Two-stage Red Sea rifting inferred from mantle earthquakes in Neoproterozoic lithosphere

Alexander R. Blanchette, Simon L. Klemperer, Walter D. Mooney, Hani M. Zahran

A R T I C L E   I N F O

Article history:
Received 29 January 2018
Received in revised form 24 May 2018
Accepted 30 May 2018
Available online xxx
Editor: R. Bendick

Keywords:
mantle earthquakes
Red Sea rifting
Saudi Arabia
cratonic rejuvenation
lithosphere–asthenosphere boundary

A B S T R A C T

We use earthquake geothermometry, measured heat flow, and structural constraints from P-wave receiver functions to model the thermal evolution of the lithosphere beneath Harrat Lunayyir. We suggest that the lithosphere thinned to its present 60-km thickness in a second stage of lithospheric thinning at 15–12 Ma following initial Red Sea extension at ∼27 Ma. Harrat Lunayyir is an active volcanic field located in the Arabian Shield >150 km east of the Red Sea rift axis. In the lithospheric mantle beneath Harrat Lunayyir we locate 64 high-frequency earthquakes at depths of 42–48 km, all with \( m_L < 2.5 \).

These brittle-failure earthquakes must have nucleated at relatively low temperatures, based upon global maximum nucleation depths and temperature-dependent-deformation experimental results. Therefore, the mantle earthquakes show that the upper mantle lithosphere is not in thermal equilibrium with the shallow (60 km) underlying asthenosphere. Our thermal modeling indicates that the lithosphere beneath Harrat Lunayyir thinned to its current 60-km thickness at 12 ± 2 Ma, as constrained by thermal modeling of: (1) surface heat-flow; (2) the depth to the mid-crustal brittle–ductile transition, and (3) the depth to the upper-mantle brittle–ductile transition.

Published by Elsevier B.V.

1. Introduction

Earthquakes within the continental mantle lithosphere are sufficiently rare as to excite intensive study. These events are mostly found within zones of continental convergence or subduction (Chen and Molnar, 1985; Sloan and Jackson, 2012), and beneath non-magmatic rift zones (Yang and Chen, 2010), and plate-boundary strike-slip faults (Inhal et al., 2016). The seismicity of the Arabian plate is highly concentrated along its borders (Youssef, 2015), and the seismic quiescence of the Arabian interior conforms to the very definition of a craton (Mooney et al., 2012). Precambrian shields typically have sufficiently thick and cold lithosphere that mantle earthquakes could occur. However, their exceedingly low strain-rates, such as the very low rates within Arabia (Reilinger and McClusky, 2011), cause mantle earthquakes to be particularly rare beneath shields (Frohlich et al., 2015). Volcano-seismicity, with mantle and lower-crustal earthquakes likely caused by early-stage pooling and mixing of asthenospheric melt at the base of or within the lower-crust (McCausland et al., 2017), should also be rare in cratons, including Arabia, given their low heat-flow (Gosnold, 2011). However, here we report 64 high-frequency (brittle-failure) earthquakes within lithospheric mantle beneath the western margin of the Arabian Shield.

Understanding of Red Sea rift history has evolved beyond early ‘two-stage’ models for the Red Sea that include two stages of ocean spreading (at 41–34 Ma and at 5–4 Ma; Girdler and Styles, 1974) or two pulses of ‘initial’ or ‘early stage’ rifting (at ∼34 Ma and at 25–21 Ma; Omar and Steckler, 1995) to widespread acceptance that the first formation of Red Sea rift basins was in Late Oligocene time (28–24 Ma) (Bosworth, 2015). Our two-stage rifting model distinguishes the Late Oligocene initiation from a mid-Miocene second stage of major lithospheric thinning culminating at 12 ± 2 Ma and has much in common with a recent theoretical model that identifies two-stage evolution as a common feature of successful rifts (Brune et al., 2016). In this latter model, lithosphere that is initially cold and thick, hence strong, begins to rift slowly with full plate separation rates <10 mm/yr. As the lithosphere is extended further, it is thinned and heated from below, leading to a nonlinear decrease in lithospheric strength. This thinning and weakening – if it can outpace conductive cooling of the lithosphere...
2. Geologic/geophysical framework

The Arabian–Nubian Shield was assembled via repeated terrane accretion during the Neoproterozoic (Stoeser and Camp, 1985). After 500 Myr of relative stability, the shield began to break apart due to impingement of the Afar Plume at ∼30 Ma, which triggered extension first in the Gulf of Aden and then in the Red Sea by 25 Ma (Stern and Johnson, 2010). Initiation of shear on the Dead Sea transform in the middle Miocene and penecontemporaneous initiation of full ocean spreading in the Gulf of Aden facilitated a second phase of Red Sea rifting with rotation of extension from rift-normal (WSW–ENE) to Dead Sea transform-parallel (SSW–NNE) (Bosworth, 2015) at 11 ± 2 Ma as constrained geotectically (Reilinger and McClusky, 2011). Seafloor spreading began within the southern Red Sea at 5 Ma and has propagated northward to ∼20°N, but it has not yet reached the latitude of Harrat Lunayyir (25°N) (Bosworth, 2015).

Camp and Roobol (1992) report two distinct phases of volcanism within the Arabian plate. From 30–20 Ma tholeiitic-to-transitional lavas were emplaced along rift-parallel dikes extending the entire length of the Red Sea rift flank, contemporaneous with the impingement of the Afar plume on the base of the African–Arabian plate. From 12 Ma to the present, transitional-to-strongly-alkalic lavas have been emplaced along a N–S trend known as the Makkah–Madinah–Nafud (MMN) line of volcanism forming the younger harrats (Fig. 1) (Camp and Roobol, 1992; Stern and Johnson, 2010). Two stages of rifting and volcanism are also potentially reflected within the uplift history of the Arabian rift flank, with early uplift of a broad region of the Arabian–Nubian Shield progressively focusing to a narrow region around the Red Sea rift (Szymanski et al., 2016).

Mooney et al. (1985) analyzed seismic wide-angle reflection/refraction travel-times through heat-flow points 1, 2, 3 (Fig. 1) to the Red Sea and constrain the Arabian Shield upper-crustal P-wave velocity of 6.3 km/s and mid-to-lower crustal velocities of 6.5–6.7 km/s, with the Moho at ∼40 km depth. Not far below the Moho, both body-wave (e.g. Koulakov et al., 2016) and Rayleigh-wave (e.g. Yao et al., 2017) tomography show a strong decrease in S-wave velocity along the entire western margin of the Arabian craton, below ~60 km beneath Harrat Lunayyir and gradually deepening toward the plate interior. Comparison with S-wave receiver function converters (Hansen et al., 2007; Fig. 1) suggests this S-wave velocity decrease marks the lithosphere–asthenosphere boundary (LAB).

Measured heat flow within the Arabian Shield is ‘cratonic’ at ∼34 mW/m² despite the proximity of the Red Sea rift and spreading center, implying temperatures of ∼450–600°C at 40–50 km
depth (Gosnold, 2011). However, geothermometry on xenoliths from Harrat Kishhb (Fig. 1), on the eastern flank of the MMN line, suggests temperatures of ~900–1000 °C within this same depth range (McGuire and Bohannon, 1989). A single heat-flow measurement located near Harrat Lunayyir (#4, Fig. 1) is slightly elevated at 50 ± 4 mW/m², implying temperatures of 800–1000 °C at 40–50 km depth (Gosnold, 2011).

Volcanism within Harrat Lunayyir began at ~600 ka sourced by melt segregation within the asthenosphere at ~60–75 km depth, and since then eruption rates have decreased and depth of melt segregation may have somewhat decreased (Duncan and Al-Amri, 2013). Harrat Lunayyir last erupted during the 10th century C.E. In 2009, a roughly N–S striking dike (Fig. 2) of ~0.13 km³ reached within 1–2 km of the surface, and the associated >30,000 earthquakes, all with m_b ≤ 5.7, caused evacuation of the nearby town of Al-Ays (Pallister et al., 2010). This volcano-tectonic crisis led to densification of the existing seismometer array from three sites before 2009 to 18 sites by August 2010, which provide the data for our study.

3. Data and methodology

3.1. P-wave receiver functions

The Harrat Lunayyir seismic array now consists of 18 Trillium 120 broadband stations (Supplementary Table 1) and spans a region ~50 × 50 km (Fig. 2). We used data from this array to create P-wave receiver functions (PRFs) from 857 earthquakes recorded during 2008–2015 with epicentral distances of 30–90°, largely from back-azimuths of 020–120° and with moment magnitudes of 5.5–8.3 (Fig. 3).

We use the common conversion-point stacking method to obtain a 3D image of the discontinuity structure beneath Harrat Lunayyir (see Supplemental Materials). Our common conversion-point “cube” (Fig. 3) shows a clear, coherent and continuous positive polarity discontinuity (i.e., an upward decrease in seismic impedance for the P–Ps conversion) at 38 km below the surface. The average half-width of the pulse at half-maximum amplitude provides our uncertainty estimate of ±2 km. This positive polarity (Fig. 3) represents the Moho, in agreement with previous crustal thickness estimates (Hansen et al., 2007; Park et al., 2008; Mechie et al., 2013; Tang et al., 2016). At 60 km we observe a seismic converter with opposite polarity to that of the Moho (Fig. 3) that we interpret as the lithosphere–asthenosphere boundary (LAB), also consistent with previous lower-resolution estimates (Hansen et al., 2007; Park et al., 2008; Koulakov et al., 2016). The average half-width at half-maximum of the LAB signal is also ~2 km, but due to its less-focused appearance we use a more conservative uncertainty estimate of ±5 km. It is uncommon to image the LAB using P-wave receiver functions due to the interference of crustal multiples (Kind et al., 2012), but here we benefit from dense station coverage, numerous teleseismic events, and a mantle-lid that is significantly thinner than the crust. In stark contrast to the 60-km thickness beneath Harrat Lunayyir, a lithospheric thickness of 100–150 km is reported at the transition from the shield to the platform beneath central Saudi Arabia (Hansen et al., 2007) and even thicker lithosphere (≥ 200 km) beneath the eastern Arabian Peninsula (Koulakov et al., 2016).

3.2. Earthquake spatio-temporal patterns

The Saudi Geological Survey (SGS) provided an event-catalog for 2009–2014 (Supplementary Fig. S2) and continuous digital seismic data for the entire calendar year 2014. The 2014 event-catalog contains 72 earthquakes with unusual source depths of 40–50 km beneath Harrat Lunayyir and with local magnitudes of 0.2–2.5 (Fig. 4). We re-located the mantle earthquakes using hypocentre (Klein, 2002) to improve absolute locations and HypoDD (Waldhauser, 2001) to reduce relative errors between events, using the IASP91 velocity model (Supplementary Table S2; Kennett, 1991). Our relocations confirm that 64 of these catalog events are indeed below the 38-km deep Moho, all at depths of 42–48 km and all located ~30 km southeast of the 2009 dike (Figs. 2 and 4; Supplementary Table S3). Reported relative depth uncertainties from HypoDD (Waldhauser, 2001), using the SVD method, are ± 100 m. Absolute uncertainties depend on array geometry. Our preferred earthquake depths minimize the observed S- minus P-wave travel-time residuals with a median RMS error of <0.23 s (Fig. 5) across all 64 sub-Moho events. Because there exists only one station east of these earthquakes (TRAS, Fig. 5), we calculate the RMS of S-minus-P residuals (i.e., observed-minus-calculated travel-times) from all stations for each event, on a 1-km grid across a 1000 km² N–S swath encompassing the hypocenters and at depths every 0.33 km from the surface to ~70 km (i.e. 210,000 trial locations) (Fig. 5). A histogram of the number of source locations yielding a given RMS S-minus-P residual, plotted against the difference
in depth between the tested source location and our preferred depth (Fig. 5), shows a narrow minimum in the distribution of residuals, indicating that our better constrained earthquakes have depth uncertainties <5 km. Hence, we confirm that all 64 events are likely in the mantle, though the 2 ± 4 km gap between the Moho and earthquake locations (Fig. 4b and d) is likely apparent rather than real. Also during 2014, ~6000 earthquakes occurred at depths between 0–20 km, with the vast majority of events spatially associated with the 2009 dike (Figs. 2 and 4). There are three earthquakes in the 2014 SGS catalog with lower-crustal depths of 20–38 km. The waveforms for all three “lower-crust” events have emergent/noisy first arrivals, making precise locations difficult to determine. We conclude that the lower-crust at >20 km depth beneath Harrat Lunayyir is essentially aseismic.

The mantle earthquakes beneath Harrat Lunayyir are dominated by high-frequency (>5 Hz) energy, with a peak at ~30 Hz (Fig. 6). They have clear, impulsive P- and S-wave arrivals and match the characteristics of “brittle-failure” earthquakes (Hill and Prejean, 2005). Brittle-failure earthquakes close to sites of active magmatism are commonly referred to as volcano-tectonic because they are indistinguishable from tectonic earthquakes despite occurring around volcanoes. Brittle-failure earthquakes contrast with both long-period (LP) volcano seismicity (0.5–5 Hz), caused by resonance within fluid-filled cracks, and with very long-period (VLP) seismicity (0.01–~0.5 Hz), associated with the unsteady transport of magma and gases through conduits (Chouet and Matoza, 2013). High-frequency, LP and VLP crustal earthquakes all occurred during the 2009 diking episode (Pallister et al., 2010), but none of the 2014 mantle earthquakes we examined have LP or VLP characteristics.

The 2014 Harrat Lunayyir earthquakes occurred within the mantle lithosphere and follow a three-part temporal evolution (Fig. 7b and d) along a plane striking ~N103°E and dipping ~43°S (Fig. 4). First, between January and early May 2014, events with magnitudes between 0.5–1.5 occurred at a mean depth of ~45 km. Then, during an eight-day window in May 2014, ~50% of the total known earthquakes occurred, 0.6 ≤ mL ≤ 1.5, and quickly migrated to the shallowest observed depth of ~42 km. The subsequent month-long quiescence was broken in June 2014 when the deepest sequence began at ~46 km, increasing to 48 km, and encompassing the entire magnitude range (0.2–2.5) of the 2014 mantle earthquakes.

This three-part temporal trend does not correlate with the upper-crustal seismicity (Fig. 7d), which is roughly constant throughout the entire year. The temporal evolution of the earthquake locations matches “stage 1” volcano seismicity caused by magma-driven fracturing and pooling between the LAB and Moho (McCausland et al., 2017). Similar patterns of seismicity are present in the SGS catalog between August 2010 and the end of 2013 (Fig. 7a and c), indicating that the 2014 sequence is not an isolated occurrence and that there are likely numerous pulses of magmatism intruding close to the base of the crust.

3.3. Derivation of finite-differences thermal modeling scheme

We used a one-dimensional finite-differences code (see Supplementary Materials) to model the thermal evolution of the lithosphere as it is eroded from below, following Turcotte and Schubert (2002) and Recktenwald (2004). We start with the one-dimensional time-dependent heat-conduction equation (Turcotte and Schubert, 2002, eqn. (4.68)), with the inclusion of radiogenic heat production:

$$\frac{\partial T}{\partial t} = \kappa \left( \frac{\partial^2 T}{\partial z^2} + \frac{A(z)}{k} \right)$$

where $k = \frac{k_{nt}}{\rho_c}$ is the thermal diffusivity and $A(z)$ is the volumetric heat production of the medium. If we assume that the concentration of heat producing elements decays exponentially as a function of depth from the surface with a characteristic decay constant $h_c$, we can write the volumetric heat production as:

$$A(z) = A_0 e^{-z/h_c}$$

Substituting equation (2) into equation (1) we obtain:

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial z^2} + B e^{-z/h_c}$$

where $B = \frac{k A_0}{\rho_c}$ is a constant related to the medium. This is the one-dimensional heat-flow equation with internal heat production that we next solve using the centered-space finite-differences approximation for the term involving the second spatial derivative of the temperature:

$$\left. \frac{\partial^2 T}{\partial z^2} \right|_{z=m} \approx \frac{T_{m-1} - 2T_m + T_{m+1}}{\Delta z^2}$$

where $m$ indicates the spatial node on a discrete mesh. Next, we use the forward-time finite-differences approximation for the left-hand side of equation (3):

$$\frac{\partial T}{\partial t} \left|_{t=n} \right. \approx \frac{T_{n+1} - T_n}{\Delta t}$$

where $n$ indicates the temporal node on a discrete mesh. In equations (4) and (5), $\Delta z$ and $\Delta t$ are the distances between spatial and temporal nodes, respectively. Substituting equations (4) and (5) into equation (3) produces:

$$\frac{T_{m+1} - T_m}{\Delta t} = \kappa \frac{T_{m-1} - 2T_m + T_{m+1}}{\Delta z^2} + B e^{-z_m/h_c}$$

where superscripts denote temporal nodes and subscripts are spatial nodes. Rearranging the terms to solve for the temperature at spatial node $m$ in the next time step, $n+1$, we get:
Fig. 5. Test of absolute depth uncertainty. a) Map showing a portion of Harrat Lunayyir, the easternmost station (TRAS), and the mantle earthquake epicenters. Black box marks the bounds of the North–South grid search of S minus P RMS residuals. b) Residuals for all events at the inverted locations. Red line at 0.23 s is the median of the residuals. c) Two-dimensional histogram of the number of occurrences of S minus P RMS residuals in 0.025 s bins at all possible Δ depths, in 0.33 km bins. The Δ depth indicates the difference between the depth of each event location in our grid search and the preferred depth of the event, where a negative change in Δ depth indicates a test location shallower than the preferred location. The grid search was performed over a latitude range of 24.8–25.5°N and longitude range of 38.0–38.1°E. The color-scale reflects a count of the number of residuals at a given change in depth across all events and test locations. The inset is a zoomed view of the minimum of the residuals.

\[
T_{m}^{n+1} = T(z_m, t_{n+1})
= T_m^n + \frac{\kappa \Delta t}{\Delta z^2} (T_{m-1}^n - T_m^n + T_{m+1}^n) + \Delta t Be^{-z_m/h_t}
\]

We simplify the equation by setting:

\[
r = \frac{\kappa \Delta t}{\Delta z^2}; \quad s = 1 - 2r; \quad C = \Delta t B
\]

This is an explicit equation for the temperature at spatial node \( m \) and time step \( n + 1 \) only in terms of the previous time step. The constant \( r \) is the stability parameter for the finite differences scheme. The solution is stable on the condition that \( r \leq \frac{1}{2} \) (Recktenwald, 2004).
To obtain an input temperature profile, we use the solution for the steady-state heat-flow equation with internal heat generation (Turcotte and Schubert, 2002, eqn. (4.30)):

$$T(z) = T_0 + \frac{q_m z}{k} + \frac{A_0 h_r^2}{k} (1 - e^{-z/h_r})$$

To solve for $q_m$, the reduced heat flow, we apply the boundary condition at the surface ($z = 0$) as shown in equations (10) and (11) below:

$$-q(z = 0) = q_0 = q_m + A_0 h_r$$

$$q_m = q_0 - A_0 h_r$$

In order to use equation (8) to solve equation (3), we need two boundary conditions: temperatures at the surface ($T_0$) and at the lithosphere–asthenosphere boundary ($T_{LAB}$ at depth $z_{LAB}$). The surface boundary condition is simply the average temperature at the surface (25°C, see Supplemental Table S4). Mantle temperature depends on the mantle potential temperature, $T_p$, and the adiabatic temperature gradient within the mantle:

$$T = T_p + \frac{z}{dT/dz}; \quad z \geq z_{LAB}$$

We assume that the seismic lithosphere–asthenosphere boundary is coincident with the thermal definition of the lithosphere–asthenosphere boundary as the intersection between the conductive lithospheric thermal profile and the adiabatic mantle–asthenosphere thermal profile. If, as we show in this paper, the lithosphere is not in thermal equilibrium with the underlying asthenosphere due to geologically recent thermal erosion, then we
would indeed expect a narrow seismic transition zone (seismic LAB) at the thermal lithosphere–asthenosphere boundary (Aulbach et al., 2017).

3.4. Thermal constraints

The maximum depth at which intra-continental earthquakes nucleate is controlled by strain-rate, lithology and temperature (e.g. Chen and Molnar, 1983). Reilinger and McClusky (2011) report that strain-rates within the Arabian Shield are below their detection threshold (<10^{-15} s^{-1}). Due to our interest in the mantle lithosphere, we restrict our analysis to olivine-dominated lithologies. The clear cut-off in earthquakes at 48 ± 2 km depth below Harrat Lunayyir shows that at this depth the modern temperature is at the maximum at which brittle-failure earthquakes can occur, which must be <900–1000 °C as measured from xenoliths at Harrat Kishib (McGuire and Bohannon, 1989).

Chen and Molnar (1983) suggest that crustal and mantle seismicity are limited to temperatures less than ~350 ± 100 and ~700 ± 100 °C, respectively, thought to represent brittle–ductile transitions in the crust and upper mantle. More recently, a 600 °C cut-off in seismicity in the mantle has been estimated for oceans (McKenzie et al., 2005) and continents (Sloan and Jackson, 2012). However, in volcanic areas, due to elevated strain-rates, lower-crustal brittle-failure earthquakes are well known from regions thought to be hotter (700–750 °C; Shelly and Hill, 2011; White et al., 2011). Ágústsson and Flóvenz (2005) used geothermal data to constrain the maximum nucleation depths of earthquakes beneath Iceland to temperatures of 600–900 °C. Although lithological considerations alone suggest mantle-lithosphere brittle-failure earthquakes should therefore be common in continental volcanic areas, they are only rarely reported. For example, there are no brittle-failure events beneath the Cascade volcanic arc where LP earthquakes are recognized around Moho depth at temperatures in excess of ~800 °C (Hansen et al., 2016). Beneath Hawaii, tectonic earthquakes thought to be caused by volcano loading and lithospheric flexure are known to occur as deep as 50 km along a long-lived tectonic fault zone (Wolfe et al., 2003). Recently, Inbal et al. (2016) located small earthquakes in southern California along the Newport–Inglewood fault below the Moho where temperatures are ~800 °C, correlating with geochemical evidence of a fluid pathway from the mantle (Boles et al., 2015). Prieto et al. (2017) suggest that a mantle earthquake beneath the Wyoming craton at 75-km depth occurred at temperatures of 750–850 °C due to strain localization in the nominally ductile regime, ~100 km away from the nearest volcanic center.

Laboratory experiments to determine rate-and-state friction laws for olivine are consistent with these earthquake observations, suggesting a temperature cut-off of ~600 °C at strain rates of 10^{-15}–10^{-14} s^{-1} for typical earthquakes, with the possibility that low-asperity stresses could extend this temperature to ~800 °C (Boettcher et al., 2007; King and Marone, 2012). Ohuchi et al. (2017) detected faulting associated with accelerated strain localization and localized heating in very fine-grained shear zones in dry dunite and extrapolate their laboratory observations to imply the possibility of faulting in the mantle lithosphere at temperatures up to 900 °C at strain rates as low as 10^{-14} s^{-1}. Thus, earthquake geothermometry suggests temperatures of 800 °C, or even 900 °C, as the cause of our observed 48-km seismicity cut-off.

There are no xenoliths reported from Harrat Lunayyir, but at Pleistocene Harrat Kishib, 150 km east of the MMN line (Fig. 1), spinel peridotites suggest temperatures of 910–970 °C at 13–20 kb (McGuire and Bohannon, 1989), corresponding to the entire depth range from the Moho to the LAB as imaged in this paper. Seismic wave-speeds, generally only a weak indication of temperature, suggest the mantle lithosphere beneath Harrat Lunayyir is at 800 ± 200 °C (Tang et al., 2016). Finally, our observation of seismograms with spectrogram peaks >15 Hz (corresponding to short-wavelength S-waves, ~300 m) (Fig. 6) indicates moderate to high Q in the source region. The sharp increase in seismic attenuation within olivine-dominated rocks at ~1000 °C (Artemieva et al., 2004) allows us to confidently exclude temperatures ≥1000 °C at depths shallower than the seismicity cut-off beneath Harrat Lunayyir. We conclude that the ambient temperature at 48 ± 2 km depth beneath Harrat Lunayyir is 850 ± 50 °C.

In order to deduce the timing of lithospheric thinning beneath our study area, we assume the thermal structure beneath Harrat Lunayyir prior to any Red Sea rifting or plume impingement matches that in the center of the Arabian craton, which we derive from measured surface heat flow and heat production (measurement #1, Fig. 1; Gosnold, 2011) and assuming a mantle potential temperature of 1354 °C (as petrologically estimated on the flanks of the MMN line; Camp and Roobol, 1992). We solve for the depth to the LAB numerically and find the original depth to be 113–156 km, in accord with the modern 100–150 km values from S-wave receiver functions at the transition from the shield to the platform (Hansen et al., 2007). We use 1D finite-differences to model the temperature evolution beneath Harrat Lunayyir as and after the lithosphere is thinned to our observed 60 km thickness (Fig. 8). We assume the lithosphere is replaced from below by asthenosphere with an adiabatic gradient of 0.3 °C/km, and we include depth-dependent radiogenic heat production and thermal conduction throughout model time and space. Our simple thermal model is insensitive to the duration of thinning so long as the duration is less than ~70 Myr, so that conduction has enough time to impact isotherms shallower than those being eroded from beneath.

Using the 350 ± 50 °C and 850 ± 50 °C isotherms as cut-off temperatures for seismicity within the upper-crust and mantle lithosphere, respectively, along with a present-day surface heat flow of 50 ± 4 mW/m² (measurement #4, Fig. 1; Gosnold, 2011), we can place upper (14 Ma) and lower (10 Ma) bounds on how long ago the LAB must have reached its present, shallow depth of 60 km.

4. Results and discussion

Fig. 8d shows that by using the full range of geothermal models from Gosnold (2011), a temperature of 800 °C (900 °C) at 48 ± 2 km depth is reached 2–5 Myr (4–11 Myr) after the lithosphere has thinned to 60 km depth. Although this time range is broad (2–11 Myr), it is nonetheless instructive. The youngest volcanotectonic event that can possibly have caused local lithospheric thinning, the onset of Harrat Lunayyir volcanicism, is dated at 600 ka (Duncan and Al-Amri, 2013). However, if the lithosphere thinned to 60 km as recently as 600 ka, the temperature at 48 km depth would now be only 600 ± 50 °C (geotherm at time = 0.6 Myr, Fig. 8b), and we would expect to see deeper earthquakes within the mantle lithosphere. The surface heat flow would also not have had time to increase from the initial 30 mW/m² to 50 mW/m². Hence, the onset of Harrat Lunayyir volcanicism at 600 ka was not the cause of local thinning of the lithosphere to the modern LAB depth.

The oldest magmatic event likely to have thinned the cratonic lithosphere was the impingement of Afar plume-head asthenosphere at ~30 Ma and the initial extension of the Red Sea by 25 Ma (Stern and Johnson, 2010; Bosworth, 2015). However, if the lithosphere beneath Harrat Lunayyir had thinned to its present 60 km as early as 25 Ma, the temperature at 48 km depth would now be 1050 ± 50 °C (geotherm at time = 25 Myr, Fig. 8b), and we would not have brittle-failure earthquakes at that depth nor most likely anywhere in the mantle lid. Furthermore, we would ex-
pect high seismic attenuation, inconsistent with our observation of high-frequency earthquakes, as well as significantly elevated surface heat flow values of $\sim 70$ mW/m$^2$, far higher than observed (Fig. 8a). Hence, early Red Sea extension cannot have thinned the Arabian lithosphere to its present 60-km thickness beneath Harrat Lunayyir.

The implication of the brittle-failure mantle earthquakes is that a second phase of lithospheric thinning, as well as uplift and likely extension, must have taken place in the Middle and/or Late Miocene. The younger harrat volcanism along the MMN line began at $\sim 12$ Ma (Camp and Roobol, 1992) including other harrats at the latitude of Harrat Lunayyir (e.g., Harrat Kura, 11 Ma; Stern and Johnson, 2010).apatite and zircon (U–Th)/He ages from the rift margin across Harrats Lunayyir and Khaybar (Fig. 1) suggest renewed extension between 15–12 Ma (Szymanski et al., 2016). If the Arabian lithosphere thinned to 60 km at 12 Ma, then the coldest of our initial geothermal profiles would only now have reached the highest permissible temperature (900 °C) at the shallowest permitted seismic cut-off depth (46 km). Our observed earthquake cut-off depth in the mantle is only defined by a relatively short duration of recording; however, if we had observed deeper earthquakes, it would require more recent thinning of the lithosphere. Our choice of a limiting temperature of 850 ± 50 °C is higher than the 600–700 °C that is often considered the cut-off of brittle behavior in the upper mantle, but using the lower temperatures would only permit a far more recent thinning of the lithosphere. Hence both scenarios (deeper earthquakes; lower temperature limit) would imply that the lithosphere is even more out of equilibrium and would more strongly support two-phase rifting.

The cut-off in crustal seismicity at 20 km depth is far less diagnostic of the timing of the completion of lithospheric thinning because the 350 ± 50 °C isotherms in our models are so much flatter than the 850 ± 50 °C isotherms (Fig. 8c). Even modest uncertainty in the appropriate cut-off isotherm that arises due to crustal lithologic variability (Magistrale and Zhou, 1996) leads to significant uncertainty in the timing of crustal heating. However, our initial geothermal models, prior to thinning, place the 350 ± 50 °C isotherms at $\sim 30$–45 km depth, clearly deeper than our mid-crustal cut-off in seismicity. Considering the full range of acceptable initial heat flow and heat production estimates (Gosnold, 2011) permits a crustal cut-off in seismicity at 20 km depth. Hence, we prefer models in which shallowing of the LAB is sufficiently far in the past (before $\sim 11$ Ma; Fig. 8c) for mid-crustal temperatures to be sufficiently elevated.

Acceptable solutions using the mean initial geothermal parameters for the time at which lithospheric thinning reduced the lithosphere to the observed 60 km thickness must fit our heat flow constraints (8–14 Ma; Fig. 8a), the maximum depth of crustal earthquakes (11–23 Ma; Fig. 8c), and the maximum depth of mantle earthquakes (2–13 Ma; Fig. 8d). Therefore, acceptable solutions must be at least 11 Ma, but can be no older than 13 Ma. Thinning ages <11 Ma would imply that the crustal seismicity is limited to regions at temperatures <300 °C (implausible in anhydrous Proterozoic basement; Chen and Molnar, 1983). Thinning ages >13 Ma would imply that the mantle earthquakes occur at temperatures...
in excess of 900 °C, hotter than previously observed or proposed (Chen and Molnar, 1983; McKenzie et al., 2005). Tests of model sensitivity to various thermal parameters (see Supplemental Materials) suggest that thinning completed at ~12 ± 2 Ma.

Although our modeling cannot formally exclude continuous thinning from 30 Ma (or earlier) to ~12 Ma, this ‘continuous’ scenario offers no ready explanation for the hiatus in magmatism and extension from 20–15 Ma (Camp and Roobol, 1992; Stern and Johnson, 2010). Indeed, a common feature of continental rifts that progress to ocean spreading is that they evolve in two separate phases (Brune et al., 2016), where initially strong lithosphere is slowly thinned and dynamically weakened through rift-induced heating, initiating a second and much faster phase. The second, faster rifting phase usually begins ~10 Myr before inception of breakup of the continental lithosphere and persists until complete plate separation is achieved. The second stage is not observed in failed rifts, where conductive cooling strengthens the lithosphere faster than it can be dynamically weakened, leading to rift failure (Brune et al., 2016).

In our preferred scenario the first stage of extension, uplift, and magmatism between 30–20 Ma was contemporaneous with modest lithospheric thinning but failed to rupture the continent and represents the slow phase of rift-evolution. The thermal constraints are, unfortunately, insensitive to the intermediate lithospheric thickness achieved during the first stage given the timescale at which heat diffuses. The change in plate boundary conditions in the middle Miocene (i.e., the onset of full ocean spreading in the Gulf of Aden and of the onset of Dead Sea transform motion) allowed rotation and an increase of the extensional stresses along the Red Sea rift to be more favorable for the second-stage of rifting (Reilinger and McClusky, 2011). Lithospheric thickness beneath Harrat Lunayyir decreased to 60 km between 14–10 Ma, largely by replacement of the lithosphere with asthenosphere, initiating the younger harrat volcanism. This second phase of lithospheric thinning must then have essentially stopped beneath Harrat Lunayyir to preserve the observed gap in seismicity between 48 km and the 60-km deep LAB. We suggest thinning beneath Harrat Lunayyir concluded as rifting became more focused within the Red Sea (cf. Szymanski et al., 2016).

5. Conclusions

Our seismic imaging locates the Moho beneath Harrat Lunayyir at 38 ± 2 km depth and the LAB at 60 ± 5 km depth, despite being located within a Precambrian shield. Observations of earthquakes within the lithospheric mantle beneath Harrat Lunayyir span the entire time period for which a suitably dense seismic array was active (i.e. since August 2010). During 2014, there were 64 high-frequency “brittle-failure” type earthquakes whose hypocenters, well constrained to the upper 10 km of the lithospheric mantle (depths of 42–48 km), lie on a plane striking ~N103°E and dipping ~43°S, located roughly 30 km to the southeast of the 2009 dike intrusion. These brittle-failure earthquakes are likely caused by melt migration within the lithospheric mantle. The existence of these brittle-failure earthquakes constrains the temperature at 48±2 km depth to be 850 ± 50 °C, which in turn requires the LAB to have reached its present shallow level in middle Miocene time (12 ± 2 Ma). Our results are consistent with a two-stage model for the evolution of continental rifting, as suggested by Brune et al. (2016) globally and as proposed by Szymanski et al. (2016) for the Red Sea rift. Our identification of high-frequency earthquakes in the shallow lithospheric mantle provides a powerful means of estimating the thermal structure of the lithosphere and consequently the timing of continental rifting in the Red Sea.

Acknowledgements

We thank Jeanne Hardebeck, David Shelly and Tom Parsons for their invaluable comments during the USGS internal review; Gary Chulick and David Hill for informal review comments on this manuscript; and Bernard Chouet and Katie Keranen for insightful discussions. We would like to thank Derek Keir and an anonymous reviewer for their thorough reviews and excellent suggestions that have substantially improved the manuscript.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.epsl.2018.05.048.

References


