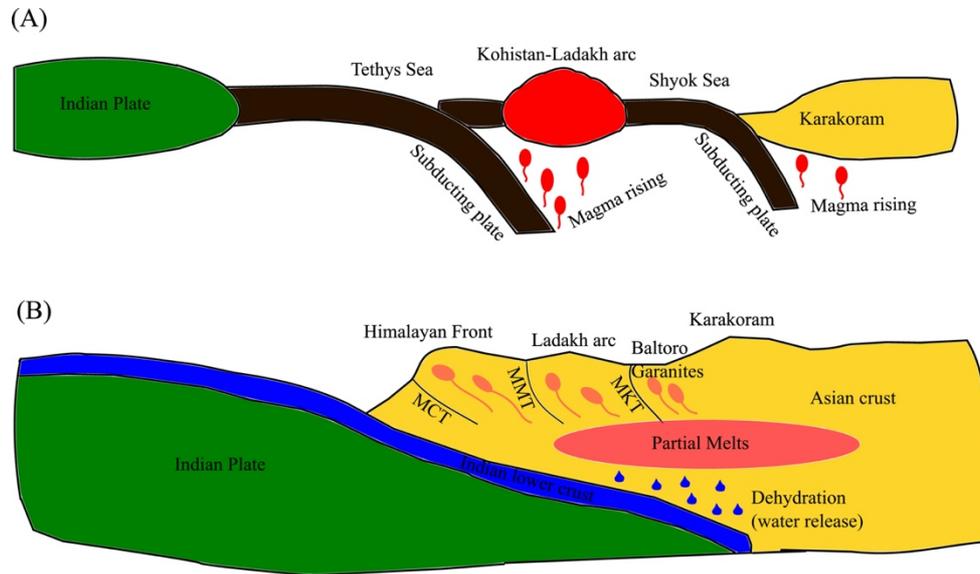
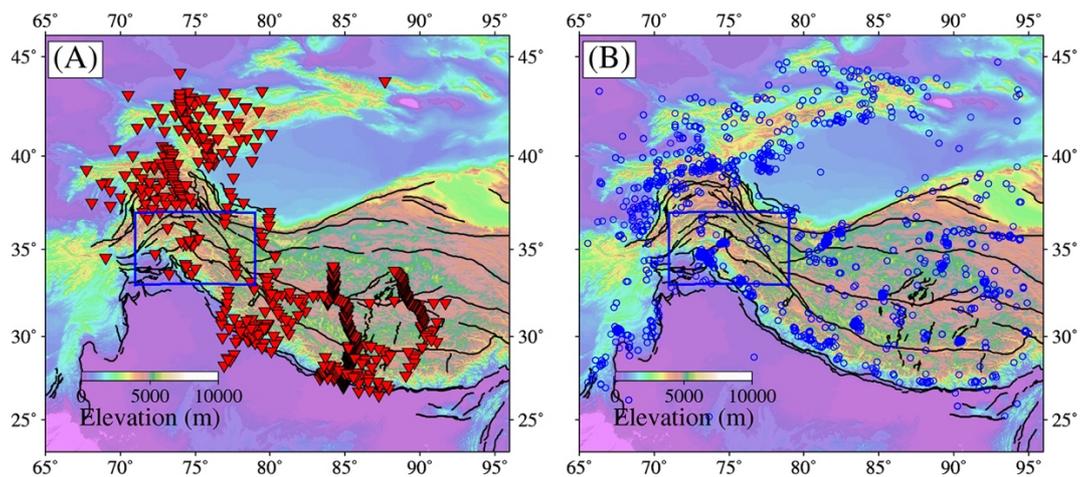


## Supplementary Material

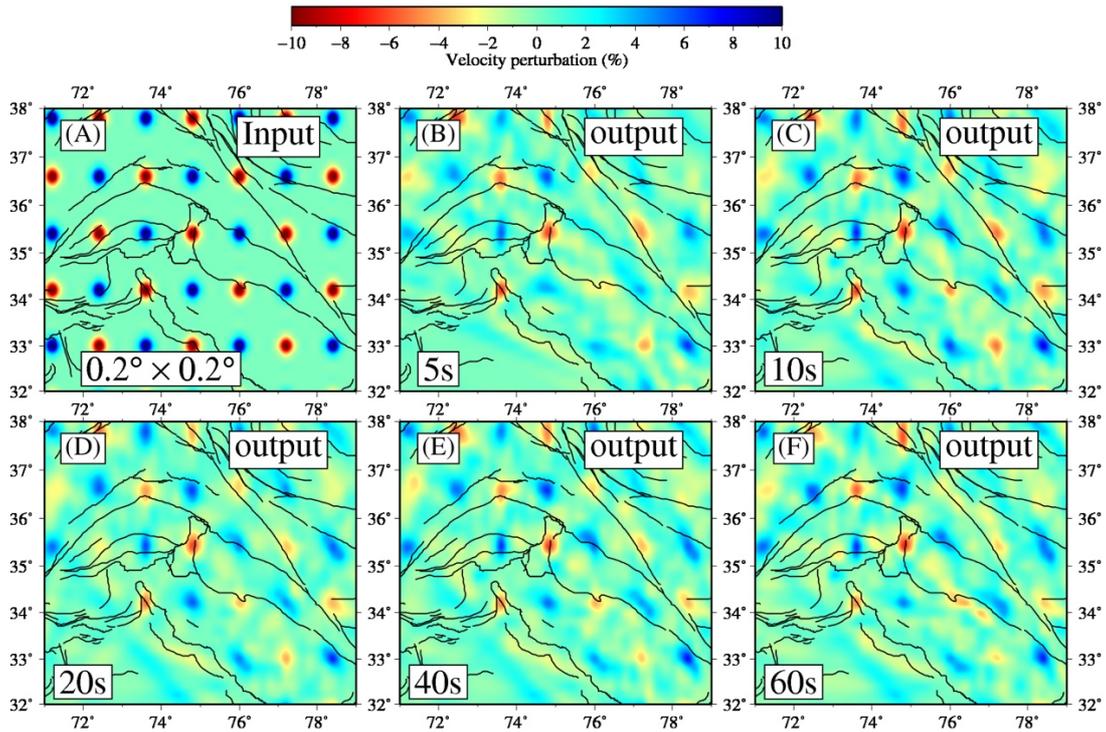
### 1 Supplementary Figures



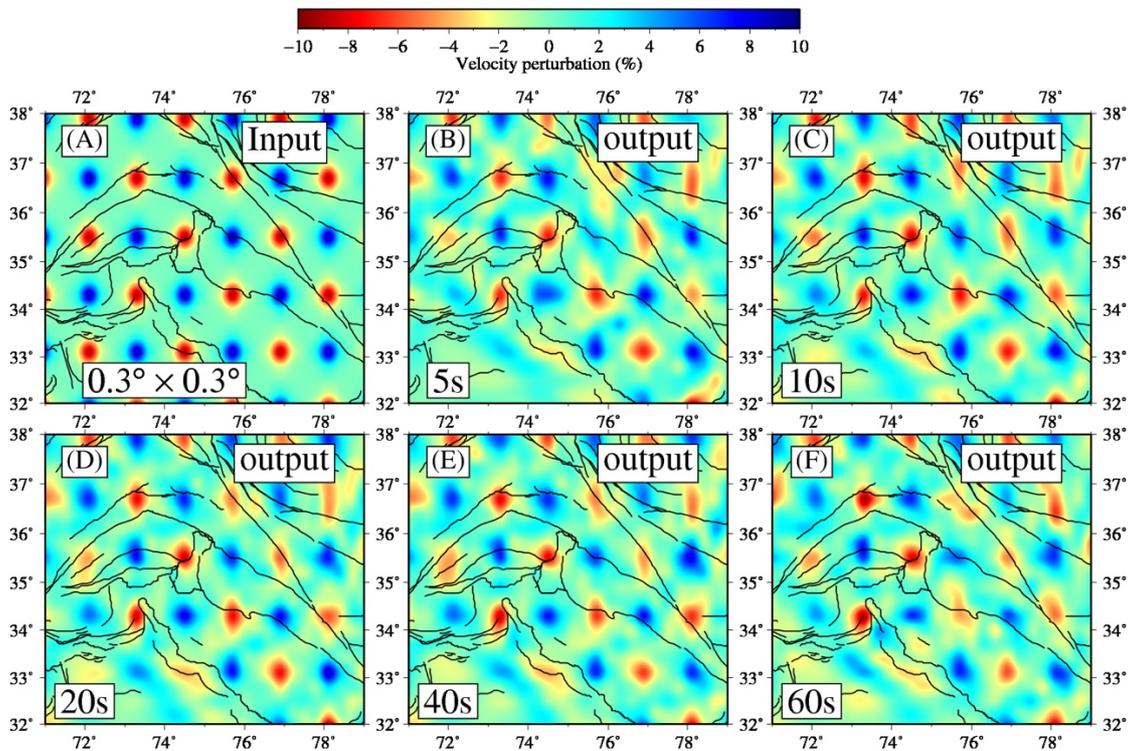
**Supplementary Figure S1.** A schematic diagram (A) shows two north-dipping subduction zones, which happened during the Early Cretaceous, resulting in the formation of the Kohistan-Ladakh arc and the Karakoram batholith (Burg, 2011). (B) The underthrusting Indian crust beneath the Asian crust and the formation of mid-crustal partial melts driven by the metamorphic dehydration reaction releasing water into the crust during the Miocene.



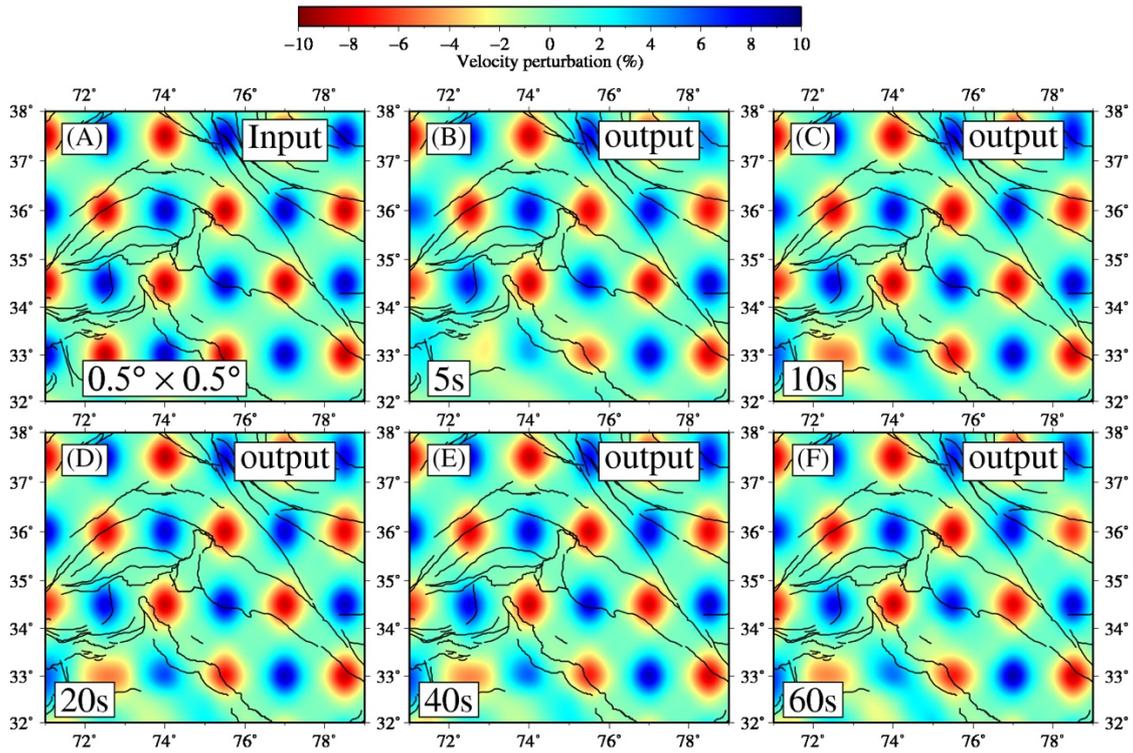
**Supplementary Figure S2.** Distribution of data used in the study. (A) Location of broadband seismic stations, shown as red inverted triangles. (B) Event locations for dispersion analysis (Depth  $\leq 30$  km; Mag  $\geq 4.5$ ). The blue box marks the study region.



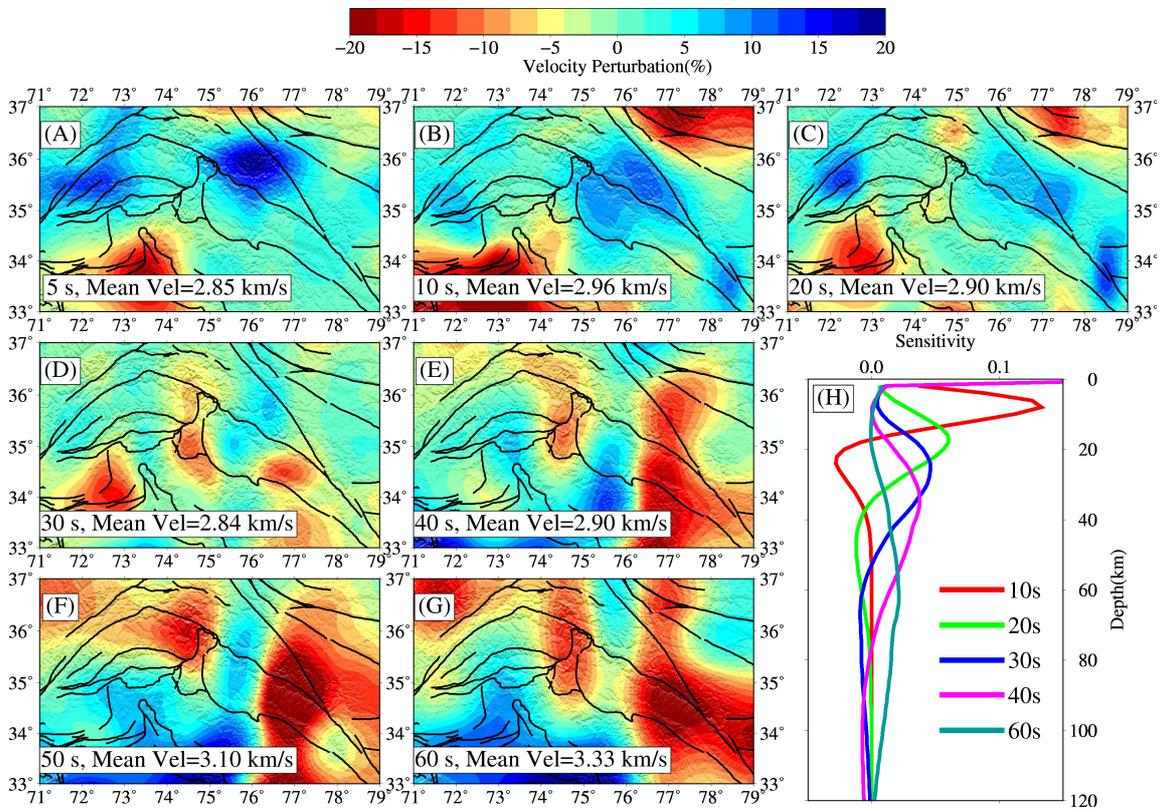
**Supplementary Figure S3.** Checkerboard test of alternating positive and negative anomalies separated by one-degree spacing between two anomalies. (A) Input checkerboard model of size  $0.2^\circ \times 0.2^\circ$ . (B-F) The corresponding output at different periods.



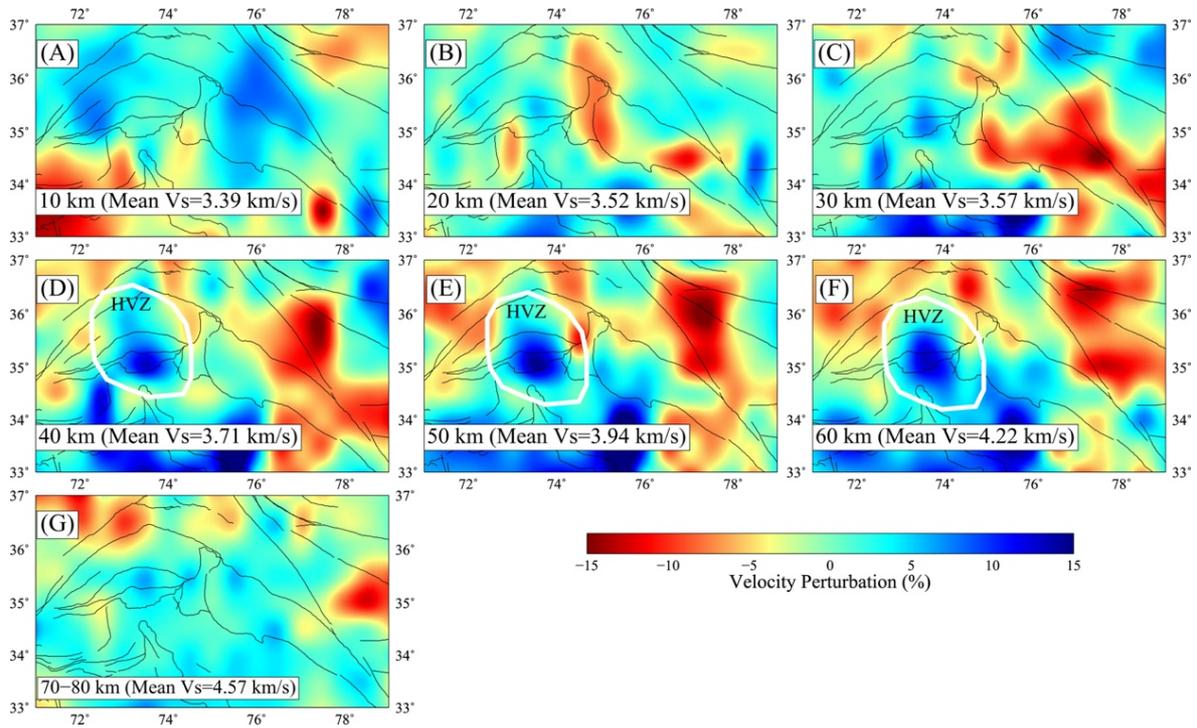
**Supplementary Figure S4.** Same as Supplementary Figure S3. Input checkerboard model of size  $0.3^\circ \times 0.3^\circ$ .



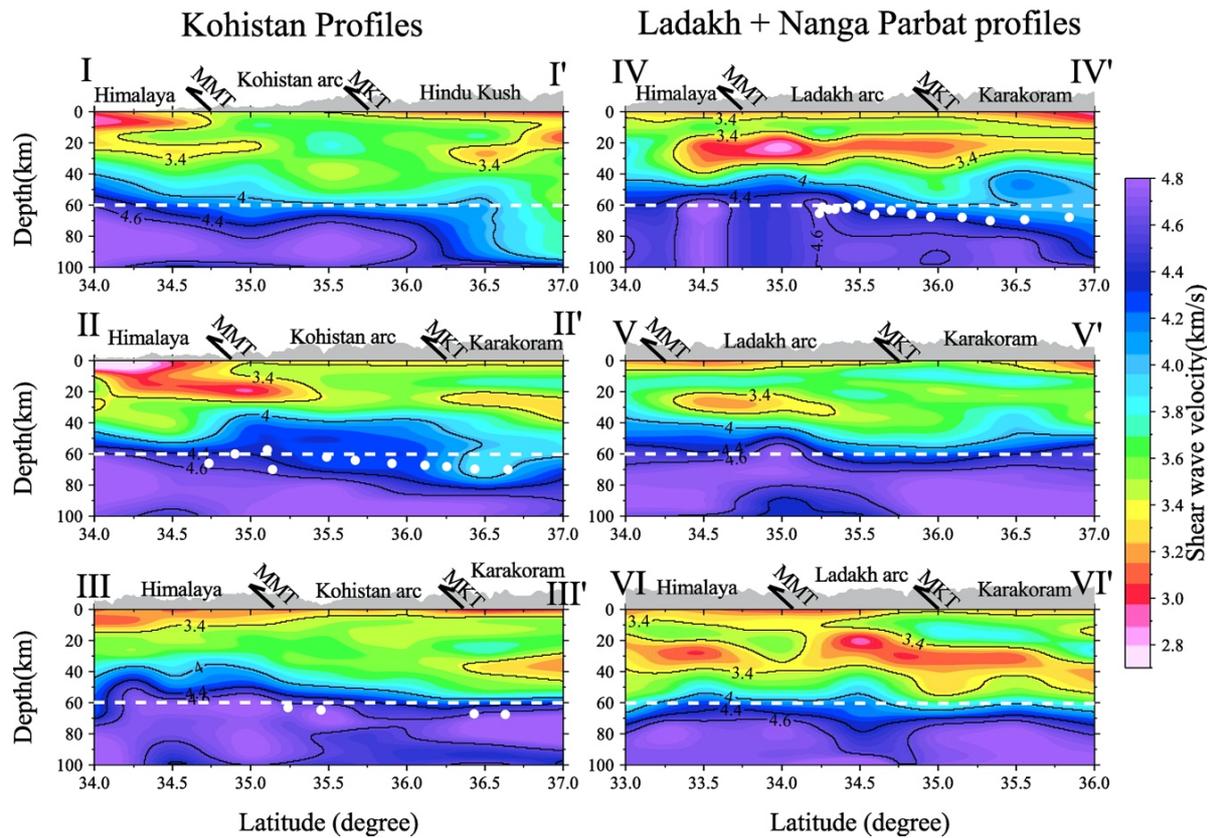
**Supplementary Figure S5.** Same as Supplementary Figure S3. Input checkerboard model of size  $0.5^\circ \times 0.5^\circ$ .



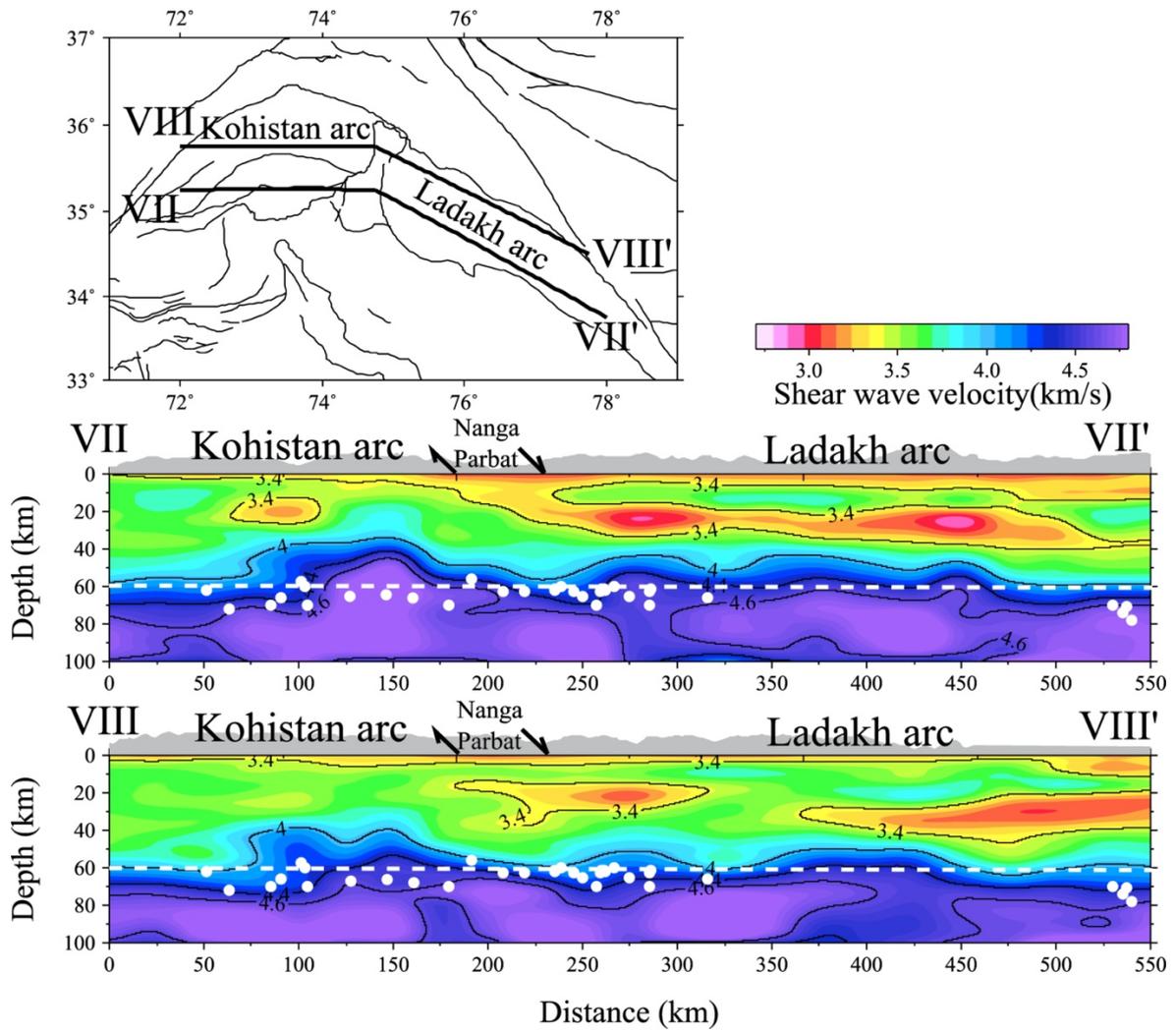
**Supplementary Figure S6.** (A-G) 2-D tomographic maps at different periods, shown as perturbations from regional mean. (H) Depth sensitivity kernel at different periods.



**Supplementary Figure S7.** (A-G) Depth slices of shear wave velocity values as perturbations from the regional mean. The white circle indicates a high-velocity lower crust observed in the Kohistan arc. The corresponding depth (in km) and the regional mean velocity ( $V_s$ ) are also indicated at the bottom of each panel.



**Supplementary Figure S8.** Velocity-depth profiles, which are shown as dashed lines in Figure 1 of the main text. Velocity contours at Vs 3.4, 4, 4.4, and 4.6 km/s are shown as black lines. Known Moho depths from previous studies are shown as white dots. A lower bound on the Moho is plotted using a dashed white line. Surface topography is shown on top of each profile, with major faults or tectonic units marked as black lines.

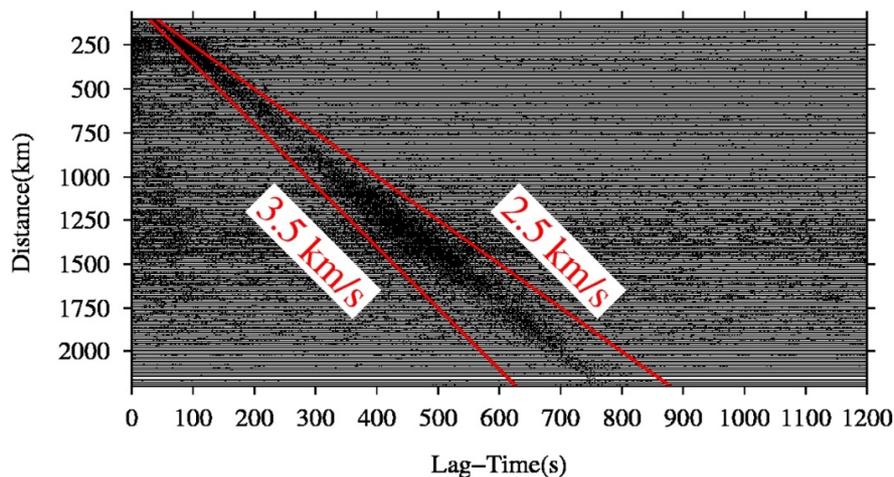


**Supplementary Figure S9.** Same as Supplementary Figure S8. The profile locations are shown at the top.

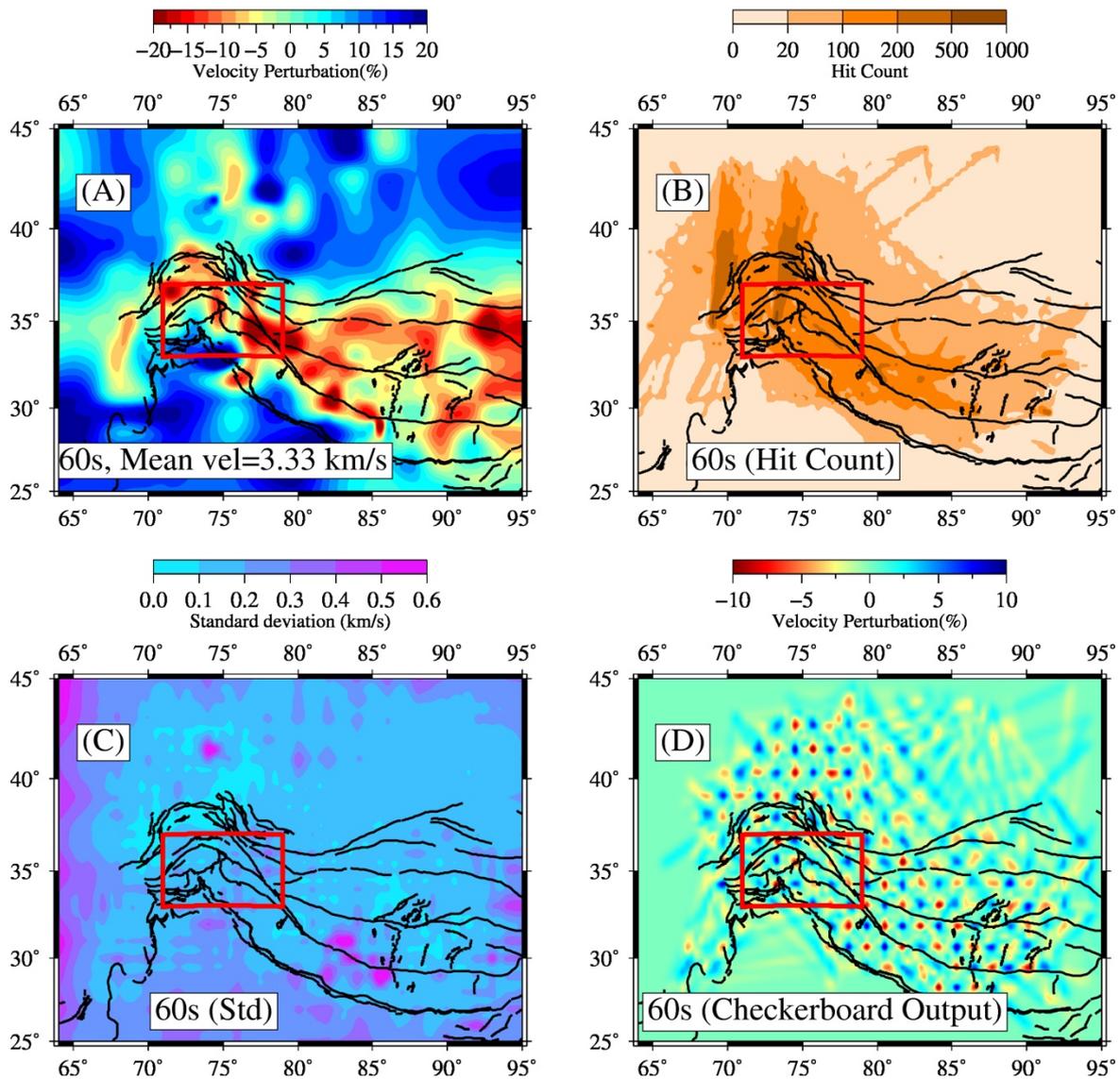
## 2 Supplementary Text S1

### Summary of data and methodology

A total of 530 broadband seismic stations, located in the Pamir, Tibet, and the Himalayas, are used for ambient noise analysis. The pre-processing follows Bensen et al. (2007) and Schimmel et al. (2011), which includes the removal of trend, mean, and instrument response from single-day-length seismic waveforms, which are band-passed at 0.01-0.25 Hz. For a given pair of stations, processed waveforms are cross-correlated using the phase coherence scheme of Schimmel et al. (2011), which improves the signal-to-noise ratio (SNR) without the explicit need for spectral and temporal normalization (e.g., Bensen et al., 2007). Finally, the daily cross-correlations are stacked together. Supplementary Figure S10 presents all the stacked cross-correlations with increasing inter-station distances. The dispersion analysis is performed using the frequency-time-analysis (FTAN), and fundamental mode group velocities for Rayleigh waves are measured. Dispersion data with  $\text{SNR} \geq 10$  and an inter-station distance greater than three times the wavelength are selected for the tomographic inversion. To improve the ray coverage, event dispersion is added for earthquakes of  $\text{mag} \geq 4.5$  and depths  $\leq 30$  km, which makes a total of 22,726 ray paths having group dispersion from period 5 to 60 seconds.



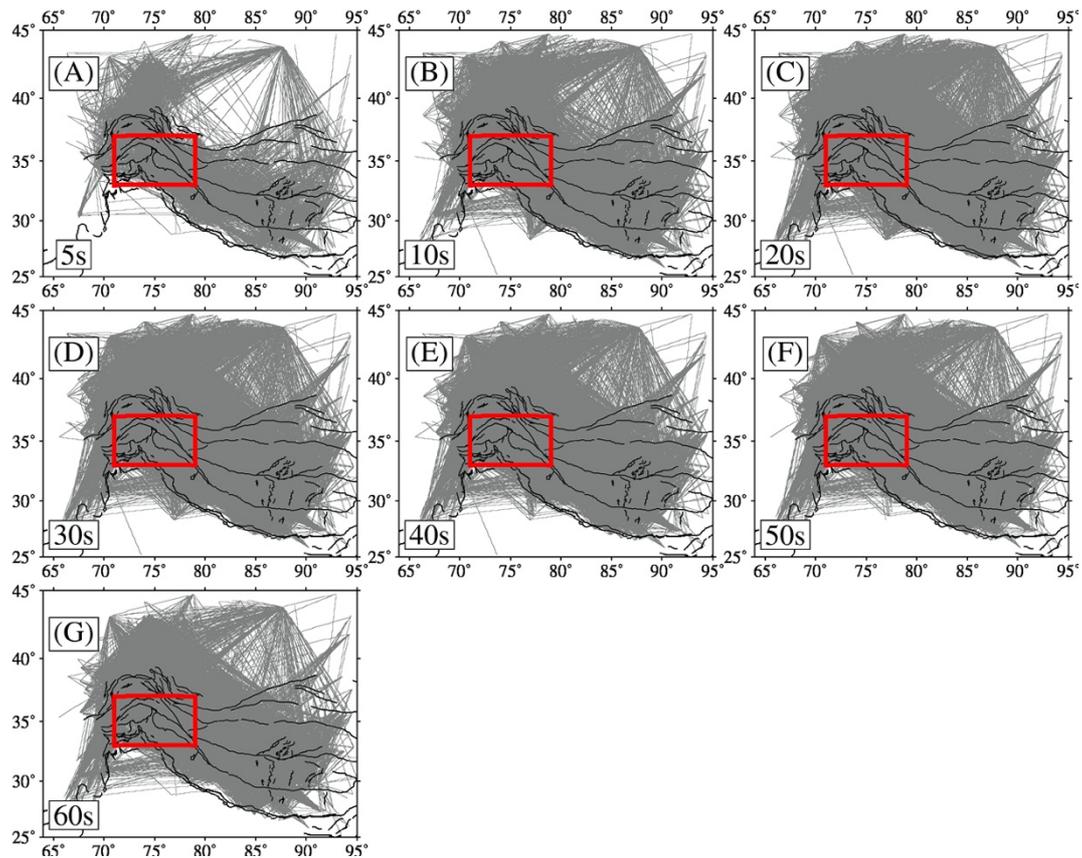
**Supplementary Figure S10.** Stacked cross-correlations are shown with increasing distance and lag-time. The red lines mark the signal window.



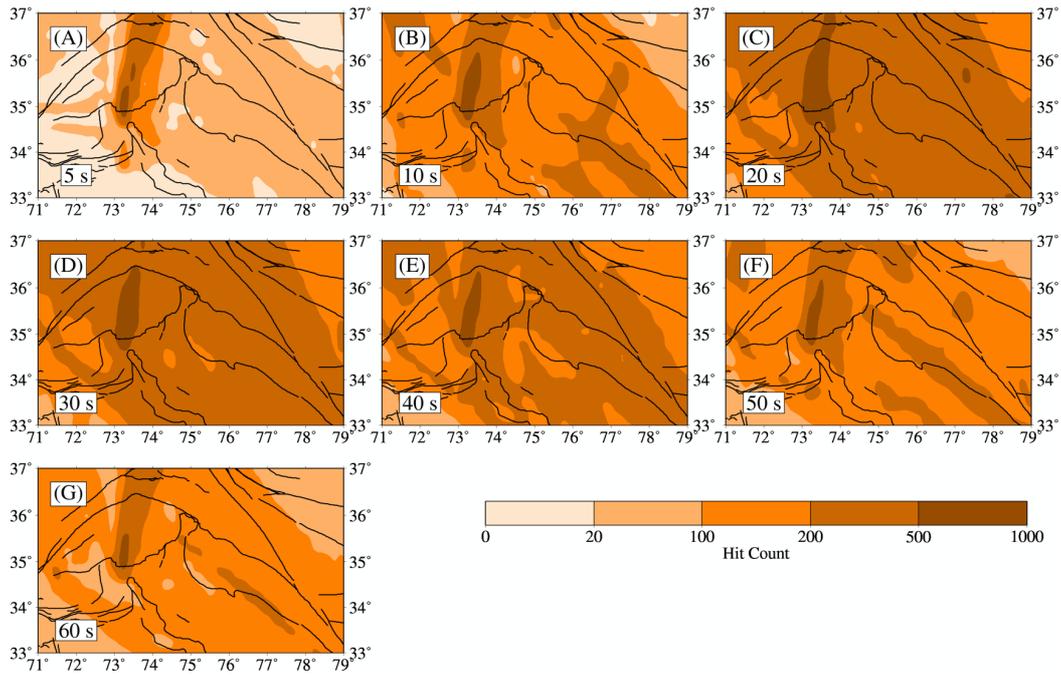
**Supplementary Figure S11.** Full grid tomography results at 60s period. (A) group velocity map as perturbations from the regional mean. (B) Ray path hit-count. (C) The standard deviation of ensemble models represents the uncertainties. (D) Full grid checkerboard result at input mode of size  $0.3^\circ \times 0.3^\circ$ , same as shown in Supplementary Figure S4.

Group velocity dispersion measurements are used for 2-D tomographic inversion using Bayesian Trans-dimensional tree tomography based on the wavelet parametrization of Hawkins and Sambridge (2015). The method represents a velocity model as a multi-resolution wavelet tree with a single pixel mean at the root of the tree and a hierarchy of wavelet coefficients from coarse to fine scales. The solution model is presented in terms of posterior probability distributions based on the Bayes theorem. For the initial model, mean group velocities at each period are used. The prior, which is the range of allowed group

velocity values, is provided in terms of a Laplacian distribution of wavelet coefficients. The final model is obtained after 1.5 million iterations, of which the first one million are discarded as “burn-in” samples. The mean and standard deviations of the last 500,000 samples are used to represent the final group velocity and its associated uncertainties. For example, the full tomography results at 60 s with checkerboard test output are shown in Supplementary Figure S11. Ray path distributions and hit-count at selected periods are presented in Supplementary Figures S12-S13.

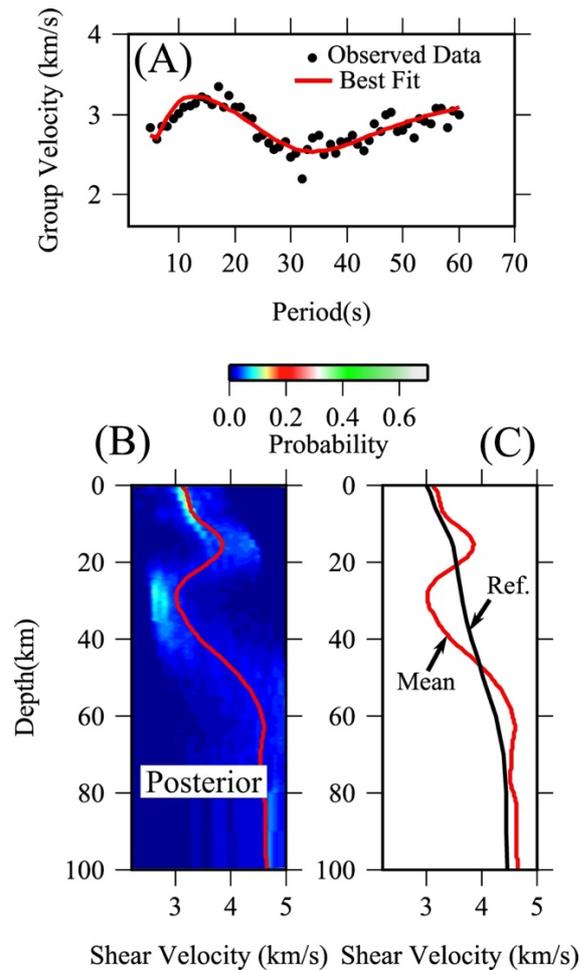


**Supplementary Figure S12.** (A-G) Ray path distributions, observed in the group velocity tomography at periods 5–60 seconds. The red box indicates the study region. Black lines are major tectonic boundaries and/or faults.



**Supplementary Figure S13.** (A-G) Ray path distribution (hit count) in the  $0.25^\circ \times 0.25^\circ$  grid-size used in the group velocity tomography.

The individual group velocities at a given grid-node in the tomographic model are further inverted to produce the 1-D shear wave velocity models ( $V_s$ ) using the Bayesian Trans-dimensional inversion of Bodin et al. (2012b). For a maximum depth of 100 km, a total of 480 parallel Markov chains are run for 100,000 (50,000 as “burn-in”) iterations to provide the final ensemble of models. The  $V_s$  model is computed by the mean of the ensemble models. An example of the 1-D inversion of data from the Ladakh region is presented in Supplementary Figure S14. The 3-D model is finally computed by interpolation of individual 1-D models, which are presented in Figure 2 of the main text.



**Supplementary Figure S14.** An example inversion of group velocity data from the Ladakh region. (A) Observed group velocity (in black dots) and best fit (red line) after the inversion. (B) A depth section showing the posterior probability distribution in the color map. Dark blue color indicates the lowest probability, while the lighter colors show increasing probabilities. The mean of the posterior, which is taken as the final  $V_s$  model, is shown in red. (C) The mean of the posterior (red line) and the reference model (black line) used in the inversion.

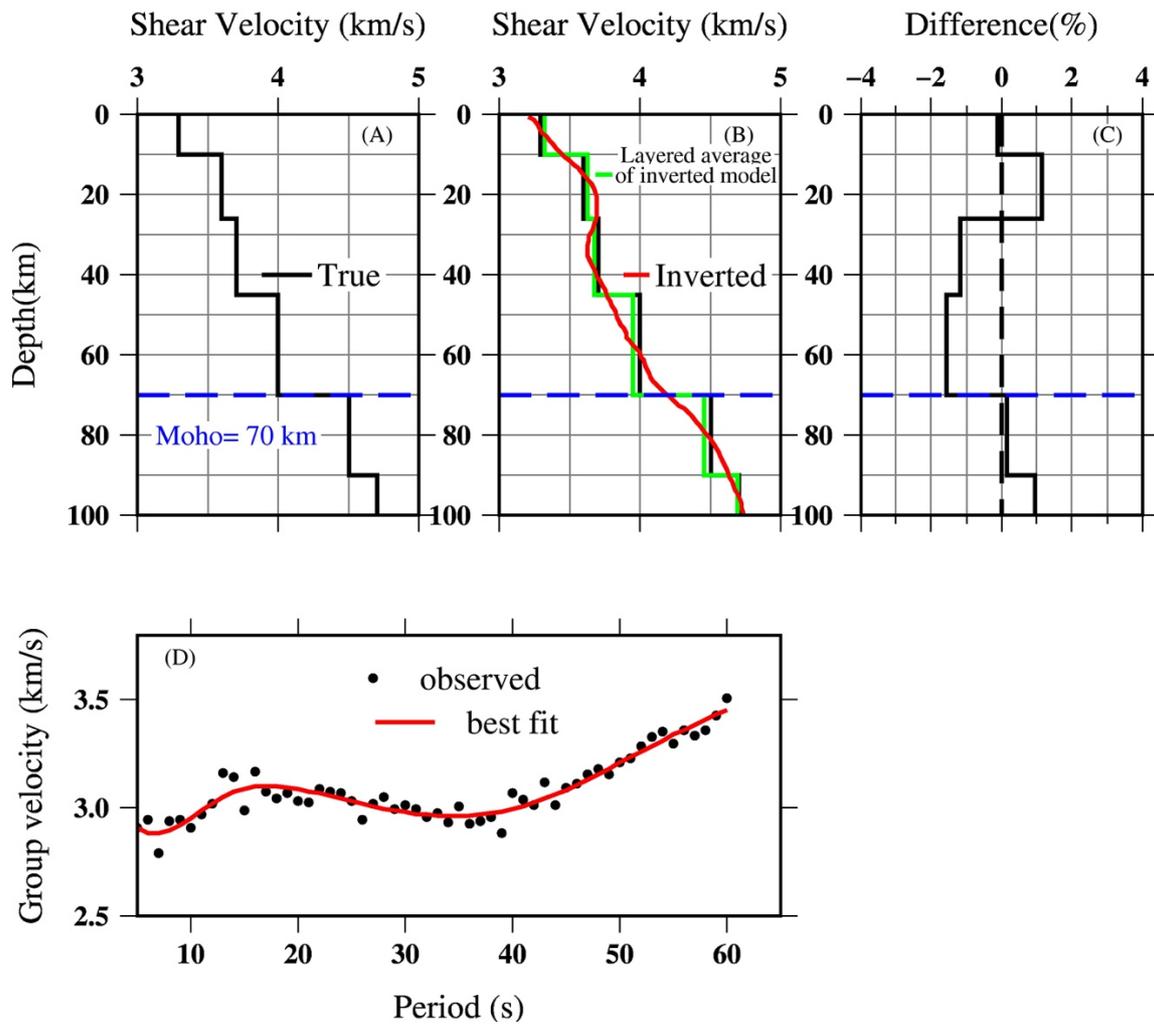
### 3 Supplementary Text S2

#### Moho depths in the Kohistan-Ladakh arc

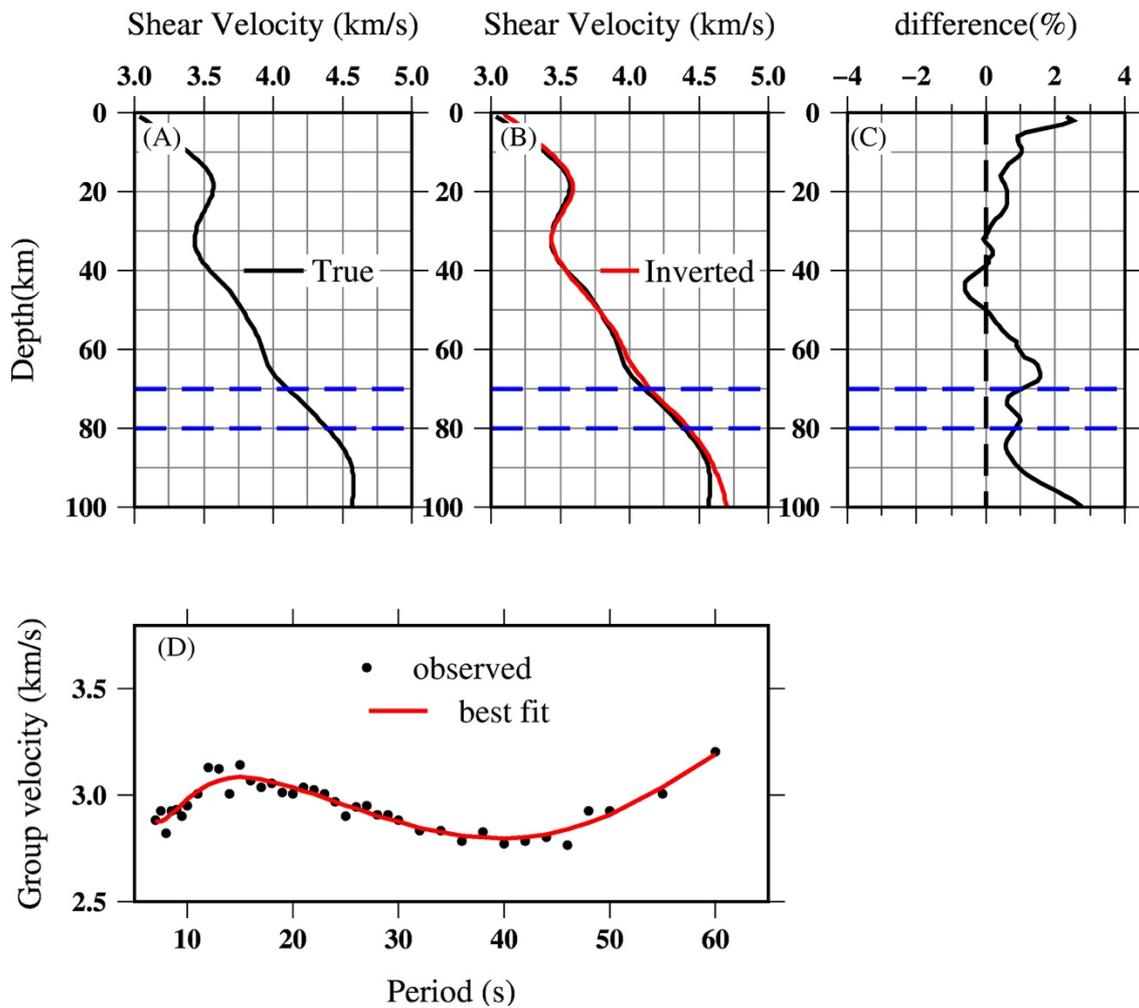
Knowledge of the variation of Moho depths in the Kohistan-Ladakh arc is important to interpret the shear wave velocity model presented in this study. Although the characteristic signature of the Moho depth can be seen in surface wave data e.g., group velocity used in this study, there exists a strong trade-off between the Moho depth and the shear wave velocity in the lower crust and upper mantle, resulting in large uncertainty in the Moho depth (Lebedev

et al., 2013). An increase or decrease in the Moho depth can be compensated, as much as 90% of the Moho signal, by an increase or decrease in the lower crust and upper mantle velocities. Because of the broad depth range of surface wave depth sensitivities (Supplementary Figure S6), the distinction between a sharp Moho and a gradational Moho becomes difficult.

To further understand this trade-off, two synthetic experiments are performed in Supplementary Figures S15-S16. The first experiment considers a sharp Moho at 70 km, and synthetic group velocities are computed in the period range of 5-60 s. A random noise (2%) is added and then inverted to recover the input velocity model. A comparison of the true and inverted models is shown in Supplementary Figure S15(B). Clearly, the inverted model is a smoothed version of the true model, and the sharp Moho signal is not well reflected. However, the average velocities of individual layers in the crust and upper mantle are reliably recovered, as the difference is less than 2 % (Supplementary Figures S15C). Note the relatively large uncertainty in the lower crust (above 70 km), which may indicate the result of velocity trade-off above and below the Moho. In the next experiment, an input model having smoothly varying velocities with depth is considered, and the same procedure is repeated, as discussed above, to calculate the inverted model. Although it is difficult to assign a single Moho depth, this experiment assumes that the Moho transition takes place at  $V_s$  of 4-4.4 km/s. The inverted model recovers the true velocity pattern but suffers from relatively large uncertainties (2-3 %) at the upper crustal depths (< 10 km), the lower crustal depths (50-70 km) and upper mantle depths (80-100 km). The large errors at the uppermost crust may be due to a lack of short-period data (< 5 s) in the dispersion. At the lower crust and upper mantle, the large uncertainties may indicate the crust-mantle trade-off (< 3 %), as discussed earlier, and decreasing sensitivity of dispersion data at depths beyond 80 km (Supplementary Figure S6). These experiments suggest that the inversion with surface wave data may only recover the average velocity values, but computing a single Moho depth is difficult.



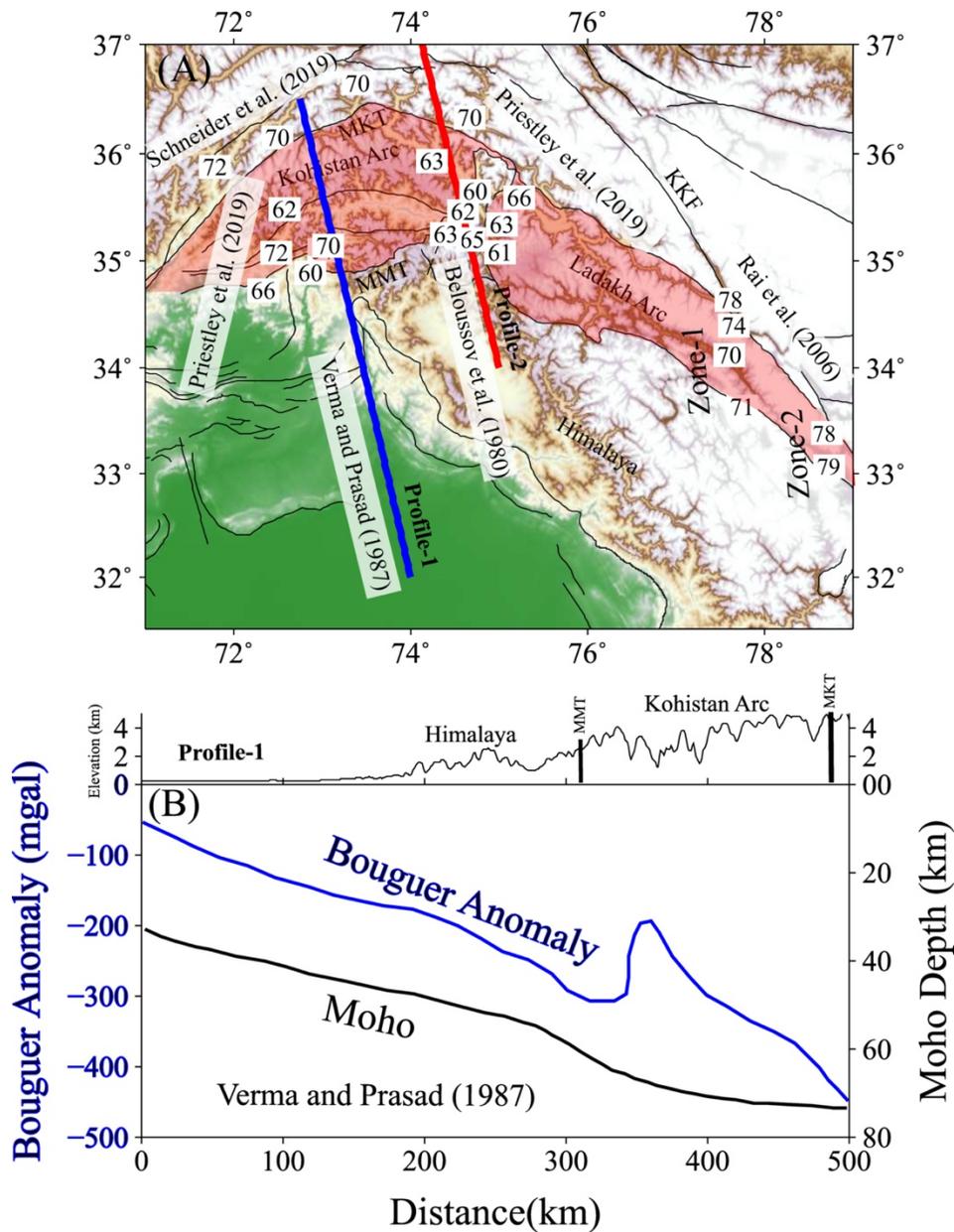
**Supplementary Figure S15.** Synthetic test for a sharp Moho. (A) An input shear wave velocity model with a sharp Moho transition at 70 km. (B) Comparison of the inverted model (red line) with the true model (black line). The green line is the layered average model of the inverted model. (C) The difference between the true model and the average inverted model is presented in terms of percentage (%). (D) Observed and best fit dispersion.



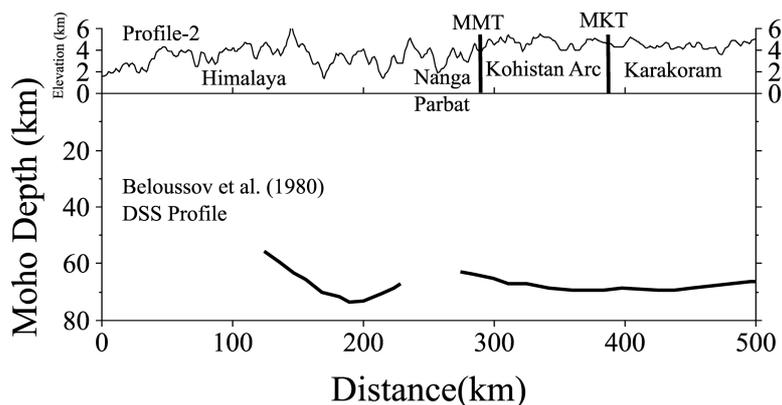
**Supplementary Figure S16.** Same as Supplementary Figure S15. The input model contains a gradational Moho transition at 4-4.4 km/s (blue dashed line).

Calvert (2011) studied the seismic characteristics of island arcs and showed that the sharply defined crust-mantle transitions are rarely observed or absent in island arcs. Similar observations, i.e., a gradational Moho, are also made in the Himalayas (Gilligan et al., 2015). Furthermore, the role of increasing thermal gradient on the shear wave velocity needs to be carefully investigated while computing the Moho depth. Diaferia and Cammarano (2017) investigated the effect of thermal gradient on the shear wave velocity model by computing dispersion and receiver functions considering a low thermal gradient (10 K/km), and a high thermal gradient (40 K/km). Their results showed that a shear wave velocity of 4.5 km/s corresponds to the Moho depth at low thermal gradient (10 K/km) but decreases to around 4 km/s at high thermal gradients (40 K/km). Due to the absence of a detailed knowledge of geothermal gradient in the Ladakh-Kohistan region, a lack of Receiver function studies in the interior of Ladakh and Kohistan, and the crust-mantle trade-off in the surface wave inversion,

as discussed earlier, this study relies on the lower and upper bound on the Moho depths in the Kohistan-Ladakh arc rather than a single Moho depth.



**Supplementary Figure S17.** Compilation of previous studies providing Moho depths. (A) A map showing the spatial distribution of local Moho depth estimates inside a white box based on previously published results (Verma and Prasad, 1987; Belousov et al., 1980; Rai et al., 2006; Schneider et al., 2019; Priestley et al., 2019). The Ladakh-Kohistan arc is shaded with pink color. (B) Variation of Bouguer anomaly (blue) and computed Moho depths along the profile-1 (blue line) of Verma and Prasad (1987).



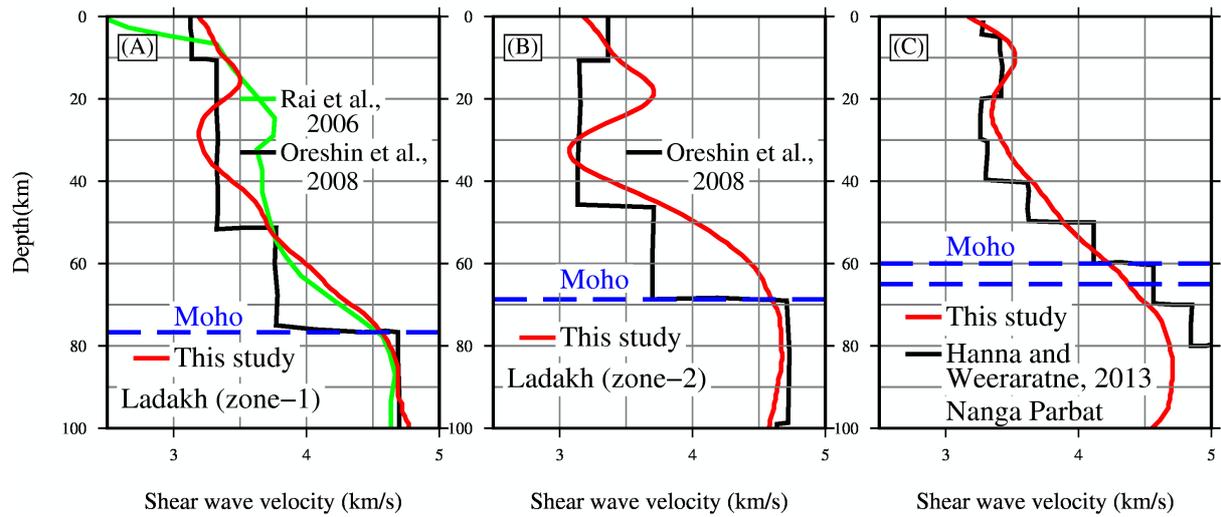
**Supplementary Figure S18.** Moho depth computed from the DSS study by Belousov et al. (1980). For the location of the profile, see profile-2 (red line) of Supplementary Figure S17.

To investigate the range of observed Moho depths in the Kohistan-Ladakh region, Supplementary Figures S17-S18 present a compilation of previous studies providing local Moho depth estimates. Priestley et al. (2019) reviewed the crustal structure of the Himalayas and surrounding regions and compiled Moho depths which are primarily derived from receiver functions and/or their joint inversion with surface waves. The Moho depths in the Nanga Parbat and southern Kohistan are from Priestley et al. (2019). In the easternmost part of the Ladakh arc, Rai et al. (2006) provided Moho depths using joint inversion of surface waves and receiver functions. A deep seismic sounding (DSS) profile from the Kashmir Basin in the south to the Pamir range in the north (red line in Supplementary Figure S17A) was provided by Belousov et al. (1980). Another linear profile used gravity modelling (Verma and Prasad, 1987) and provided Moho estimates in the western Himalaya and Kohistan region (blue line in Supplementary Figure S17A). In the Karakoram region, north of the Kohistan arc, Schneider et al. (2019) estimated Moho depths from receiver function modelling. These studies provide first-order estimates of the Moho depths in the Kohistan-Ladakh region. In the Kohistan arc, there are relatively good samples of the Moho estimates compared to the Ladakh arc. The Moho depths range from 60 km to 72 km in the Kohistan region. In the DSS profile, the Moho depth is 60 km in the Nanga Parbat area and increases to 70 km just north of the Kohistan arc (Supplementary Figure S18). Similarly, the Moho depth, computed from the gravity data, starts at 60 km below the MMT, the southern boundary of the Kohistan arc, and increases to 70 km in northern Kohistan (Supplementary Figure S17B). In eastern Ladakh, the Moho depths estimates can go up to 80 km (Rai et al., 2006). From these published values, it is fair to say that the Moho lies in the depth range of 60-72 km in the most part of the Kohistan-Ladakh arc, except in eastern Ladakh where a

deeper Moho (up to 80 km) can be expected. Note that the eastern end of the Ladakh arc represents a transition from the island arc system in the west and the Andean-type southern Tibet in the east, possibly separated by the Karakoram Fault (Burg, 2006), which may be the cause for the increased Moho depths.

Supplementary Figure S19 presents a comparison of this study's velocity model to published shear wave velocity models from the Ladakh and Nanga Parbat regions. This comparison helps to assess the reliability of this study's model and also provides first-order information on the expected range of shear wave velocities around the Moho in the Kohistan-Ladakh region. In the Ladakh region, two velocity models from eastern Ladakh are taken from Rai et al. (2006) and Oreshin et al. (2008), where velocity models are computed by receiver functions modelling and their joint inversion with surface waves. There is a general consistency between the velocity model of the crust and upper mantle from the present study and the published models. Some discrepancies in the crustal velocities may indicate differences in the data, methodology, and spatial resolution of the derived model. The upper mantle velocities in these studies are  $\sim 4.6$  km/s, which are also recovered in this study's model. In the Nanga Parbat region, Hanna and Weeraratne (2013) computed a shear wave velocity model from teleseismic surface waves. They didn't provide a Moho depth, however, the superposition of observed Moho depths from Priestley et al. (2019) shows that a  $V_s > 4.4$  km/s corresponds to the base of Moho in the Nanga Parbat region. Although the comparison is done at a selected few locations of the Ladakh and Nanga Parbat area due to a lack of published models, a reference value of  $V_s \sim 4.4-4.6$  km/s can be regarded as indicative of the base of the Moho for this study. The depth range of the  $V_s \sim 4.4-4.6$  km/s has previously been used to mark the Moho in Ladakh and the Himalayas (Rai et al., 2006; Hazarik et al., 2017). In the velocity profiles, presented in Figure 2 of the main text and Supplementary Figures S8-S9, the contours of  $V_s \sim 4.4-4.6$  km/s lie in the depth range of 60-70 km for the Kohistan and Ladakh region, except in the eastern Ladakh, this depth range is around 70-80 km, which is consistent with the inferred range of the Moho depths, discussed in the previous paragraph.

In summary, for this study, a lower and upper bound of 60 and 72 km, respectively, on the Moho depth can be taken for the whole Kohistan-Ladakh region. The upper bound can reach 80 km at the eastern end of the Ladakh arc.



**Supplementary Figure S19.** Comparison of this study's shear wave velocity model with published velocity models (Rai et al., 2006; Oreshin et al., 2008; Hanna & Weeraratne, 2013). The red line indicates the velocity model of the present study. The zone-1 and zone-2 are marked in Supplementary Figure S17(A).