REVIEWER #1

I appreciate the authors addressing my concerns, particularly for trade-offs between Vs and thickness in the inversion. However, there are still unclear points that should be clarified. I hope that the following comments contribute to improve the manuscript.

L111
There are two “similar” in this sentence, and please remove either.
Reply: Thank you for pointing this out. We have revised the sentence accordingly and removed the redundant use of “similar” to improve clarity.

L114–119
Thanks for adding explanations on the azimuthal corrections and processing. However, I do not still understand what the authors did on the clock corrections. It seems that Yang et al. (2023; 2024) were written in Chinese, and I do not understand whether these papers are related to clock deviations for the OBSs used in this study. To avoid misleading readers, what the authors did step-by-step should be more explained.
“For example, in the case of OBS stations…”
This sentence explains the preprocessing to the continuous records, and Fig. S1 maybe shows clock corrections using the cross-correlation functions. The sentence and citation are inconsistent. More explanations on Fig. S1 should be needed. Also, it is necessary to define “a sliding time-window normalization” in details.

Reply: We appreciate the reviewer’s careful reading and constructive suggestion. We have now substantially expanded the description of the OBS clock-drift correction in the revised manuscript (Data and Method section) to provide a clear step-by-step explanation.

In general, there are two factor affect the clock drift since OBS internal clocks are based on quartz crystal oscillators whose actual vibrating frequencies is not always the nominal frequency as it claimed. It is sensitive to the large pressure and temperature changes at the seafloor. As a result, the clock frequency deviates from the nominal value, leading to cumulative timing errors of up to several seconds over a year-long deployment. Specifically, we use the real-time clock signal $\mathbf{PClk_0}$ recorded by the instrument to obtain the actual oscillator frequency after it has been affected by in-situ conditions. This measured frequency is directly related to the true time stamp of the OBS data. Using Eq. (S1), we integrate the deviation between the nominal and measured frequencies over the recording period to estimate the cumulative oscillators vibration frequency error and correct the sample rate accordingly.

Then it is the second part, clock drift due to oscillators ageing. We use our OBS data from multiple-components to get noise cross-correlation function between both land station TOLI2 and OBSs so we can estimate and correct the drift caused by the quartz oscillator aging. The workflow nowadays is a common practice which is similar to Hable et al. (2018).

By computing cross-correlation functions between station pairs and analyzing asymmetries in causal versus acausal parts, they inferred relative timing offsets. These pairwise measurements were combined in a least-squares inversion to estimate individual station clock drifts, typically modeled as linear or spline over the deployment period. Observed drifts could reach several seconds per year, but the method achieved sub-millisecond precision (~0.2–0.5 ms) in the estimated corrections. Applying these corrections aligns OBS and land station data, significantly improving the accuracy of seismic event locations and waveform-based studies.

The cross-correlation functions shown in Fig. S1 are intended to verify and validate the effectiveness of our time correction. For two OBS stations without relative clock drift, the causal and acausal phases of their noise cross-correlation functions remain stable and symmetric in time. When relative clock errors exist, the positive and negative branches of the cross-correlation function shift asymmetrically with time. Therefore, comparing the cross-correlation functions before and after time drift correction provides an independent means to confirm whether the timing drift has been adequately removed. We have added this explanation to the supporting text.

Last, we remove the wording about the procedure about cross correlations. Such as “sliding time-window normalization” in the preprocessing section. But I will put some explaination here for our reviewer. Specifically, it refers to the normalization of continuous seismic records within short overlapping time windows to stabilize noise amplitude variations prior to computing the cross-correlation functions. This step is to suppress transient, high-amplitude events—such as earthquakes, instrument glitches, or local noise bursts—before computing noise cross-correlation functions so that it ensures that the resulting cross correlations reflect the stable background noise field

In this procedure, the continuous waveform is divided into short, overlapping time windows of fixed length (typically tens of seconds). In our case, we used is 120 s which is default in MSNoise package (http://www.msnoise.org/).

For each window centered at time t, the absolute amplitudes of the seismic trace $|d_j|$ within the window are averaged to obtain a local normalization factor:

$$w_n = \frac{1}{2N+1}\sum_{j=n-N}^{n+N} |d_j|.$$

The raw waveform value d_n at time n is then normalized as:

$$\hat{d}_n = \frac{d_n}{w_n}.$$

This process adaptively scales the data, reducing the influence of large-amplitude, short-duration signals while preserving the relative phases of the ambient noise field. Compared with one-bit normalization, the sliding time-window method retains more amplitude information and is less likely to distort the ambient-noise wavefield. It also effectively suppresses earthquake signals because their amplitudes exceed the window-based threshold and are consequently down-weighted.

L122

whose -> with

Reply: Thank you for your suggestion. We have revised the wording accordingly and replaced "whose" with "with" to improve grammatical accuracy and readability.

L257–265

I thought that the authors estimated Vp/Vs and crustal thickness that are averaged over a top sediment layer and underlying igneous crust, so in my previous comment, I requested that this point is explicitly stated. The authors stated in the rebuttal letter that "We have explicitly stated that the estimated crustal thickness includes both the marine sediment and the igneous rock layers above the Moho, and that the reported Vp/Vs values represent averages over these two layers in the revised manuscript at Lines 238-240". However, according to these lines, the authors estimated thicknesses of overlying sediment layers (L262). I am very confusing. It would be really helpful if the authors explain how sediment thicknesses were estimated in the H-kappa method, and show the resulting values in a table. Also, are those values consistent with the Vs profiles estimated from the inversion?

Reply: We appreciate the reviewer's thoughtful follow-up and the opportunity to clarify our methodology. First, among all stations, C21F, located in the central Celebes Sea, is situated on young oceanic crust with 0.8 sediment cover (and our forward modelling, as supported by regional seismic studies (Kopper et al. 1999 DOI: 10.1016/S0264-3707(98)00004-0). We therefore adopt the crustal thickness beneath C21F (7.83 km) minus the sediment 0.8 as the reference igneous basement thickness (7.0 km) for this region. Under the assumption that adjacent stations located along the same magnetic

lineation formed during a similar spreading episode share comparable igneous basement thickness, we estimate the sediment thickness at each station by subtracting this reference value (7.0 km) from the total crustal thickness obtained from the H-κ stacking.

We would also like to clarify that the sediment thickness estimates are not presented as standalone quantitative results in this study. Instead, they are used only as an internal diagnostic tool to verify the internal consistency of the H-κ-derived crustal thickness and the Vp/Vs values across stations. Because our focus is on the overall crustal structure rather than the detailed sedimentary cover, we do not attempt to derive or interpret sediment thicknesses in detail. The simple subtraction approach (using C21F as the reference crustal thickness) is therefore intended solely as an approximate method to assess whether the spatial variations in Vp/Vs and H-κ results behave coherently.

Eq. S2 in Text S2

Please define the weight, w, and E_T. What time window and frequency range did the authors take for E_T? Does i correspond to i-th event for a total of N? After the summation of E_T with weights, why is E_T estimated? Is this a recurrence formula? I do not understand this equation at all.

Reply: Thank you for the helpful comments. We have revised Eq. (S2) and clarified all related definitions in Text S2. The updated expression is:

$$\boldsymbol{E(\theta_a) = \frac{\sum_{i=1}^{N} w_i E_T^i(\theta_a)}{\sum_{i=1}^{N} w_i}}.$$

Here, $\boldsymbol{i}$ denotes the i-th earthquake event among a total of N events used for energy stacking. $\boldsymbol{E_T^i(\theta_a)}$represents the transverse-component energy of the i-th event as a function of the back-azimuth $\boldsymbol{\theta_a}$; it is measured within a fixed time window around the S-wave arrival and within the frequency band specified in the revised text. The weight $\boldsymbol{w_i}$is defined as the inverse noise level (or an SNR-based weight) of the i-th event so that higher-quality events contribute more strongly to the stacked energy. To compute each $\boldsymbol{E_T^i}$, we selected events with magnitudes greater than 5 and epicentral distances between 5° and 90° from the continuous waveform records of OBS M01G and applied a 0.02–0.5 Hz band-pass filter to the extracted seismograms. The purpose of Eq. (S2) is to compute the weighted mean transverse energy for a given back-azimuth $\boldsymbol{\theta_a}$, and the summation does not re-estimate $\boldsymbol{E_T}$ but simply forms a stable weighted average of the individual-event energies. We also clarify that Eq. (S2) is not a recurrence formula; it contains no recursive dependence on previous values and is purely a weighted-average stacking equation. All of these explanations have been added to Text S2 in the revised manuscript. These

clarifications have now been added to Text S2 in the Supporting.

Text S2

Thanks for adding descriptions on time corrections. However, as mentioned above, I do not understand what the authors indeed did for clock corrections. Step-by-step explanations on the clock corrections conducted in this study are needed, particularly how to estimate Time_Err (s) in Table 1. In the current manuscript, the authors only cite Hable et al. (2018).
Also, please remove the sentence of "This information has been added…".

Reply: Thank you for the comment. We understand the reviewer's concern regarding the step-by-step procedure of the clock-drift correction. As stated in our initial response, the complete correction workflow—including how the cumulative timing error Time_Err (s) reported in Table 1 is derived—is now described in detail in the supporting document Text S2: Time Correction.

In this study, clock correction is first achieved by determining the in-situ oscillator frequency of each OBS from the recorded PClk_0 signal and integrating its deviation from the nominal sampling frequency over the entire deployment to obtain the cumulative timing error. In addition to this instrument-based calibration, we further applied a cross-correlation approach using coherent seismic phases to refine residual drift, ensuring that both absolute clock offset and time-dependent drift are fully corrected.

Moreover, because this study relies primarily on receiver functions—which are sensitive to relative rather than absolute arrival times—the impact of residual clock drift on our key results is inherently limited. Even if a small uncompensated drift were present, it would not affect the differential timing that controls receiver-function waveforms and subsequent structural interpretations.

Text S3

Thanks for adding descriptions on Voronoi cell. However, I would like to read what the Voronoi cell correspond to in this study. Does the Voronoi cell correspond to layer thicknesses or shear-wave velocity in this study? How is this parameter linked to physical quantities used in this study?

Reply: Thank you for the question. In this study, Voronoi nuclei are used to jointly control the geometry and the shear-wave velocity of the 1-D layered structure. Their specific roles are as follows.

(1) Layer geometry determined by nuclei: Each Voronoi nucleus represents a control point in depth. The layer boundaries are placed at the midpoints between adjacent nuclei, so

the depth intervals of the Voronoi cells directly define the layer thicknesses. When two nuclei move closer or farther apart, the thickness of the corresponding layers adapts accordingly.

(2) Shear-wave velocity within each layer: Each nucleus carries a shear-wave velocity parameter v_i, and this value is uniformly assigned to its entire Voronoi cell. Thus, a Voronoi cell defines a layer whose velocity is constant and equal to the value associated with its nucleus.

In the rebuttal letter, the authors mention that "These differences are mainly attributed to the limited number and quality of waveforms recorded by OBSs, as well as the inherent trade-offs between Vs and layer thickness when dealing with complex crustal structures." and "We believe that the integration of both methods offers a balanced and reliable characterization of the crustal structure." These should be stated in the text. As I stated in my previous comment, it seems that readers do not understand which results the authors think more reliable. The above points would help readers understanding the authors' intentions.

Reply: Thank you for the insightful comment. We agree that readers should clearly understand our assessment of the reliability of the different results. In the revised manuscript, we now state explicitly that the S-wave velocity structures inverted from receiver functions carry certain uncertainties, which is expected due to the limited number and quality of OBS waveforms and the trade-offs between Vs and layer thickness in complex crustal settings. In this study, however, we do not interpret the absolute values of the Vs profiles; instead, we focus on the relative low-velocity anomalies that are consistently resolved across stations and are robust to different inversion parameterizations. By contrast, the crustal thickness estimates obtained from both the H-k stacking and waveform fitting are more reliable, as demonstrated by the stable H-k peaks and the good agreement between observed and synthetic receiver functions. We have incorporated these clarifications into the main text so that readers can better understand which aspects of the results we consider most robust and how they should be interpreted.

REVIEWER #2

General comments:

The revised manuscript is considerably improved -- a lot of missing details have been filled in. With the additional details, I remain confident that the results are robust and interesting, and that the authors are focusing on the most reliable aspects of their measurements. Some minor presentational issues remain, as detailed below -- nothing that a minor revision wouldn't fix. I've annotated the manuscript and supplement with a number of minor grammatical edits (some of the newly added text needs a bit of work). The main points are below, but please consider the annotations on the PDF files as well.

Specifics:

1. Figure 1: as noted last time, the caption still refers to three microblocks, but only two are listed.

Reply: Thank you for pointing this out. The caption of Figure 1 has been revised accordingly, and the incorrect reference to three microblocks has been corrected to list only the two microblocks shown in the figure.

2. Lines 93-95: This sentence is ungrammatical and I'm not 100% certain what was intended. Rephrase.

Reply: Thank you for pointing this out. We have remove the sentence accordingly. As the review suggested, the sentence is not necessary.

3. Figure 3: I think my comment from last time may have been misunderstood. The best-fit (H,kappa) pair predicts arrival times for the phases used in the stack -- it would be helpful to indicate those on the RF traces, in order for the reader to see which arrivals are the ones used for the (H,kappa) estimate.

Reply: Thank you for the suggestion. We use grey bar to show the predicted arrival times of the Ps, PpPs, and PsPs + PpSs phases corresponding to the best-fit (H, κ) pair on the receiver-function traces. This modification has been made in Figure 3 in the revised manuscript.

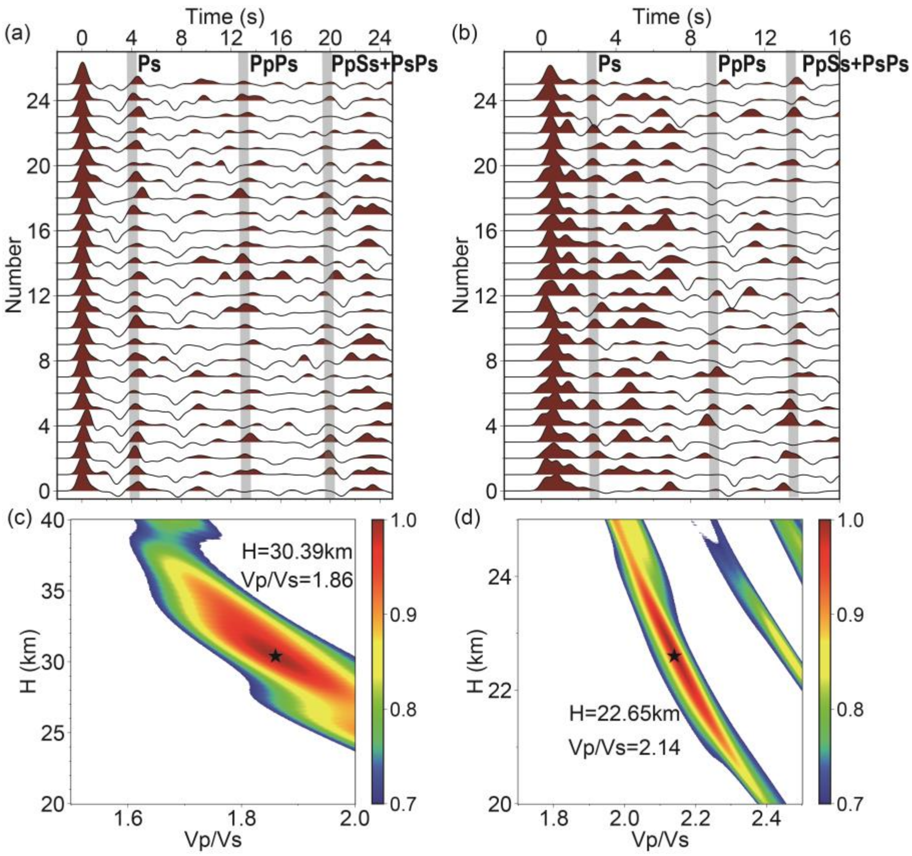

Figure 3: (a) Receiver functions at the land station TOLI2; (b) Receiver functions at the OBS site M03F; (c) H-κ stacking result of TOLI2; (d) H-κ stacking result of M03F.

4. Line 164 refers to "crustal" thickness when I think basement thickness is meant.

Reply: Thank you for the comment. In this study, it should be "basement thickness" instead, we have modified accordingly.

5. Lines 164-167 suggest that thicknesses for sediments were calculated, but they aren't shown anywhere. Could they be added to Table 2?

Reply: Thank you for the suggestion. In this study, the sediment thicknesses inferred from comparing each station's H–κ result with the reference total crustal thickness from the C21F's RF result, then the basement thickness (7.8 km) was used as reference for the oceanic crust thickness for the (C28F, C12F, C18F, C09G, C08F). We have included the sediment thickness in Table 2 in the revised manuscript.

6. Line 205: The authors never answered my previous query about whether a moveout was applied before stacking the RFs. It's not a huge problem if this wasn't done, but in that case the authors should specify "stacked without moveout".

Reply: Thank you for pointing this out. We have now clarified this in the manuscript by stating

explicitly that the RFs were “stacked without moveout.”

7. Line 292: "The presence of such a feature" -- please specify which feature -- is it the sharp lateral change from C28F to C08F, or something else?

Reply: Thank you for the comment. We have revised the text to specify the feature being referred to. In this context, “such a feature” denotes the pronounced lateral change in crustal thickness between stations C28F and C08F. This sharp mid- to lower-crustal contrast supports our interpretation that the Palu–Koro fault cuts through the entire crust and offsets the Moho discontinuity. The wording in the manuscript has been updated accordingly.

8. Supplement: it would be best for the text descriptions of processing steps to be in the same order as the supplemental figures (the figures show time, then azimuth, while the text reverses this).

Reply: Thank you for the suggestion. We have adjusted the order of the descriptions in Text S1 and Text S2 so that the processing steps now follow the same sequence as the supplemental figures, ensuring consistency between the text and the figures.