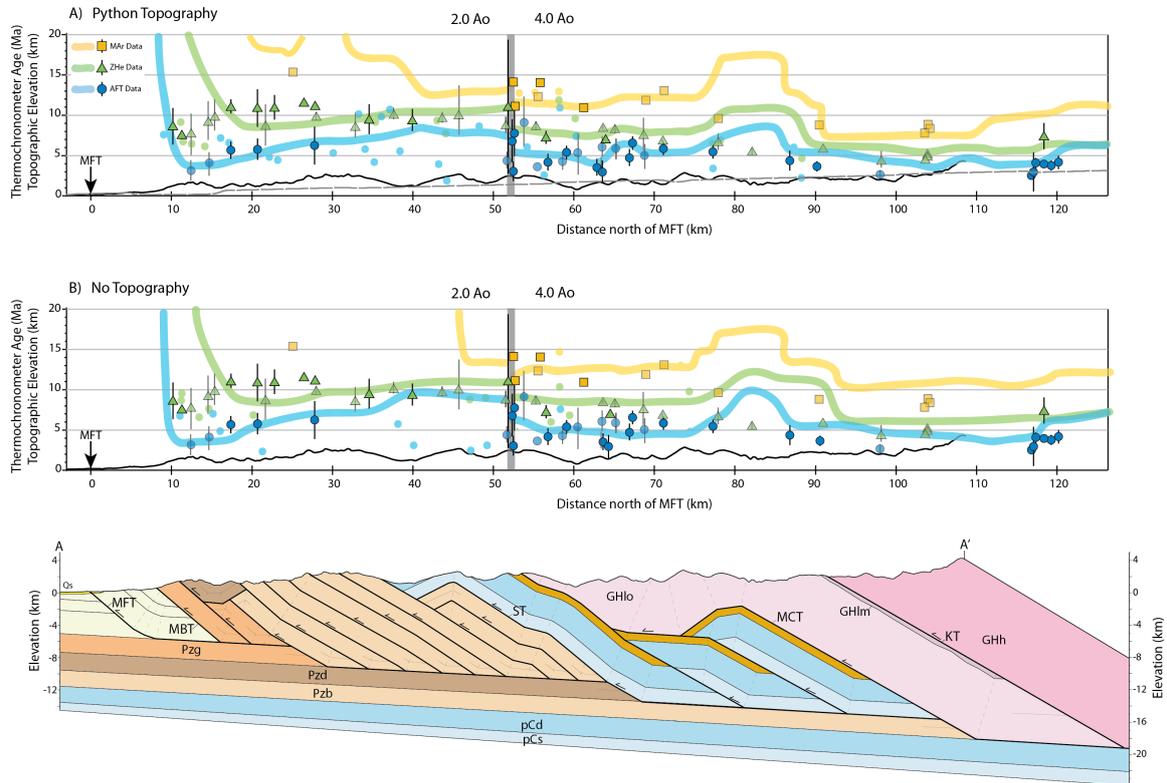
Figure 9: Predicted MAr, ZHe, and AFT cooling ages using a flexural model of the modified geometry with Python Topography (a) and No Topography (b) models compared to published thermochronometer data. The decollement ramp through the upper LH Baxa and Diuri units has been split, and the Baxa footwall ramp moved 35 km north. Published data include additional ages from the Kuru Chu line of section west of the Trashigang section (Long et al., 2012) shown in the transparent colors. The gray vertical line aligned with the location of the MCT shows the division between outputs from separate thermal models have been merged using 2.0 and 4.0 µW/m³ to the south and north of the MCT respectively. Other model variables are set as Split KT and Velocity C.
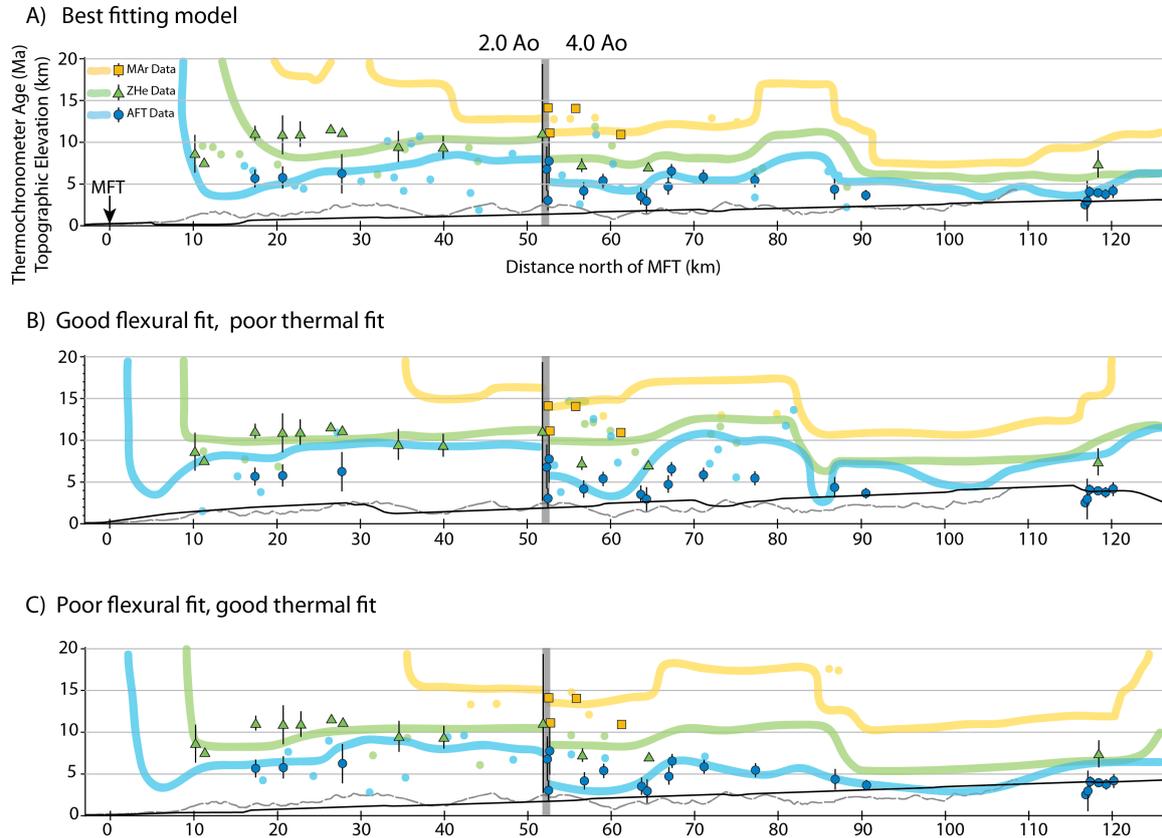
Figure 10: Comparison of predicted MAr, ZHe, and AFT cooling ages between the best fitting flexural-thermal model combination present in figure 9a (a), a well-matched flexural model that yielded a thermal model with poorly fitting predicted ages (b), and a poorly-matched flexural model that yielded a thermal model with well-fitting predicted ages (c). Published data include additional ages from the Kuru Chu line of section west of the Trashigang section (Long et al., 2012). Predicted ages are presented with combined thermal models using Ao of 2.0 and 4.0 µW/m³. Flexural models used Split KT and Python Topography.

Table 1

Kinematics & Geometry	Topography	EET (km)	Crustal Density (g/cm³)	Foreland Basin Thickness (km)	Decollement Dip (degrees)	Surface Geology
Long et al. [2011a]	-	-	-	5.6	4	-
Early KT	Python	65	2.60	4.3	5	over-eroded by 0.5 km at Diuri Fm and hanging wall of ST
Early KT	Template	65	2.60	4.5	4.4	GOOD
Split KT	Python	65	2.60	5.2	5.1	GOOD
Split KT	Template	65	2.60	4.6	3.75	over-eroded by 0.3 km at hanging wall of ST
Split KT	NoTopo	65	3.20	5.1	4.6	over-eroded by 0.9 km at Diuri Fm and by 2.3 km at hanging wall of ST
Late KT	Python	65	2.60	5.6	5.4	under-eroded by 0.4 km at Diuri Fm, hanging wall of ST, and GH synform
Late KT	Template	65	2.60	4.9	5.2	under-eroded by 1 km at Diuri Fm
Modified Split KT	Python	65-85	2.65	5.7	4.5	GOOD
Modified Split KT	NoTopo	60	3.25	4.9	4.4	over-eroded by 2.0 km at Diuri Fm and hanging wall of ST

Table 1: Comparison of the geologic constraints of the published Trashigang cross section to the final deformed cross section results of the flexural models presented in this study.

Table 2.   Numerical model parameters

Property/ parameter	Model input value
Material properties	
heat production	
crustal volumetric heat production	1.0 - 4.0 mW/m3
e-folding depth of crustal heat prod.	20 km
thermal conductivity	2.5 W/m K
specific heat capacity	800 J/kg K
crustal density	2700 kg/m3
mantle density	3300 kg/m3
Numerical properties	
temperature at base	1300°C
model base	110 km
surface temperature at 0 km	20 °
atmospheric lapse rate	0°/ km
kinematic grid spacing	0.5 km
displacement increment	~10 km
model domain	730 x 110 x 5 km
horizontal node spacing (numerical model)	0.5 km
vertical node spacing (numerical model)	1.0 km
model start time	50 Ma

Table 2: Thermal and rock property parameters assigned as input for Pecube.

Table 3.
PUBLISHED GEOMETRY FLEXURAL MODEL RATES
SPLIT KT MODELS

| | | | VELOCITY: CONSTANT | | VELOCITY A | | VELOCITY B | |
| | | | Ao (µW/m³): 1.0-3.0 | | 2.25-3.0 | | 2.25-3.0 | |
Active Structure	Slip (km)	Net Slip (km)	Time (Ma)	Velocity (mm/yr)	Time (Ma)	Velocity (mm/yr)	Time (Ma)	Velocity (mm/yr)
MCT	63.2	63.2	23.0	17.3	23.0	31.6	20.0	21.1
Lower LH Duplex	87.9	151.1	19.3	17.3	21.0	14.7	17.0	25.1
KT	25.0	176.1	14.3	17.3	15.0	37.3	13.5	74.6
Upper LH Duplex	161.4	337.5	12.8	17.3	14.3	37.3	13.2	74.6
KT	20.0	357.5	3.5	17.3	10.0	6.0	11.0	5.4
MBT	26.5	384.0	2.3	17.3	6.7	6.0	7.3	5.4
MFT	13.4	397.4	0.8	17.3	2.2	6.0	2.5	5.4
			0.0		0.0		0.0	

EARLY KT MODELS

| | | | VELOCITY: VELOCITY B | |
| | | | Ao (µW/m³): 2.25-2.75 | |
Active Structure	Slip (km)	Net Slip (km)	Time (Ma)	Velocity (mm/yr)
MCT	63.2	63.2	20.0	21.1
Lower LH Duplex	87.9	151.1	17.0	22.0
KT	45.0	196.1	13.0	69.4
Upper LH Duplex	145.9	342.0	12.4	69.4
Upper LH Duplex	15.5	357.5	10.3	5.4
MBT	26.5	384.0	7.4	5.4
MFT	13.4	397.4	2.5	5.4
			0.0	

LATE KT MODELS

| | | | VELOCITY: VELOCITY B | |
| | | | Ao (µW/m³): 2.25-3.5 | |
Active Structure	Slip (km)	Net Slip (km)	Time (Ma)	Velocity (mm/yr)
MCT	63.2	63.2	20.0	21.1
Lower LH Duplex	87.9	151.1	17.0	22.0
Upper LH Duplex	145.9	297.0	13.0	73.0
KT	45.0	342.0	11.0	45.0
Upper LH Duplex	15.5	357.5	10.0	5.5
MBT	26.5	384.0	7.2	5.5
MFT	13.4	397.4	2.4	5.5
			0.0	

NEW GEOMETRY FLEXURAL MODEL RATES

| | | | VELOCITY: Velocity C | |
| | | | Ao (µW/m³): 1.75-4.0 | |
Active Structure	Slip (km)	Net Slip (km)	Time (Ma)	Velocity (mm/yr)
MCT	58.0	58.0	20.0	29.0
Lower LH Duplex	78.0	136.0	18.0	17.3
KT	25.0	161.0	13.5	68.4
Upper LH Duplex	146.0	161.0	13.1	68.4
Upper LH Duplex	53.0	360.0	11.0	14.6
KT	20.0	380.0	7.4	14.6
MBT	29.0	409.0	6.0	6.7
MFT	11.0	420.0	1.7	6.7
			0.0	

Table 3: Tested combinations of flexural, thermal, and deformation timing and velocity parameters.