Supplement 1

1. Numerical method

In this study, we use the Arbitrary Lagrangian-Eulerian finite element code FANTOM, which solves the Stokes equation coupled to the heat-transport equation (Erdős, 2014; Thieulot, 2011) to analyze the behavior of a layered lithosphere in extensional and contractional regimes. The presented model setup builds on our previous work investigating the effects of structural inheritance (Erdős et al., 2014) and surface processes (Erdős et al., 2015) on mountain building.

While stress is below the yield criterion, deformation occurs by viscous flow and is characterized by a temperature-dependent nonlinear power-law rheology. The effective viscosity η_{eff} is specified as:

$$\eta_{eff} = f.A^{-1/n}.\dot{\varepsilon}^{(1-n)/2n}.exp\left(\frac{Q + Vp}{nRT}\right)$$

Where A is a pre-exponential scaling factor, n is the power-law exponent, $\dot{\varepsilon}$ is the second invariant of the deviatoric strain-rate tensor, Q is the activation energy, V is the activation volume, p is pressure, T is temperature and R is the universal gas constant. Different parts of the lithosphere can be described by different dominant minerals, requiring the use of different rheologies. Parameters, such as the pre-exponential scaling factor, the power-law exponent, the activation energy or the activation volume are derived from laboratory measurements (Gleason and Tullis, 1995; Karato and Wu, 1993) and are given in Table 1 of the main text for each model material. The effective viscosity of the quartz-dominated upper and lower crust is characterized by large uncertainties due to compositional inhomogeneities. Therefore, we use a scaling factor f to generate crustal materials of variable strength. Here, we use f = 100 to simulate a strong lower crust coupled to the strong upper mantle lithosphere (Table 1).

We use the pressure-dependent Drucker-Prager yield criterion, equivalent to the Coulomb yield criterion in two dimensions to model frictional-plastic deformation. Yielding occurs when:

$$(J_2')^{1/2} = p \, sin\phi_{eff} + C \, cos\phi_{eff}$$

Where J_2' is the second invariant of the deviatoric stress, p is pressure, C is cohesion and φ_{eff} is the effective internal angle of friction. In the above equation, φ_{eff} is given as $p \, sin(\varphi_{eff}) = (p-p_f) \, sin(\varphi)$ for pore-fluid pressure p_f. Mechanisms such as fluid-pressure variations (Sibson, 1990) grain-size reduction and mineral transformations (Bos and Spiers, 2002) may reduce the frictional strength. The effect of these plastic strain-softening mechanisms is introduced into the model through a linear decrease of the internal angle of friction from 15° to 2° and a simultaneous decrease of cohesion from 20 MPa to 4 MPa as the second invariant of deviatoric strain varies from 0.05 to 1.05 (Figure 1b).

The use of temperature dependent densities and rheologies means that we need to solve for the heat-transport equation as well as for the Stokes equations when modeling viscous-plastic slow creeping

flows. The mechanical and thermal systems are solved sequentially for each model time step. The initial temperature field laterally uniform in the model domain. The surface temperature is $T_0 = 0$ °C and the $T_m = 550$ °C Moho-temperature is reached through a parabolic geothermal gradient. Between the Moho and the LAB ($T_b = 1330$ °C), the temperature increases linearly, while temperature of the sub-lithospheric mantle is initially uniform at $T = 1330$ °C. The geothermal gradient is affected by crustal heat production ($h_c = 0.8$ μW m^{-3}). The lateral thermal boundaries are insulated, while the basal boundary is set to a constant temperature $T_b = 1330$ °C.

The densities of the modeled materials depend on temperature: $\rho(T) = \rho_0(1 - \alpha(T - T_0))$, with the thermal expansion coefficients α and the reference densities at temperature $T_0 = 0$ °C given in Table 1. Elasticity is not included in our models, as it is generally not expected to affect large-strain deformations. Moreover, the inclusion of the sub-lithospheric mantle in the models allows for self-consistent isostatic compensation. As the topography is controlled by the density distribution and the inherent strength of the modeled materials, our viscous-plastic models in effect exhibit regional "flexural" isostasy.

2. Model Setup

The models are set up to represent an idealized continental lithosphere situated on top of the sub-lithospheric mantle in a 600-km high and 1400-km wide box (Figure 2 of main text). The lithosphere consists of the following rheologically distinct layers: (1) a 3 km thick pre-orogenic sedimentary package; (2) a 1 km thick weak layer representing a salt décollement horizon that can be taken to represent the widely present Triassic evaporite deposits in western Europe; (3) a 21 km thick quartz dominated upper crust; (4) a 10 km thick, strong lower crust; (5) an 85 km thick dry olivine dominated lithospheric mantle; (6) and a 480 km thick, wet olivine dominated sub-lithospheric mantle. The 1 km thick weak décollement in our model is thicker than that observed in the Alps (Philippe et al., 1996) but is limited by the resolution of the model, which is not sufficient to accurately track thinner layers. A Eulerian grid consisting 2800 nodes in the horizontal and 300 nodes in the vertical dimension is used to carry out the calculations. The nodes are distributed irregularly in the vertical direction, with 125 nodes covering the upper crust, 125 nodes covering the lower crust and the mantle lithosphere, and an additional 50 nodes covering the sub-lithospheric mantle. Consequently, the horizontal resolution is 500 m for the entire model-domain, while the vertical resolution is changing between 200 m in the upper crust, 800 m in the lower crust and the mantle lithosphere, and 9.5 km in the sub-lithospheric mantle.

Constant, depth-dependent velocity boundary conditions are imposed on both vertical model boundaries (v = ± 0.5 cm/yr on each side), while a free slip condition is implemented on the basal model

boundary. The velocities are prescribed so that the net amount of material passing through the sides of the box is zero. All three models were run in extension mode for 15 Myr before the boundary conditions were reversed, creating a contractional setting for the remaining 50 Myr model time. Note that we did not explore the effects of a post-rift thermal relaxation phase (approximately 70 Myr in case of the Alps for example).

To ensure that the deformation initially localizes in the center of the model rather than close to the boundaries, a strain-weakened rectangular seed of 6 by 6 km is positioned at the top of the strong, frictional-plastic lower crust (Figure 2 of main text, pink square). This seed allows for the development of single-sided subduction without the need for any additional forcing. Note that, since the initial conditions are symmetrical, subduction initiates with a random polarity in each model. For ease of comparison, however, we show all models with the same polarity.

3. Surface Processes

When the surface process algorithms are not activated (i.e. in the reference model) the top of the model acts as a free surface, with only minimal surface smoothing applied to prevent numerical instabilities.

In Model 2, a simple sedimentation algorithm is applied. Starting after 45 Myr and Δx = 150 km of crustal shortening, all topography below a reference level is filled with sediments at the end of every time step (5000 yr). The reference level is prescribed as the elevation of the top left node of the model. After the onset of sedimentation there is a 0.5 Myr transitional period during which, at the end of every timestep, only 10% of the prescribed accommodation space is filled up with sediments. The implementation of a transitional period serves to attenuate the strong effect of sedimentation on the mechanical evolution of the foreland fold-and-thrust belt and to prevent numerical instabilities.

In Model 3, the sedimentation scenario is slightly modified. Initial sedimentation starts earlier, at 35 Myr and Δx = 50 km, while the base level of sedimentation is changed from -500 m to 0 m at 50 Myr (and Δx = 200 km) to mimic the change in deposition rate and the transition from an underfilled to an overfilled basin observed in the Western Alps.

The sediments have the same material properties as the upper crust except for their lower density (ρ_{sed} = 2300 kg m^{-3}). The above described sedimentation algorithm is very simple, but the resulting basin-fill geometries are consistent with observations from natural foreland-basin systems (DeCelles and Giles, 1996). The delayed onset of sedimentation allows for the rise of an orogen with a more than 2-km high mean topography, which could provide a realistic source for the sediments. Moreover, if applied during the rifting phase, our sedimentation algorithm would quickly create a sedimentary basin with an unnaturally high thickness (i.e. over 30 km).

Models 2 and 3 also employ elevation-dependent erosion following $\Delta h/\Delta t = h.E_r$, where h is elevation, t is time and E_r is an erosional time constant (yr^{-1}). E_r is set so that a 2-km high topography erodes by 1 km in 1 Myr. The erosion algorithm – when used – is activated together with the sedimentation algorithm.

Note that as a result of this simplistic approach, the two surface-process algorithms do not conserve mass. However, material that was initially deposited in the model can be eroded when elevated above the reference base level.

Supplement 2

1. Model evolution of additional models 1.1 and 2.1

We present two additional model experiments that help in deciphering the effects of sedimentation and erosion on the evolution of the of surface slope (α) and décollement angle (β) of the orogenic pro-wedge over time. In Model 1.1 there is no sedimentation; only the erosional routine is active throughout the entire model run (figure A1 a-c). In Model 2.1 there is no erosion, and only the fixed-baselevel sedimentation routine is active from 45 My onward (figure A1 d-f).

The first two phases of the model evolution, described in the first paragraph of section 3.1 of the main text, are characteristic for the two complementary models as well, even though in case of Model 1.1 erosion is active from the model onset.

2. Model 1.1

In the third phase of model 1.1, wedge building starts predominantly in the pro-wedge, initially by formation of a crustal-scale pop-up structure, and then primarily through an outward-propagating sequence of basement thrust sheets (figure A1 c) with an average length of 46 km. Superimposed on this basement deformation, the pre-orogenic sediments are also deformed, creating a complex thin-skinned fold-and-thrust belt (figure A1 a and b).

The lack of sedimentation results in a narrow zone of foreland deformation, focused around the frontal basement thrust, while the presence of erosion results in a narrower (290 km width at 60 My, compared to 375 km in the reference model) and lower (maximum height of 6.4 km at 60 My, compared to 7.5 km in the reference model) orogen. The presence of erosion also means that large areas of the orogenic hinterland are stripped from the pre-orogenic sedimentary cover, exposing upper- and middle-crustal rocks, as well as a narrow sliver of continental mantle lithosphere in the retro-wedge that was originally part of the uplifted keystone structure. The retro-wedge of the model remains relatively undeformed throughout the model run, with the main retro-thrust in the foot of the

130   above-mentioned keystone structure accommodating almost all the displacement in the crust (figure A1b). The retro-basement shows little thin-skinned deformation, with only four minor back-thrusts activated throughout the model run.

3. Model 2.1

The general development and the major features of Model 2.1 are very similar to those of Model 2

135   (figure A1 d-f). Following initiation of sedimentation at 45 My, sediment-loaded foreland basins form on both sides of the orogen, with more intense thin-skinned deformation on the pro-side. The sequence of outward-propagating basement-thrust sheets in the pro-wedge is disrupted as deformation remains localized on the active frontal basement thrust for about 6 My before a new basement-thrust sheet is formed. The absence of erosion leads to the creation of a thick, highly

140   deformed cover of shortened pre-orogenic sediments on top of the orogenic hinterland, burying the progressively rotated central basement-thrust sheets. The orogen is wider (380 km width at 60 My, compared to 340 km in model 2) and higher (maximum height of 8.4 km at 60 My, compared to 6.4 km in model 2), with thicker foreland basin cover on both sides (figure A1 d and e). The lack of erosion means that there is no chance for buried and heated upper- and middle-crustal material to be

145   exhumed; hence the deep structure of the central part of the orogen is less hot than in model 2 (compare the location of the 550 °C isotherms in models 1, 2 and 2.1).

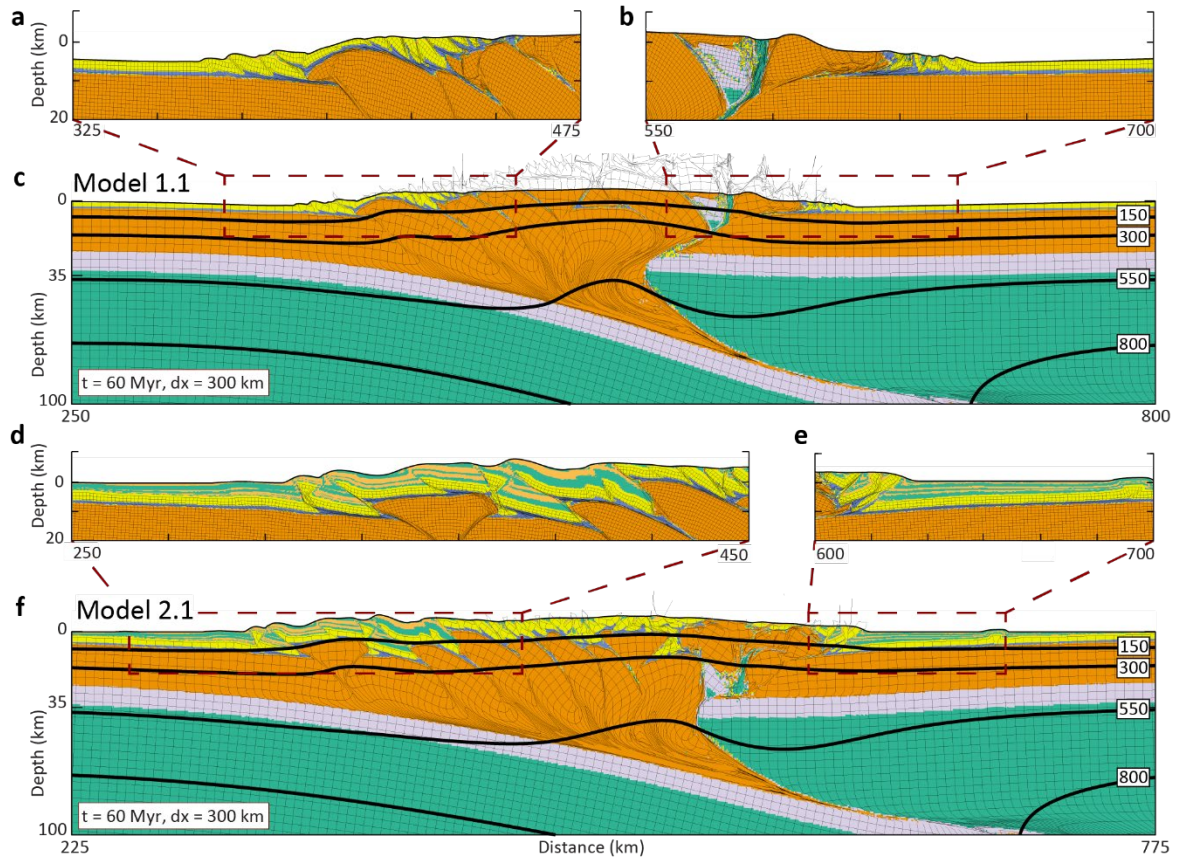



Figure A1. Supplementary Model results. The material coloring scheme is identical to that used in Figure 1. Models are run for 60 My: 15 My (150 km) extension followed by 45 My (450 km) contraction for a total net contraction of 300 km. (a)-(c): Model 1.1 with only erosion, showing deformed Lagrangian mesh and isotherms after (c) 60 My (Δx = 300 km). (a) and (b) are extracts from panel (c) showing the small-scale deformation patterns in the foreland fold-and-thrust belts. (d)-(f): Model 2.1 including a simple sedimentation algorithm filling up accommodation space until a baselevel of -500 m, showing deformed Lagrangian mesh and isotherms after 60 My (Δx = 300 km). (d) and (e) are extracts from panel (f) showing the small-scale deformation patterns in the foreland fold-and-thrust belts.

Supplement 3

Detailed analysis of surface slope (α) and detachment angle (β) of the presented models

1. Model 1 (Reference model)

All models start with an early wedge-building phase, in which widening of the orogen results in a

160   prolonged and significant decrease in β and a highly variable trend in α. This initial phase reaches its

conclusion around 38 My, with the activation of the first major basement thrust.

After the wedge-building phase, the (α + β) value in the reference model slowly stabilizes, in

accordance with the predictions of critical taper theory, at a value of ~13° over time. This balance is

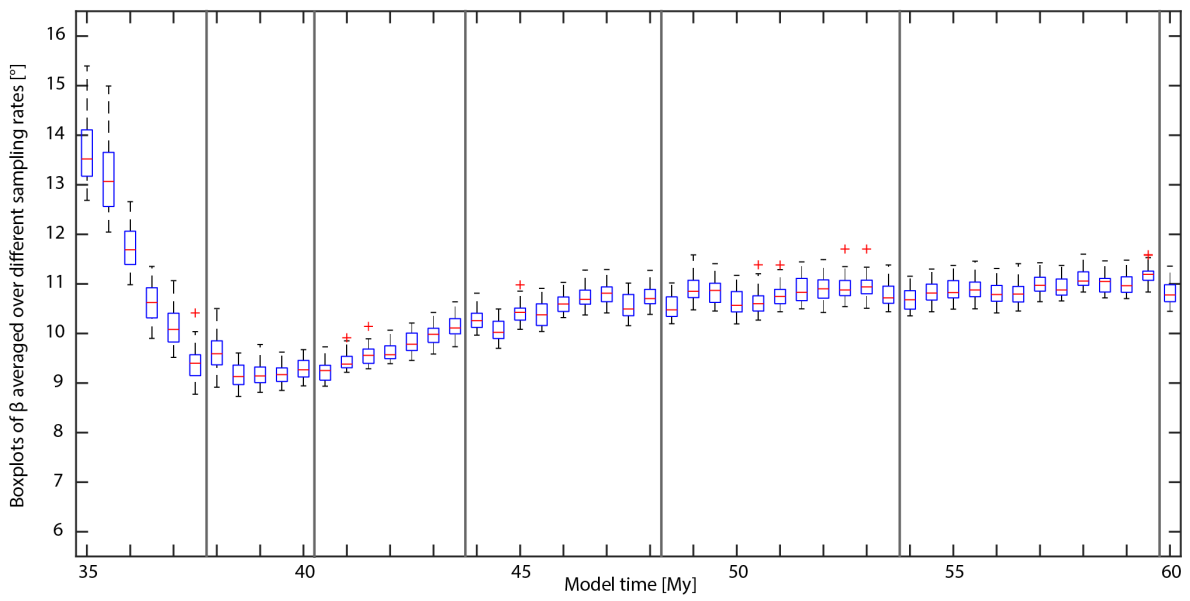
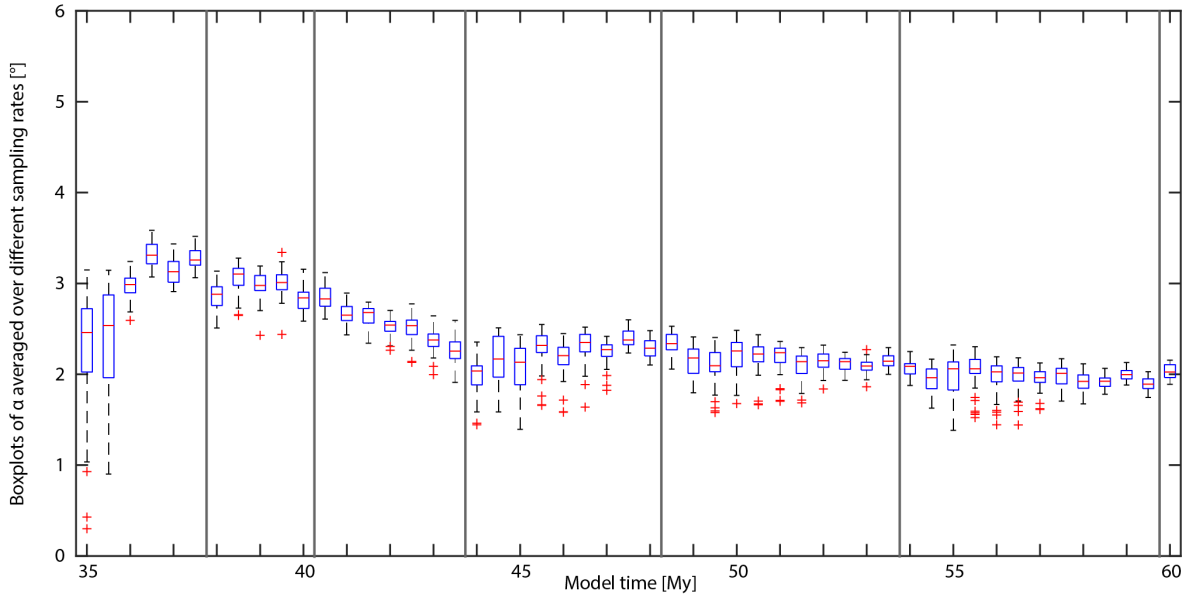
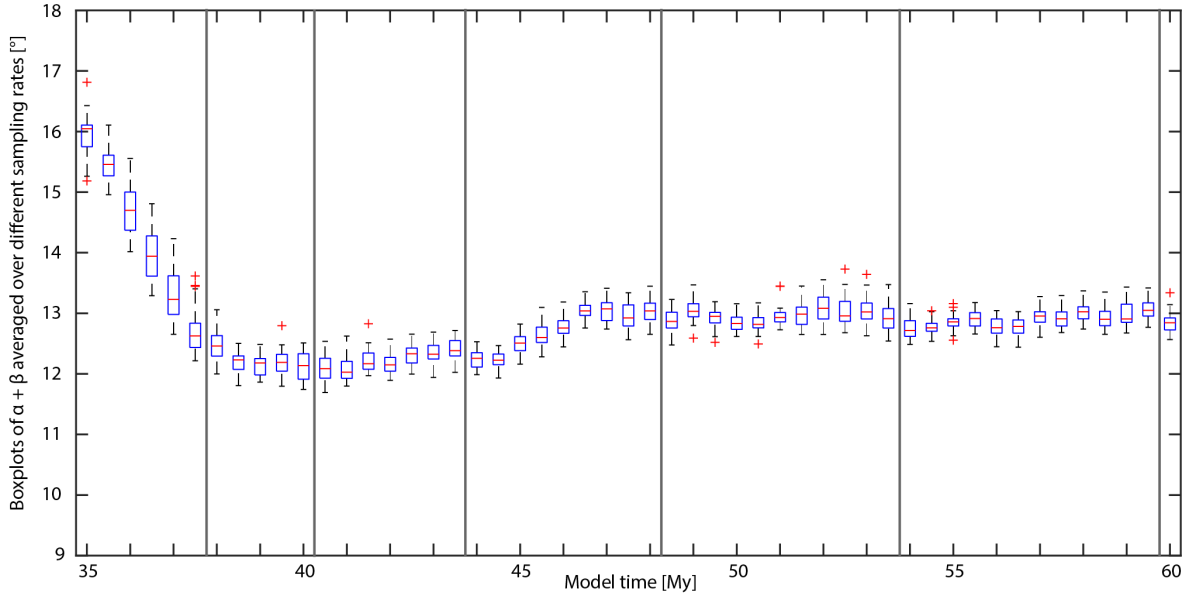
achieved through a constant, slow decrease in α and an opposing trend in β. In detail, the α + β plot

165   shows slight drops around the initiation of new basement thrust sheets, followed by slow increasing

trends as internal deformation accumulates before stepping out again in a manner very similar to that

described by Hoth et al. (2007) and Naylor and Sinclair (2007).

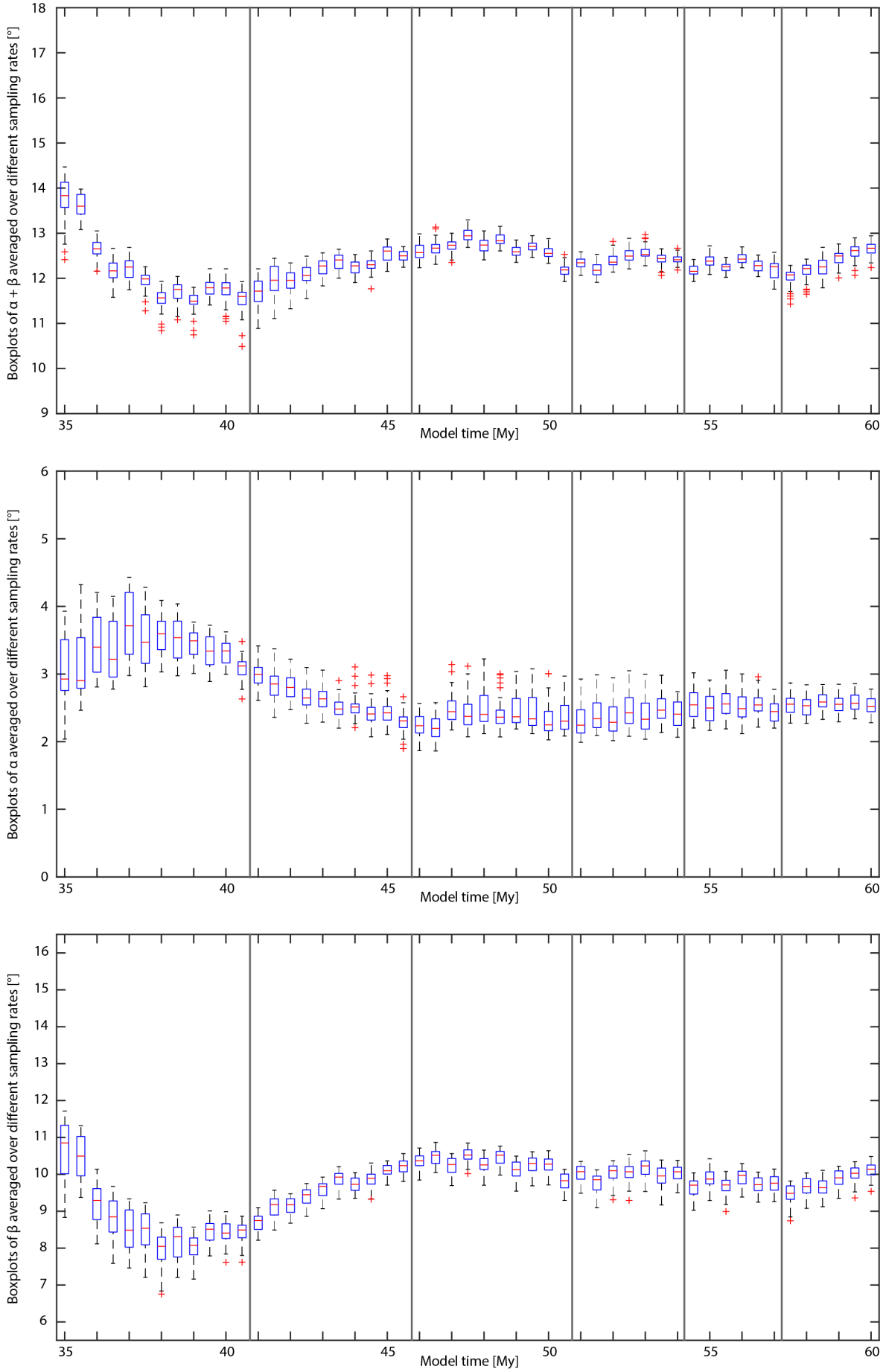
Figure A2. Plots of α + β, α and β vs. model time for Model 1. For each time-slice, the α and β values were determined using a
range of sampling intervals. The boxplots present the average α + β, α and β values of these individual sampling intervals
170   *calculated for the entire wedge. On each box, the central mark is the median, the edges of the box are the 25th and 75th*
percentiles, the whiskers extend to the most extreme data-points considered not to be outliers. The outliers are plotted
individually. The thick grey lines mark the initiation of thick-skinned thrusting events at the distal edge of the wedge.

Evolution of α, β and α + β values of Model 1.0 pro-wedge over time

2. Model 1.1

In Model 1.1 there is no sedimentation, but erosion is active from the start of the model. The pattern emerging in the α, β and α + β plots is rather similar to that observed in case of the Reference model. After the initial wedge-building phase, α + β stabilizes to a slightly lower value (~12,5°). The variations in the α + β plot observed in the reference model (i.e. drops at the formation of new basement thrust sheets followed by slowly increasing trends) are slightly more pronounced here.

The trends in the α and β plots are different from those observed in the reference model. α stabilizes at a somewhat higher value (2.5° compared to 2°) while β stabilizes at a somewhat lower value (10° compared to 11°) than the final values observed in the reference model. This is likely the result of the wedge-narrowing effect of erosion. The foreland basin is narrower; hence the wedge does not include a wide zone of gently dipping topography at its tip.

Figure A3. Plots of α + β, α and β vs. model time for Model 1.1. For each time-slice, the α and β values were determined using a range of sampling intervals. The boxplots present the average α + β, α and β values of these individual sampling intervals calculated for the entire wedge. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data-points considered not to be outliers. The outliers are plotted individually. The thick grey lines mark the initiation of thick-skinned thrusting events at the distal edge of the wedge.

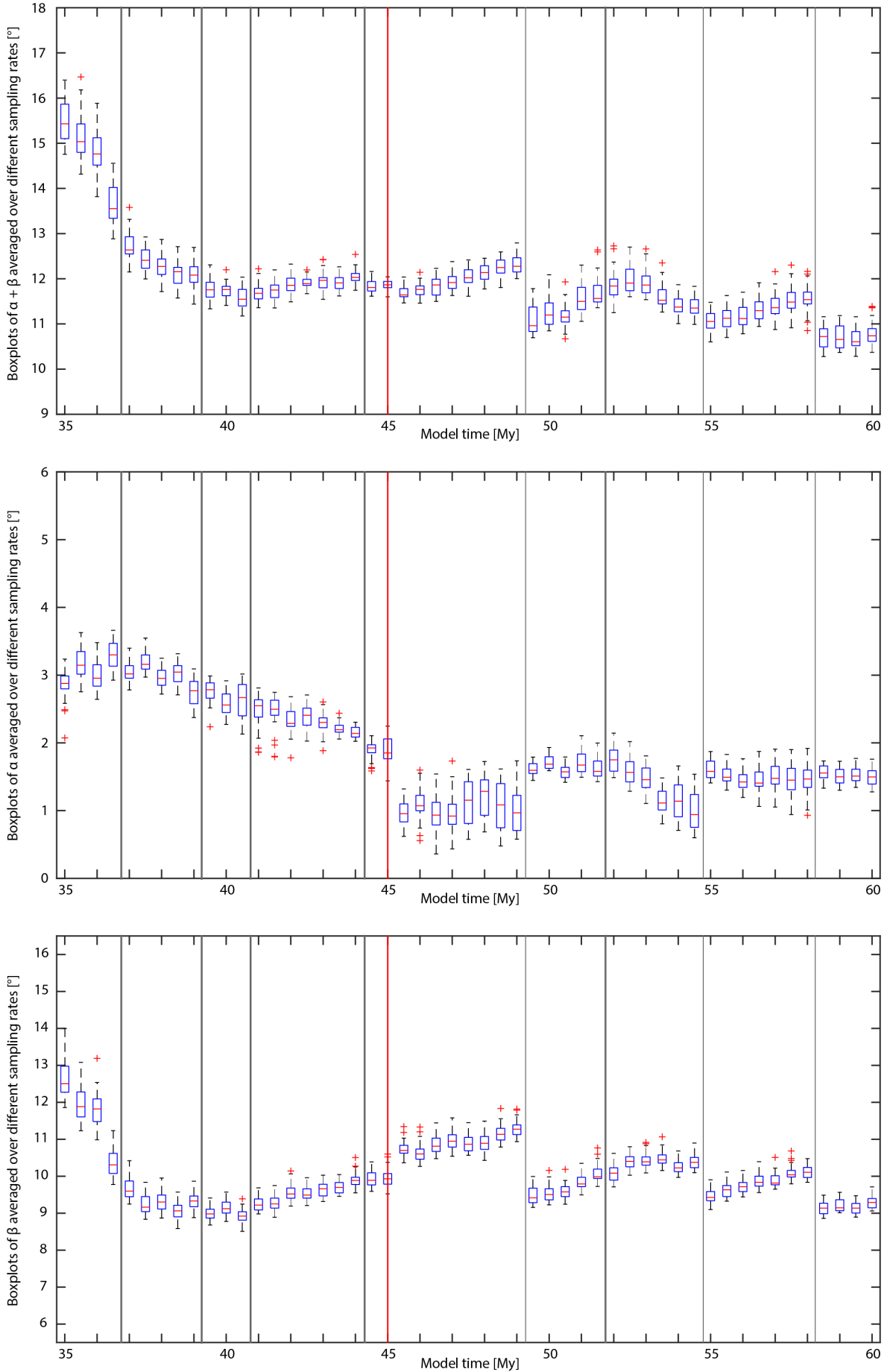
Evolution of α, β and α + β values of Model 1.1 pro-wedge over time

3. Model 2

In Model 2, sedimentation and erosion both initiate at 45 My (thick red line on figure A4). Similar to the previous models, the α + β plot stabilizes at a constant value of around 12° before the initiation of sedimentation. The subsequent variations in the α + β plot are mainly driven by variations in β.

After the initiation of sedimentation, there is a slight jump in β, which results from burial of the previously active frontal thrust under the newly deposited foreland sediments. As a result, the wedge temporarily narrows and a narrower gently dipping foreland is incorporated in the β-calculations. After this initial jump, β steadily increases as a result of the increasing crustal load exerted by the growing orogen. The loading increases β until deformation steps out to a new (thin-skinned) frontal thrust, widening the wedge and incorporating a previously undeformed gently dipping basement, which instantaneously reduces β.

The surface slope α takes up values between 1° and 2° (with relatively high uncertainty on the smaller values). Every time the wedge widens due to the emergence of a new (thin-skinned) frontal thrust, α jumps back up to a value around 2° (with a considerably lower statistical uncertainty).

Figure A4. Plots of α + β, α and β vs. model time for Model 2. For each time-slice, the α and β values were determined using a range of sampling intervals. The boxplots present the average α + β, α and β values of these individual sampling intervals calculated for the entire wedge. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data-points considered not to be outliers. The outliers are plotted individually. The thick red line marks the initiation of sedimentation. The thick grey lines mark the initiation of thick-skinned thrusting events and the thin grey lines mark the initiation of thin-skinned thrusting events at the distal edge of the wedge.

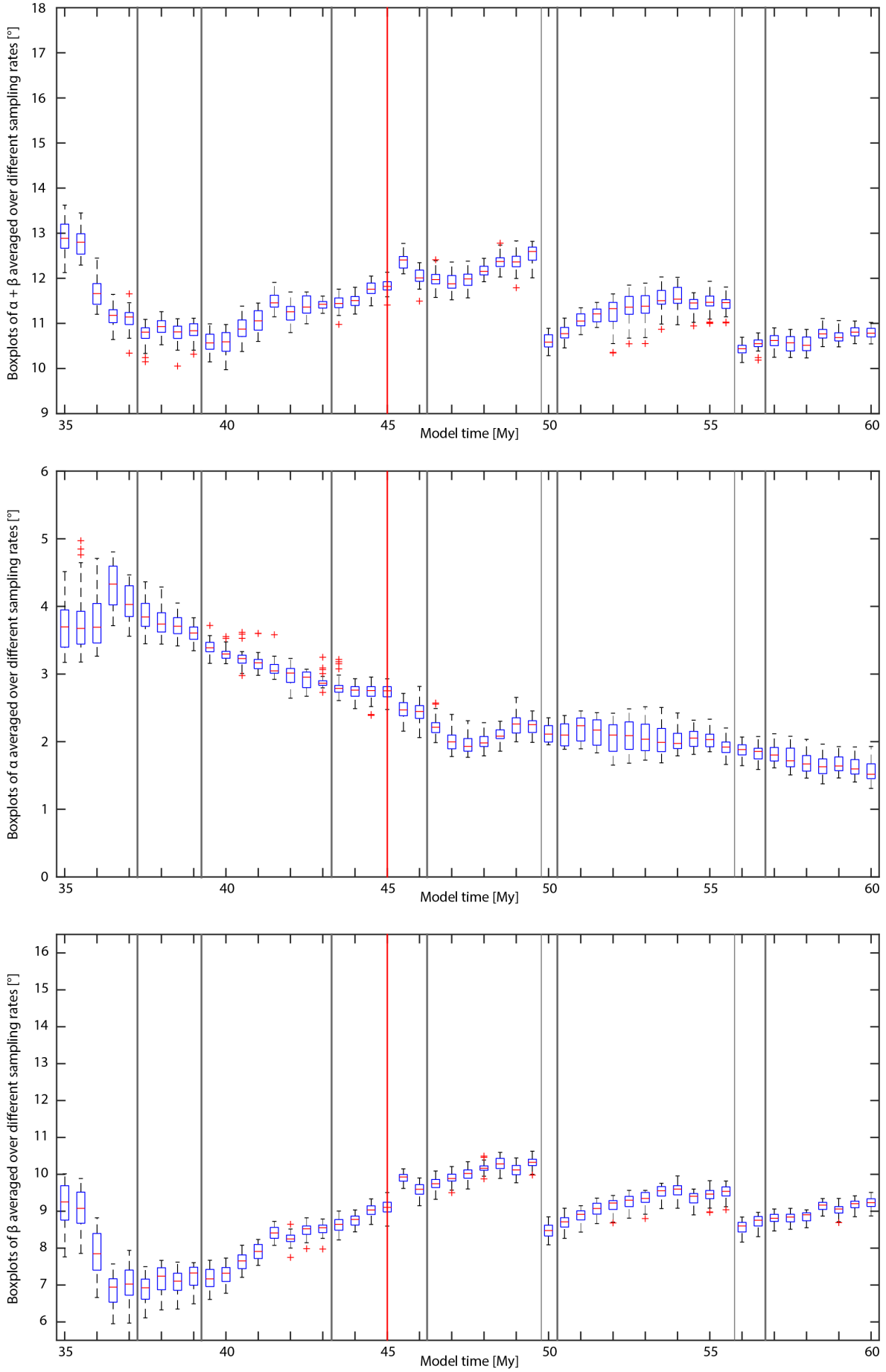
Evolution of α, β and α + β values of Model 2.0 pro-wedge over time

4. Model 2.1

In Model 2.1 erosion is not active, while sedimentation initiates at 45 My (thick red line on figure A5). Much like for Model 2, the α + β plot tends towards a value of around 12° before the initiation of sedimentation. The subsequent variations in the α + β plot are mainly driven by variations in β, apart from the long-term overall decrease that seems to be driven by a long-term decrease in α.

The relationship between the α + β plot and the β plot described for Model 2 (cf. Section 2.3) holds true for Model 2.1 as well. After the initiation of sedimentation, there is a slight jump in β as a result of the burial of the previously active frontal thrust, resulting in the temporary narrowing of the wedge. Subsequently, β steadily increases as a result of increasing orogenic loading. This loading increases β until the deformation steps out to a new (thin-skinned) frontal thrust, instantaneously reducing β.

The long-term decrease in α is only observed for models without erosion. We speculate that since models with erosion do not produce topography higher than 6 km, while models without erosion can grow topography as high as 8 km, this decreasing trend might be a result of the more pronounced increase in orogenic loading. The additional thickening of the wedge likely results in a hotter orogen which in turn is more prone to internal (ductile) deformation. This effect is likely enhanced by the increased width of the orogen, compared to models where erosion is active.

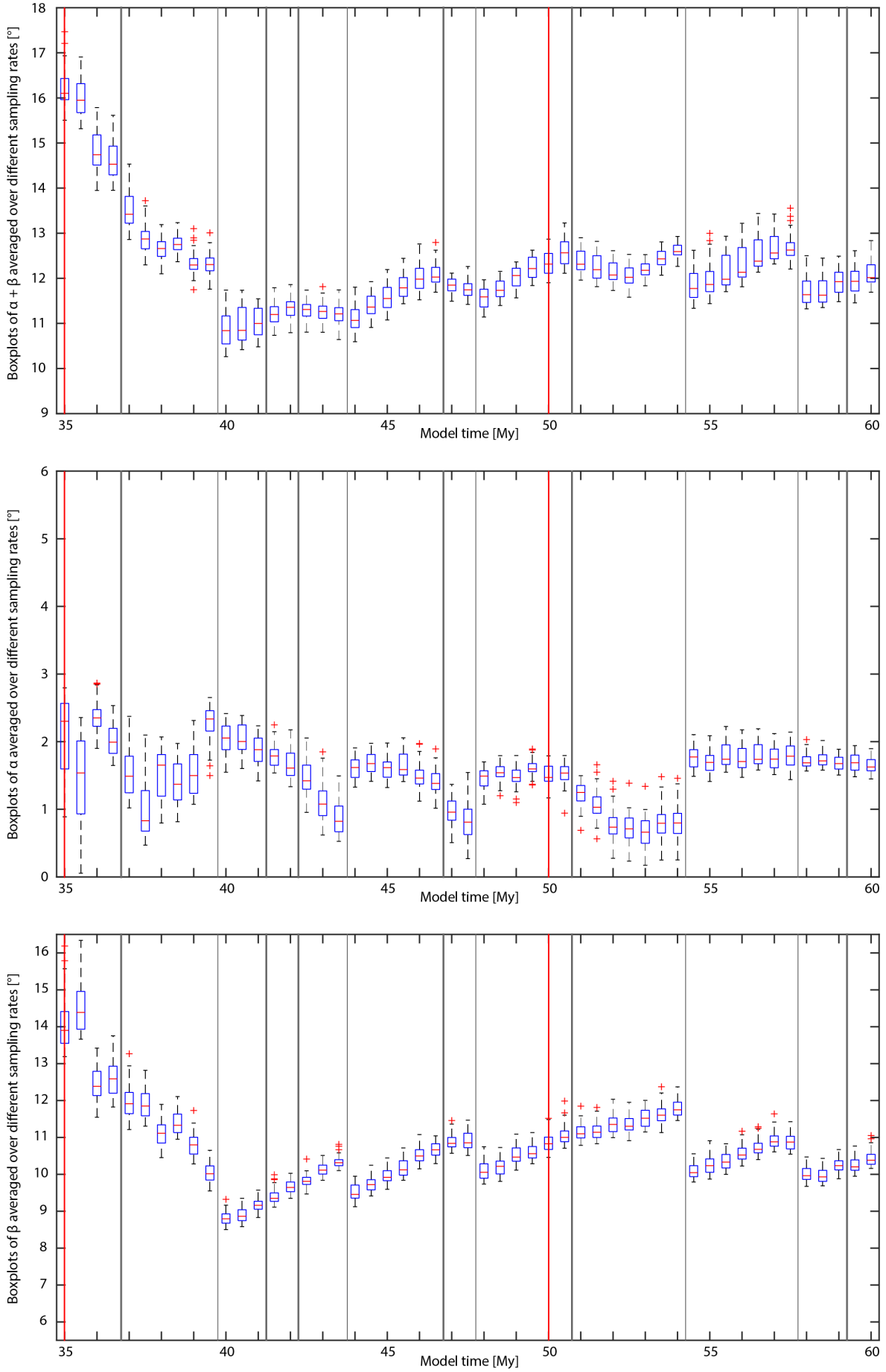
Figure A5. Plots of α + β, α and β vs. model time for Model 2.1. For each time-slice, the α and β values were determined using a range of sampling intervals. The boxplots present the average α + β, α and β values of these individual sampling intervals calculated for the entire wedge. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data-points considered not to be outliers. The outliers are plotted individually. The thick red line marks the initiation of sedimentation. The thick grey lines mark the initiation of thick-skinned thrusting events and the thin grey lines mark the initiation of thin-skinned thrusting events at the distal edge of the wedge.

Evolution of α, β and α + β values of Model 2.1 pro-wedge over time

5. Model 3

Model 3 is somewhat different from the previous models, as sedimentation initiates as early as 35 My and intensifies at 50 My in this model. As a result, α values remain extremely low (between 0.5° and 2°) throughout the entire model period. As in Model 2, the uncertainty in α values increases as the absolute values decreases.

Similarly, the patterns in β, observed in all the previous models that include sedimentation, are present in this model too. The change in the baselevel of sedimentation at 50 My leads to a very similar response as the initiation of sedimentation at 35 My.

Figure A6. Plots of α + β, α and β vs. model time for Model 3. For each time-slice, the α and β values were determined using a range of sampling intervals. The boxplots present the average α + β, α and β values of these individual sampling intervals calculated for the entire wedge. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data-points considered not to be outliers. The outliers are plotted individually. The thick red lines mark changes in the sedimentation rate. The thick grey lines mark the initiation of thick-skinned thrusting events and the thin grey lines mark the initiation of thin-skinned thrusting events at the distal edge of the wedge.

Evolution of α, β and α + β values of Model 3 pro-wedge over time

Supplement 4

255   **DOI of supplementary animations uploaded to the TIB AV-Portal**

Model 1: https://doi.org/10.5446/38571

Model 2: https://doi.org/10.5446/38572

Model 3: https://doi.org/10.5446/38576

Model 1.1: https://doi.org/10.5446/38573

260   Model 2.1: https://doi.org/10.5446/38574

Model 1: α-Topography-Strain rate: https://doi.org/10.5446/38575

Model 2: α-Topography-Strain rate: https://doi.org/10.5446/38577

Model 1: β-Topography-Strain rate: https://doi.org/10.5446/38578

Model 2: β -Topography-Strain rate: https://doi.org/10.5446/38579

265

References

Hoth, S., Hoffmann-Rothe, A., Kukowski, N., 2007. Frontal accretion: An internal clock for bivergent wedge deformation and surface uplift. Journal of Geophysical Research 112. B06408–17. doi:10.1029/2006JB004357

270   Naylor, M., Sinclair, H.D., 2007. Punctuated thrust deformation in the context of doubly vergent thrust wedges: Implications for the localization of uplift and exhumation. Geology 35, 559. doi:10.1130/G23448A.1

Bos, B., Spiers, C.J., 2002. Frictional-viscous flow of phyllosilicate-bearing fault rock: Microphysical model and implications for crustal strength profiles. J Geophys Res-Sol Ea 107. doi: 275   10.1029/2001jb000301

DeCelles, P.G., Giles, K.A., 1996. Foreland basin systems. Basin Research 8, 105-123.

Erdős, Z., 2014. Coupled surface process and tectonic modelling of extension-inversion tectonics in the Pyrenees. he University of Bergen, Bergen.

Erdős, Z., Huismans, R.S., van der Beek, P., 2015. First-order control of syntectonic sedimentation on 280   crustal-scale structure of mountain belts. Journal of Geophysical Research: Solid Earth 120, 5362-5377. doi: 10.1002/2014jb011785

Erdős, Z., Huismans, R.S., van der Beek, P., Thieulot, C., 2014. Extensional inheritance and surface processes as controlling factors of mountain belt structure. Journal of Geophysical Research: Solid Earth 119, 9042-9061. doi: 10.1002/2014jb011408

285   Gleason, G.C., Tullis, J., 1995. A Flow Law for Dislocation Creep of Quartz Aggregates Determined with the Molten-Salt Cell. Tectonophysics 247, 1-23.

Karato, S., Wu, P., 1993. Rheology of the upper mantle: a synthesis. Science 260, 771-778.

Philippe, Y., Colletta, B., Deville, E., Mascle, A., 1996. The Jura fold-and-thrust belt: a kinematic model based on map-balancing, in: Ziegler, P.A., Horvath, F. (Eds.), Peri-Tethys memoir 2; Structure and 290   prospects of Alpine basins and forelands. Paris, Ed. du Museum National d'Histoire Naturelle, pp. 235-261.

Sibson, R.H., 1990. Conditions for fault-valve behaviour. Geol. Soc. Spec. Publ. 54, 15–28.

Thieulot, C., 2011. FANTOM: Two- and three-dimensional numerical modelling of creeping flows for the solution of geological problems. Physics of the Earth and Planetary Interiors 188, 47-68. doi: 295   10.1016/j.pepi.2011.06.011