Response to reviewers

Dear Reviewers,
Dear Topical Editor and Editor-in-Chief CharLotte Krawczyk,

We sincerely thank both reviewers for their fair and constructive reviews, which greatly improved our manuscript. We appreciate the feedback given on the manuscript and carefully incorporated all points risen. Please find below our answers for each comment in green coloured text.

Kind regards,
Yueyang Xia on behalf of all co-authors

Comment: During the review process we first corrected the sticky notes from the PDF-Version and then started the main correction and reorganised the manuscript.

We realized that the calculation of the depth errors from the NRM displacement field needed more explanation (both reviewers).
To avoid an incomplete description, we are explaining the application by mathematic formulas, illustrated the additional individual steps, and added an online repository with the programming scripts and data examples.

New Chapters:
2.1.1 NRM synthetic data example
2.1.2 Depth variant alignment from relative displacement correction
2.1.3 RMO automatic picking by tracking through NRM displacement field
2.1.4 Effective RMO selection based on semblance analysis

This part increased the length of the manuscript significantly.

Referee 1: Nathan Bangs

1. The description of the technique is reasonably well presented and illustrated, but there are two things that are not very clear. First, what is the need for deriving the NRM displacement field and then calculate an NRM residual depth error?
The NRM is comparing two sections, in our case two CIP gathers. It will find by minimizing the difference of the two sections a smooth displacement field. By trace shifting, we create a reference gather. The displacement field is a lateral and vertical smooth version of the vertical shifts from trace to trace. By tracking through the displacement field, a smooth depth error is estimated.

It seems like this is simply flattening the arrivals in the gathers. Isn't this effectively the same as picking the arrivals and flattening them to a common depth?
Yes, in principle. But automatic picking or tracking needs additional parameters, especially for non-hyperbolic or weak events. Typically, coherence measurements are used and additional outliners must be detected and removed. Even though, a vertical smoothness between depth error branches is not guaranteed. As written in the updated manuscript for the accretionary wedge structure, we used 11000 depth error branches, each branch with 30 picks. The smoothness of the depth errors in space and depth will stabilize the linear equations and the inversion result.

Is the NRM technique mostly a convenient way to derive the residual depth errors, or is there an analysis here that makes it superior?
First of all, it is convenient and does not need any detailed quality check. We analysed in more detail the differences between NRM and PWD methods in the manuscript. The main difference (more details in the manuscript) is, that PWD detects low amplitude events with the correct dip (length of few samples and traces) whereas NRM is showing them with reduced (wrong) dip values. Due to the strategy by the NRM method to minimize the energy difference, the strongest amplitude events will dominate this inversion result. If small weak amplitudes need to be tracked by NRM correctly, a time-variant gain application before the NRM calculation will reduce this effect.

More explanation is needed for what the NRM does beyond simply picking the events to derive a depth offset.
We hope that the comments gave an explanation.

2. Secondly, I did not find an explanation for how the residual depth errors are used to update the model. Table 1 describes the second to last step as "Update the Tomography Model Properties that will Minimize the CIP-gather RMO". I did not find any description of this. How exactly is this done, and is this the only way in which the residuals are used?
Thanks for that comment. We include a new Chapter: "2 Method: Non-rigid matching and reflection tomography" and "2.2 Methodology of the ray-based grid tomography with CIP depth errors".
Here we explain the background of the grid-based tomography and limitations and how the CIP errors are used to update the velocity.

I thought that this is usually done with some constraint such as a linear or hyperbolic relationship with offset. How is it done here?
See comment above.

3. The comparison to the initial model is a problematic basis for making any significant conclusions. I certainly see that there is value in using the initial model to demonstrate the changes over the iterations; however, it is not clear that it represents anything that would make a useful basis for conclusions beyond showing improvement. Were the velocities considered to be the best possible that could be derived from a particular technique, and can that be demonstrated?
The only thing we can say is that the velocities, based on the horizontal alignment in the CIP gathers, is one solution for the shortest smoothing scale length we used (see also Chapter "4 Discussion" and "4.2 Model uncertainties by tomography". The tomography behaves better than the classical 1D depth focusing analyses because it accounts for dipping layers. But most important is that a "best" depth focussing pick is determined by fitting parabolic curves along the complete offset (aperture) of a CIP gather for a velocity update. This is not the case for non-hyperbolic grid-based tomography. In this Chapter, we further discuss how velocity uncertainties could be quantified. A future analysis sequence should be based first on a sensitivity test (checkboard test) followed by a Monte Carlo approach.

Are the wide-angle data comparable?
Yes and No. We discuss that in "4.1 Final velocity model and reflectivity structure".
NO: see Chapter "3.2 Initial velocity building from wide-angle tomography". The OBS profile and the model were offline in respect to the reflection profile and the model did not even fit the seafloor. Additionally, there was a gap of three OBS stations at the trench axis which reduced the accuracy significantly. We discuss the original OBS model and our initial model building

for the tomography. For example, the OBS model shows a nearly constant velocity of 3800 m/s between CDP 25000-26500 at a depth of 7000 m very close to the trench axis. The final velocity in this sedimentary area increases from 1750 m/s at the seafloor to 2280 m/s at 7400 m depth.

YES: On the lower slope, we reduced the velocities in our initial velocity model compared to the OBS model. The tomography increased the velocity again to the level of the original OBS model, locally even up to 500 m/s higher.

4. The wide-angle data are likely to be affected more by anisotropy than near-zero offset data because their ray paths have a significant horizontal component. It is fine to use an initial model to show how the technique is applied and how it changes from start to finish, but it can't be used to claim that the change in velocity has any geological significance as it is for some of the conclusions. It would be best if it were stated directly that the comparisons between initial and final models are only to demonstrate how the technique improves the results over the iterations and not to show that it is superior to any other technique. You can make it clear that the initial model does not represent a best result from another technique and that you are not claiming this is the degree of improvement that can be expected beyond other techniques.

   Yes, I agree with most of that.

   But that a change in velocity by a migration velocity analysis will have no geological significance is difficult to accept. The tomography is just like the smoothed version (even better because of reflector dip corrections and 2D assumption) of massive picked CIP-gathers or depth focusing errors (1D approximation) used for velocity building during the last 30 years. The subsurface structures were mostly interpreted in consideration of the determined velocities from near vertical and/or wide-angle data.

5. There should be some discussion of how anisotropy or effects of streamer feathering may contribute to residual errors or can be accommodated with this technique. There is a brief mention of anisotropy, but no mention of streamer feathering. These are obvious factors that can impact arrival positions within the gathers and skew the velocity models.

   Thank you for pointing out the anisotropy topic. We included a chapter in the discussion: 4.3 Anisotropic tomography. This chapter is connecting the OBS model to the results from the near-vertical velocity. We explain how the grid tomography in Chapter 2.2 can be used for an anisotropic tomography inversion with minor modification if the near-vertical data would have long offsets.

   Comment to streamer feathering: On the profile there exist local areas with significant deviations of shots from a straight line, but no additional information of the streamer position is available. We used a back-tracking method to position the streamer along the crooked shot profile to minimize offset errors. Additionally, the direct wave with known shallow water velocity was used to correct the offsets to smaller values along the streamer if needed.

6. My biggest issue with this paper is with the writing, mostly in sections 4 and 5. There are many vague and unsubstantiated claims made throughout these sections, and some in section 3. For example, in lines 368-372, what determines "large-scale length" what are "significant" velocity corrections?

   We re-submit completely revised chapters 4 and 5. Chapter 4 (Discussion) now includes:

   4.1 Final velocity model and reflectivity structure

   4.2 Model uncertainties by tomography

Here we included a paragraph about the scale-length reduction and show an example of the velocity updates, from the first iteration up to the last iteration. In this example, an offline structure with velocities of less than 1300 m/s is discussed.
4.3 Anisotropic tomography

It is not clear why the scale length is appropriate for the first iteration and why there is a corresponding reduction in scale length with subsequent iterations. What makes observed velocity updates more “pronounced”?
I think we did not explain that correctly. In the method chapter: “2.2 Methodology of the ray-based grid tomography with CIP depth errors”, we explain the motivation of the scale length factor (in space and depth). If the model is unknown the tomography needs first to update the long-wavelength variations and iteratively reduce the scale length from iteration to iteration to find smaller scale length variations. If a model already exists with the long scale length variations (in our case the OBS model) each iteration internally starts from the longest scale length to the shortest scale length. A data example is shown in chapter 4.2: Model uncertainties by tomography.

What defines a “fluctuation of velocity changes” and what determines when they are or are not “related” to “subsurface structures” and what constitutes structures that relate to these changes?
Removed.

This is just three sentences. There are more such issues throughout. There are far too many claims like this that are hard to determine what is meant and what the basis for them is. I have noted many of them on the manuscript, but this section needs to be rewritten.
Done.

7. There are few topic sentences. Without them it is hard to tell what the goal of the paragraph is and where the discussion is leading. The discussion wanders through a number of topics that are not very closely related to each other or seemingly to the rest of the paper. It is not clear why these topics are being discussed and whether they are to help establish the validity of the technique or the results from its application along the Sunda trench or Hikurangi. It is not clear why a new setting (Hikurangi) is added in the discussion rather than as part of the results with the Sunda results.
   We removed the Hikurangi example.

8. Overall, this is appearing to be a valid technique and the paper does a reasonable job of presenting it, illustrating how it is applied, and showing the expected results. What I don’t understand are the goals of the discussion and how it adds to the paper. The discussion and conclusions need to be significantly revised before this is published.
   Thanks, Done.