FIRST REVIEW ROUND

Response to Reviewers for

"Characterization and validation of tidally calibrated strains from the Alto Tiberina Near Fault Observatory Strainmeter Array (TABOO-NFO-STAR)"

Reviewer A:

Hanagan et al Borehole strainmeter calibrations

This paper provides an initial look at borehole strainmeter data from a new network of six instruments installed in NE Italy. The promise is that these data will be able to record transient slip on nearby faults in this tectonically active region. This type of borehole strainmeter network has analogs in the western US where borehole strain augments other instruments that measure crustal deformation.

The goal of this paper is to calibrate in-situ these strainmeters using their measured Earth tides and comparing those with modeled or predicted tides. Although the sensors in these 4-component extensometers can measure strain to less than 0.001 ppm, the presence of the borehole distorts the regional strain. It can be assumed that the material around the borehole is an isotropic, homogeneous material, but the reality is that is not the case. Consequently, these instruments require in-situ calibration and the Earth Tides provide a good signal since they are easy for these instruments to measure and the tides well understood and are repeatable. However, it does require a means to predict the tides. There is software available and they provide “reasonably” accurate prediction of the tides, subject to various assumptions. The paper builds upon the Hodgkinson et al (KH) paper, and importantly, tests the calibrations obtained from recordings from four earthquakes; the Turkey (M7.8), a regional M5.5, and two local M4.3-M4.5 events, both for their static offsets, and the inferred direction of the P-waves.

At some point, the work should be published, but at this point, it isn’t ready for publication. Here are some of the problems that I see:

1) The paper heavily relies upon the calibration of (KH) but only uses one facet, that being the unconstrained solution to the calibration procedure. Why weren’t the other solutions pursued including the KH “constrained” and “loosely constrained” solutions? Unlike the unconstrained solutions, the other two provides a bit more physics (and, importantly, constraints,) into their solutions and, often, the fit of the constrained calibration to the tides is nearly equivalent to the unconstrained solution.

This is a good thought to include the physical constraints, but we choose to focus solely on the non-constrained solution (which we now reference several times in the main text) because of the expected significant deviation from assumptions inherent to the lab/manufacturer’s calibrations. After completing the analysis with the non-constrained solution, we show through different ways that the orientations are likely in error and that four of the 6 sites exhibit negative areal coupling, which the constrained solutions do not allow for. Solving for the installation orientations as you *mention would add additional information, but would not necessarily help the goal of the paper (to calibrate the instruments, validate that they perform better than the standard lab calibrations, and characterize the signals for the array).*

We modified the last sentence of section 1.3 to: "Given the severe deviation from what is expected for the lab calibrations, we approach calibration following the non-constrained approach of Hodgkinson et al. (2013), which drops all aforementioned assumptions of the lab calibrations."

2) Although the paper alludes that the tidal models might not be accurate (after all, these models come with certain assumptions), I would have liked to see, at a minimum, more discussion, or better, some quantification of how that uncertainty could affect the calibration.

This is a very important point, though we believe it is very well addressed in Langbein (2010, 2015), which we cite in the introduction and discussion (4th paragraph of each section). Specifically, in the discussion, we reiterate the finding that uncertainty from the tidal models have: "...a potential 10-30% uncertainty from the tidal calibrations considering the lack of precision and accuracy of the tidal models." Given that this is well-presented and studied in the existing literature, we prefer to limit adding further discussion of it to the manuscript, and instead focus on the inversion uncertainty from the model covariance as a new metric in the context of the strainmeter calibrations.

In the methods section we also cite Amoruso and Crescentini (2009), to highlight that the strainmeters should be relatively insensitive to the choice of ocean loading model, and that the earth body tide is generally better constrained, which provides additional support for the tidal models in the region.

3) Although the unconstrained calibration fits the modeled tides better than than an assumed isotropic coupling model, it appears that isotropic model performs as well, if not slightly better than the tidal-calibration when comparing static offsets from the four earthquakes evaluated in this paper. That is, the RMS misfit of the tidally calibrated strain is 10.5 ns versus 4.7ns for the isotropic calibration. On the other hand, obtaining a 10 ns fit to coseismic offset could be considered as acceptable.

You are correct that, overall, the amplitude misfit is greater for tidally calibrated strains, but we consider only strains from the near field earthquakes as representative of the static earthquake strain field. Far-field (>> 10s kms) static strain offsets are noted to be far too large (as in this study, and references within Wang and Barbour, 2017). This may result from relatively higher dynamic strain influence over static influence from those events (see paragraph 5 of the discussion section). For the Umbertide earthquakes within the BSM array footprint, the observed, tidally calibrated static strains are far closer to the elastic model predictions, indicating that the static strains dominate at these distances. TSM3 is an exception, but the station seems to be compromised by another near-field process in the presence of high strain rates, and cannot be considered in the RMSE calculation for static strain offsets, because it is measuring something else (see the last three paragraphs of the discussion). Further indication that the tidally calibrated strains are showing better results for measuring the earthquake-related strains is the fact that the principal strain axes are oriented *toward (and perpendicular) to the direction of the event for both near and far-field earthquakes, whereas the manufacterer's calibrated strains show more haphazard orientations.*

I think a more convincing argument with the static offsets is to examine the measured coseismic offsets from each gauge and compare those to the predicted offsets from each earthquake. That is, for each earthquake model of predicted tensor strain, multiply those values by A-matrix in equation 1 to provide the predicted, gauge offsets. They can be compared to the observed gauge offsets. The A-matrix can be either the one computed from the tidal calibration and one computed using the theoretical values. My guess is that the predicted offsets using the tidal calibrations will be in closer agreement that those derived from 'theory' and will provide convincing evidence that the tidal calibrations perform better than those derived from theory.

This is another way to look at the information we provide in Table S11. We originally looked at the values as predicted gauge offsets, but found that they provided no additional information for the quality of the calibrations. Stated another way, whether we look at the forward (i.e. regional strains) or reverse (i.e. gauge strains) information relies on the same calibrations, and so looking at both seem valid but redundant. We settled on focusing on the calibrated amplitude and azimuthal misfit as more physically intuitive way to understand the mismatch, and also note that plotting the modeled and calibrated observed strains on the same map provides a visual inspection of this misfit.

4) This paper seems excessively long as it doesn't necessarily lay-out new methods. On the other hand, the use of obtaining azimuthal constraints is new. However, the discussion on the negative values of areal coupling is not new, but, fortunately, for the most part, held in the supplement. (I would have like to see a plot of the areal coupling vs barometer admittance with not only the TSM results, but those from KH plotted all together (ie, update KH figure 6).

We have included quite a lot of background in the supplement to cut on the length. The length of the main text can mainly be attributed to the methods of analysis for the "validation" signals, and the results of the calibrations and validations, which we hope you might agree are best kept comprehensive and reproducible.

We have added the KH results for areal coupling and barometric pressure response in supplementary Figure S10. Thank you for the suggestion.

5). Likewise, I question the validity of using the Mandler algorithm for tidal calibrations. Specifically, that method includes not only the standard M2, and O1 components, but also uses K1 and S2, both are very close to diurnal and semi-diurnal periods that can exhibit leakage into nearby K1 and S2 components. For example, of the 24-hour, diurnal period is very close to the K1 period of 23.93 hours. In principle, the two can be separated, but it requires al least a year of data. The longest period considered in this paper is 333 days (and more typically, 200 days)

This is a very good point to raise, and one that we attempted to address more clearly in the supplemental text: "preventing a long enough timespan to fully separate the suite of tidal *constituents **(at least one year is recommended to separate the constituents used following this method).”** Nonetheless, the two calibrations are quite similar, as demonstrated, for example, in Figure S8. This, in a way, provides one more point of comparison and validation for the calibrations, though not one that is central to the study. For these reasons, we hope to keep the alternative approach in the supplement.*

6) Some shortening of the paper can be obtained by deleting most of the material in lines 301-314 and the accompanying figures and tables in the supplement. Essentially, Baytap provides the signal to noise ratio. And, since the tidal calibrations fit the tidal data, there are no surprises from the PSDs showing little, residual power in the tidal bands. Perhaps, tabulating the regression coefficient (reduction in variance) along with RMSE should show that the tides were successfully modeled

This would be a good alternative way to view successful identification of the tidal constituents. Though, we feel that these lines and accompanying figures and tables also get the point across, and provide an external validation outside of relying on Baytap alone. Because the accompanying figures and tables are in Supplement, it leaves the choice to the reader whether they care to investigate the additional information further without significantly increasing the main text length, which will be the primary concern of most viewers.

Various comments and notes made while I read the paper.

Supplement:

Include a table with site locations, elevation, install azimuth and depth of instrument; make this a part of table S1.

This information has been added to Table S1.

Lines. 185 — I don’t see that this ‘reformatting’ deviates from KH; she set-up the LS in the same manner.

It is possible to solve the least squares problem directly following the formatting in Hodgkinson et al. (2013). To include the weights as a diagonal matrix (as is standard for least squares), we reformat it as described. The matrix setup has different dimensions for the problem implemented here vs in Hodgkinson et al. (2013).

Lines 216-231 — I don’t understand what is being done and what is being quantified? This method appears to quantify the uncertainty of the pseudo-inverse/Calibration matrix, but does not factor the large uncertainty, that being the epistemic error related to the model of the earth tide; The model predictions can be in error due to the local geology not conforming to the assumption of a layered model of the earth’s structure. What is the meaning of 3%?. Table 1 lists another set of percentages which I assume are the standard deviations discussed in line 229. And, given that this is a numerical exercise, what is the justification of removing calibration values that exceeded 32% (ie, why are the outliers that are in the ‘tails’ tossed?). By tossing these outliers, doesn't this artificially decrease the 1-standard deviation percentage uncertainty listed in Table 1? Perhaps that large uncertainty is due to lack of constraints which have 4 degrees of freedom (16 observations and 12 unknowns).

Essentially, this uncertainty is that of the least squares fit uncertainty. It does not factor in the model error of the earth tides, because we do not have this information (hopefully, future work with arrays of broadband seismic station near the borehole at the surface will be used to provide external reference measurements for seismic events). We found it informative to understand the uncertainty on the model fit, which has not been reported before for each component in prior studies, and seems to corroborate some of the misfit as stated in the discussion. The effect of tidal model uncertainty is more thoroughly addressed in Langbein (2010, 2015). The percentages in table 1 are trimmed to represent the 1-standard deviation uncertainty (dropping 32% to retain 68% of the values), because, as you point out, the uncertainties are extremely large with the tails (particularly when the nominal value is near 0 and the percentage is likely to blow up). A larger uncertainty could still be calculated from this 1-standard deviation value.

*To clarify this, we added "This provides an estimate of calibration uncertainty as a percentage of the nominal strain value at the level of one standard deviation (68%) for the regional strain components at each station**, and minimizes the effect of very large percentages when the regional strain approaches 0.**" Hopefully this is clearer. We also more clearly state that the "percentage is insensitive to the range of random strain values, and stable for the number of perturbed calibrations within 3%* ***of the values reported later."***

Line 250; principle strain???

This is a good clarification we have now included.

Line 270-278; For the M7.8 with its the longer, 30-60 minutes windows, there could be some "drift" in the time-series which is usually classified as random walk. I suggest using a program like Hector (https://teromovigo.com/product/hector/) or est_noise (https://github.com/langbein-usgs/est_noise) to estimate the co-seismic offsets and their uncertainties.

This is a good point to raise, though we do not observe any drift in the time series, especially with the tidal and barometric pressure corrections, so recalculation of the cosiesmic offsets seems unnecessary here. We tested this through adjusting the averaging window, which is included in the main text description of the paragraph. We also include the timeseries with offsets marked for TSM2 from Kahramanmaras in Figure S3 to demonstrate the lack of drift.

Line 305. — reference to "standard earth scope calculations".

Our knowledge here comes from migrating the code established by K. Hodgkinson toward the current process, which primarily relies on python scripts. We do not have a good peer-reviewed reference, but can highlight that the main author and several coauthors have had a direct hand in handling the standard calculations.

Line 323 — Not, ALL, STAR site achieve.... TMS4 has a 1.3 ns RMSE misfit which exceed that KH threshold of 0.84. Also, I wouldn't place too much emphasis on fitting phases; with small amplitudes, it is harder to measure the phase; high amplitude yield batter resolution of the phase measurement. I am not clear how the numbers in Table s2 were calculated? How does amplitude relate to RMSE? And, why are the numbers associated with each gauge dimensionless? Was weighted LS used?; If going into the detail of the fits, why not break each column into two columns, with one column showing the amplitude(magnitude) of the residual vector for M2 and the second column for O1? If the amplitude of the residual vector is used, I see no need to tabulate phase.

The threshold in Hodgkinson et al. was 0.84 for the RMSE misfit calculated according to main text equation 6. The value for TSM4 is 0.77 (Table S7). The amplitude RMSE is in fact larger (the 1.3 ns value presented in Table S6), but this is a different metric than presented in Hodgkinson et al., (2013). I have included a reference to the table for clarification.

You are correct that the phases are less certain. We include the phase RMSE for completeness, to share the finding for those that would not be familiar with the concept that the phases are more difficult to constrain.

Table S2 shows the Baytap calculated amplitudes and phases with associated standard deviations, which we describe in the Table caption and briefly introduce in section 2.1 of the main text. To decrease the length of the manuscript, we included further description of the amplitude and phase estimates in Supplemental Text A. The units of the columns are at the top of the Table (nstrain, and degrees). We present the information in the way that Baytap presents the results, as separate amplitude and phase values. So, we do not need to calculate phase separately, but would need to calculate the residual vector separately.

We did use weighted least squares, but found little difference in the results, and thus chose to remain consistent with the standard for the NOTA BSMs (unweighted). The weighted results are presented in Table S7.

Lines 336-345 — see previous comment for lines 216-231.... Previously, the claim is that the calibration matrix is good to within 3% (Line 231). This needs a better explanation.

Hopefully this is clearer with the updated description in the methods section, detailed in the response to lines 216-231.

Line 346-351 — rather than using another figure, compute the RMSI between the two techniques and summarize with one line .... For the comparisons using the unweighted versus weighted LS to estimate the calibration matrices, the RMSI ranges between x and y (and hopefully, x and y are small).

Great suggestion. We removed the figure and report the RMSI values instead.

Line 389; Probably the installation azimuth is incorrect. As an example, for TMS6, with a revised azimuth for gauge0 of 31 degrees (instead of 47.3), I get the following shear coupling coefficients:

Num	azi	d1	d2.
1	59.	2.0714	2.3082
2	119.	2.1261	2.6069
3	179.	2.1983	2.0884
4	209.	1.9147	2.3592

(Azi is CCW from east)

These are closer to the expected value of the shear coefficients

I definitely agree! Through several lines of evidence we come to a similar conclusion, as stated in the second and third to last paragraphs of section 3.1.

“Magnetometer” versus “compass”; I wonder if the terminology ‘magnetometer’ is appropriate, Magnetometers can measure the magnitude of the magnetic field. They can be ‘total’ field (non directional’, or “field’ that measures the magnetic vector from which the horizontal azimuth is obtained. Although the compass may measure the magnetic field in two directions, the result that is being used is the azimuth. I don’t think the total magnetic field is being recorded in the borehole.

The instrument itself has a magnetometer. The measured value is essentially rotated until the field strength matches proper orientation. This was verified with our engineer Wade Johnson.

Equation 7 — this also could be used to compare your preferred calibration matrices with those obtained with “Weighted” uncertainties and the Mandler method

This is a good point, and we have included RMSIs for the weighted calibrations in the main text. The Mandler method goes directly from counts to nanostrain, so we cannot compute a direct RMSI. To highlight this, we also added the sentence: “One main difference with these calibrations is that they go directly from counts to strain, and therefore skip the linearization process” to the last paragraph of section 3.1.

Figures 6 and S9 — Have the tides been removed? I think the plot would be clearer if the tides were removed from each time series. An alternate presentation is to perform Baytap analysis on the tensor strain and pressure data and record the pressure admittance (sensitivity).

They had not been removed for the weekly plots, but now they are. Thanks for the suggestion.

Missing Table S10

Line 383 — Table S8 contain the Mandler analysis and NOT the estimates of Cs and Ds for each meter……

The supplemental tables mentioned in the main manuscript need to align with the numbering within the supplement. I think there is at least one missing table.

Line 396 — Again, table appears missing — paper refers to Table S9 as if it provides results from the Mandler method but instead has dynamic strain information…

Line 438 —Table S9

There was an error in the supplement where Table S7 was repeated twice. All instances should be fixed now.

Compared misfits in Table S11 between predicted coseismic strain offset and those calculated using "tide model" and "Lab modeled". Simply summing the squares of each column and dividing by 15 (column 11 and 12)

Figures vs text … Manufacturer's calibration vs "lab" calibration….Their relation needs to be defined. "Lab" calibration is a shortcut for Manufacturer's calibration….

We have corrected all instances of "Lab" to "Manufacturer's."

Line 557 — I disagree. The paper did not test the constrained solution types spelled-out in KH. I agree that the unconstrained solutions performed here better perform than the ones that use theoretical values (and azimuth) (or "lab/manufacturer's).

*To reflect this, we have softened the language to state "The tidal calibrations **more** adequately…" as opposed to "are needed."*

Line 559 "the tidal calibrations show remarkable success for resolving dynamic and static strain orientations approaching single nanostrain-scale". This doesn't quite make sense…. "the tidal calibrations show remarkable success for resolving dynamic and static strain orientations for signals of the order of a few nano-strain"?

This is a good rephrasing, we have included it.

Line 604. Errors in the tide models do not necessarily systematically bias the 'calibrated' strains across the network. Local inhomogeneities to each site could bias either "up or down" the strain measured in the borehole relative regional strain.

*Good point, we have clarified this to refer to a single station, rather than the network: "…the lack of systematic over- or under- estimation of near field strain amplitudes **for any single station**…"*

Line 611 — preface this discussion by stating that of all of the strainmeter, TSM3 tidally modeled offsets vastly exceeded the expected offsets relative to the other strainmeters.

We have changed the first sentence to: "TSM3 recorded static strain offsets that far exceed those of the other stations of models, even for the near-field events."

Line 657…. "For what it is worth"; I recall that dilatometers in response to the 2004 Parkfield EQ had a post-seismic transient, but that response was short-lived relative to the longer-term, post-seismic transient recorded by the GNSS network and creep meters. That short-lived, perhaps exponential in shape, suggests a diffusion mechanism like a poro-elastic phenomenon.

Very interesting! I think several lines of evidence point toward short-lived diffusive poroelastic influence in some strainmeters, as well. A coauthor on this paper has subsequently dived into the TSM3 postseismic transients, and believes that there is also an afterslip signal that takes over (I also saw an atmospheric pressure increase that coincides with the afterslip signal, complicating a simple tectonic interpretation). The interesting thing to point out about the strain transient vs pore-pressure sensor transient in this paragraph, is that the two do not always coincide. I believe the pore pressure transducer responds to a different hydrologic signal that is not at the same depth as the strainmeter. Moreover, the hydrologic connection of the pore pressure sensor may have changed through time, as the borehole was initially caked with mud (more so than the other stations).

Supplement: A

Nice method for detrending the data based upon Blewitt's MIDAS algorithm. However, how do you determine whether the time series has an offset, or an outlier? I suppose the outlier could be considered as two offsets that cancel each other.

We decided not to differentiate too much between outliers and offsets because, as you state, the outlier cancels itself.

Supplement C:

The vertical coefficient symbol, v, looks close to the poison's ratio symbol; (equation S5)

You are right, we have changed it to a z.

It might be worthwhile to compare the areal dilatation derived from tidal models with Rayleigh waves that are measured by a co-located, broadband seismometer. See a recent paper by Canitano , 2024

(https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/114/3/1589/635222/Constraints-on-the-Calibration-of-Borehole). This might not be possible if the instruments are not colocated.

I completely agree, this was discussed amongst coauthors as potential future work. Hopefully, we could include an array of collocated seismometers.

Manufacturer's vs Lab calibration notation. I think a better identity is either "theoretical" or "nominal" values. The use of 'calibration' implies that actual measurements were made; both 'theory' and 'nominal' imply some assumptions were made.

We have changed all instances to Manufacturer's, which is most consistent with the EarthScope Metadata nomenclature.

Recommendation: Resubmit for Review

Dear reviewer,

Thank you very much for the detailed review. Your expertise with the instruments and time series analysis are clear, and we hope that you find our edits and responses satisfactory.

Best,

Cassie Hanagan, on behalf of the coauthors

Reviewer B:

This is a straightforward, carefully done, and neatly written paper. The authors calibrate a set of strainmeters in Italy using tidal observations, and do a good job of moving the approach to strainmeter analysis forward to more sophisticated techniques. The writing is longer than it needs to be, and omitting some details would probably make the paper better, but overall the descriptions are clear, and the discussion of the results is direct and measured. I have only minor comments.

Lines 147-148.

Would be nice to describe what Baytap does a bit more in the main text in a sentence or two, as the order and approach to fitting matters, and it seems carefully done. Eg, you remove a long-term trends, iteratively fit and remove offsets, etc.

Thank you for the suggestion, we have now included the following sentence: *"Baytap08 uses a Bayesian modeling approach to estimate tidal constituents and other variations correlated with external data (Tamura, 1991; Tamura and Agnew, 2008)."*

Lines 214-231.

The descriptions that came before this made sense quite quickly, but this section is hard to read. Perhaps it needs to start with a goal: apparently "an estimate of calibration uncertainty as a percentage of the nominal strain value." Except I'm not sure why that's more useful than just the uncertainty on the calibration parameters?

We wanted to provide an uncertainty on the resulting, regionally oriented strain values rather than on the calibration coefficients themselves, which are less physically meaningful when separately considered. This way, we can analyze the effect of the calibration uncertainty on calibrated strains that most people will use.

We have starte a new paragraph with the leading sentence that should clarify our goals better: "We use the resulting model covariance (Cm) to estimate a percentage uncertainty on the calibrated regional strains."

Covariance usage.

If you want people to actually start considering uncertainties in calibration, it would be useful to make it obvious how to use the covariance matrix in equation 5. You've set it up to be quite simple---to get a realization of the calibration matrix, you add a realization from that covariance matrix to the mean.

Other researchers may well want to redo their results with lots of realizations of the calibration matrix, to see if it matters. It would be useful to explicitly say how to get one and to say explicitly that your average calibration matrix and average covariance matrix are in the supplement. Actually, comma-delimited text files would be much better if you actually intend these calibrations to be used. Copying from pdfs often leads to errors.

These are excellent points for the covariance matrices. Though, because the covariance matrices are applied to the moore-penrose pseudoinverse of the calibration matrices, I worry it would be misleading to provide them directly. This is partly why we provide a percentage estimate of uncertainty on the calibrated strains as described in the previous comment response. The information to reconstruct the covariance is available through fitting the predicted tidal constituents and observed values from baytap via least squares and computing Cm through Equation 5.

In the data availability, we state that there is now a python package to run through the calibrations, and that the calibration matrices are available in the EarthScope metadata files. We believe it would be best for people to find the calibrations through these pathways. Hopefully it is clearer in the data availability statement now.

Section 3.4

I think it's pretty clear that the modelled Kahramanmaras strain is inconsistent with the calibrated strain, but is it not worth directly comparing with your uncertainty estimates on the calibration? Or maybe it's not inconsistent---it's just small (lines 564-566)?

It is just small, and potentially influenced by other signals that occur following the earthquake (e.g. barometric pressure fluctuations). Including the percentage uncertainties is still a good suggestion, so we have included it: " The observed static strains for TSM6 from the Mw 7.8 Kahramanmaras event are one exception, though uncorrected response from environmental influences could easily bias the small (<3 nanostrain) offsets ***in addition to the 5% shear strain uncertainty from the calibrations (Table 1).****"*

Figure 1.

Very minor, but I think most people use stars for earthquakes and triangles for stations. At least that's fixed enough in my head that I read the figure wrong repeatedly.

I don't think you are incorrect, but the seismic stations are triangles, and given the name of the network (STAR), we prefer to keep them this way, and labeled in the legend accordingly.

Thank you very much for your detailed and helpful review!

SECOND REVIEW ROUND

Dear Reviewer,

I would like to thank you immensely for helping me to improve the clarity of the manuscript. I have applied nearly all of the suggestions you provided, including cutting the comparison with the manufacturer's calibrations substantially. I hope you find the reviews satisfactory - we are looking forward to getting this manuscript out there for use by the scientific community.

Many thanks,
Cassie

(all of my responses will be green)

Reviewer A:

This is my second time of reviewing this paper. Like the other reviewer, I thought that the original manuscript was too long. The revised version appears to be the same length as the original submitted to Seismica. Consequently, my review will focus on portions that can be shortened or deleted.

The paper concentrates on the calibration of six borehole strainmeters recently installed in NE Italy. The method employed for calibration is one of the methods used by Hodgkinson et al (KH) when they examined the data from roughly 70 borehole strainmeters installed in the western US. The strainmeters in Italy are the same model as used in the western US, so the problems encountered by KH (and initially, Roeloffs) are the same found for the six Italian strainmeters. Where this manuscript deviates from that of KH is the demonstration that the calibrated strainmeters provide accurate recordings of both dynamic and static strains from both local and regional earthquake sources. Likewise, deviations from the expected strains appear to provide insight into the hydrological regime of the materials adjacent to the borehole. Consequently, I suggest that the paper to be shorten to focus on these aspects in preference to comparison with calibrations using the manufacturers' values.

Also novel in this paper is a method for evaluating the aleatory uncertainty of the tidal calibrations. This dovetails with Langbein's work (2010 and 2014) that evaluated the epistemic uncertainty in tidal calibrations due to predictive modeling of Earth tides. Although in my first review I indicated that I didn't understand the method, but I do now after a few changes (and one more suggested below).

Thank you for these clarifications, we both shortened and clarified the section according to your suggestions.

I think the only reference to the manufacturers calibration should be through the RMSI index defined in the paper. The difference in RMSI between the tidal calibrations used in this paper and the manufacturers' suggested calibration is significant enough that one doesn't need to list (and plot) the differences in terms of the dynamic response, coseismic offsets, and the response to surface loads. Instead, keep the focus on the degree that the tidally calibrated strainmeters match the anticipated dynamic strains and the inferences between expected coseismic strains and the observed. Likewise, the negative areal coupling previously found by Roeloffs and KH suggests other mechanisms. Although tangential to revising the manuscript, the supplement that discusses the negative coupling terms could be deleted and simply referenced to the Roeloffs paper.

We took many of your suggestions later for eliminating the comparisons to the manufacturer's calibrations. The comparisons to the manufacturer's calibrations are largely kept in the supplement, but figures and other edits in the main text as mentioned in line to your comments below have been modified and implemented. We also removed the Roeloff's supplemental text.

In the revised manuscript, I don't see any justification for excluding the loosely and fully constrained, calibration methods provided by KH. For instance, the areal coupling coefficients estimated for the STAR strainmeters are all within the range of -1 to +4 specified by KH as a constraint. Likewise, by removing the constraint that the actual azimuths of the strainmeters as those measured in the field, then one can solve for the shear coupling coefficients either with the methods of KH, or use the coupling equations found here for the Italian array with least squares and juggle the azimuths to find a set of shear couplings that minimizes the spread (RMS) of the inferred shear coupling coefficients. Instead, one can justify using the least-squares approach used here (like the KH unconstrained method) since it allows for calculation of the covariance of the coupling model, and consequently propagating that into the aleatory uncertainty of computing regional strain.

You are correct, we justified the non-constrained solution based on prior knowledge that the installation conditions were not isotopic, but your description provides further justification for implementing the non-constrained solution. We have added the sentence "We select this non-constrained approach, as opposed to solutions obtained with gauge orientation and coupling coefficient constraints from Hodgkinson et al. (2013), because it allows us to quantify the covariance of the coupling matrix, which we propagate into an estimate of aleatoric uncertainty for the regional strain solutions." following equation 1.

I will list some places where the manuscript can be shortened. I will reference the pdf file that contains "tracked changes".

Lines 188-212. I would not call equation 2 a deviation from KH. Instead, state that equation 1 can be rewritten as a standard least squares problem by sorting both the coupling matrix as the model vector, m, and the observation matrix as the data vector, d. This becomes equation 2 and the usual LS relations apply, that being equation 4 and equation 5 — delete the discussion about equation 3.   Actually, as written, equation 3 doesn't make sense. One can rewrite equation 1 by sorting the data and model matrix into 4 different LS formulations per equation 2. Or, combine the G matrices of the four LS problems into a 16x12 **block** diagonal matrix. I think this is what is meant with equation 3.

This section has been substantially shortened with your suggestions. Thank you for clarifying the block diagonal matrix aspect, that is a much more concise way to put it.

Searching on the word “manufacture”, I get:

Line 437-38 delete sentence

This change has been addressed.

Figure 6; remove time series that use the manufacture’s calibration; likewise, remove the “lab” calibration, principle strains in Figure 7.

This change has been addressed.

Liness 489-92 delete

This change has been addressed.

Lines 503-04 delete

This change has been addressed.

Lines 544-52 delete

This change has been addressed.

Figure 8; remove coseismic offsets derived using the manufacture’s calibration.

This change has been addressed.

Lines 576-597 — attempt shorten this paragraph

This change has been addressed.

——

Other comments:

Line 220 — I think an introduction to uncertainty method is needed: “Although the uncertainties in the estimates of the coupling terms are provided via the model covariance, application of the Moore-Penrose inversion eliminates a direct estimate of the uncertainties in the calibration matrix. This is overcome with simulation"……

This is an excellent way to phrase the uncertainty analysis. I have added "Although the uncertainties in the estimates of the coupling terms are provided via the model covariance, application of the Moore-Penrose inversion eliminates a direct estimate of the uncertainties in the calibration matrix. This is overcome through the simulation of thousands of plausible calibration matrices…"

Lines 309 to 322 — I would cut this paragraph and state that the Baytap analysis indicates that there is high signal to noise with the exception of….. Correspondingly, eliminate figure S4 and Table S5

This change has been addressed.

Line 332 to 339. Consider deleting; interesting, but not required.

This detail was included based on the concerns of a coauthor, and we would prefer to leave them in because it provides more meaningful numbers for comparison (given that amplitude and phase are different units).

Figure 5 "reasonable", what does that mean? (Or, is that a typo for "regional"?) Also, what is the trimmed, 1-standard deviation? Or, what is the standard deviation before 'trimming'?

We have changed this to "plausible," with more description in the methods section for how the calibration uncertainty is estimated. Because we pulled from full distributions as described in the methods section, the 1-standard deviation value becomes significantly inflated by the tails and near-zero values. Therefore, we discard 32% of the percentages and take the mean. This is more thoroughly described in the methods section 2.1. We have added a reference to the section in the caption.

Line 378, define "I" as the identity matrix….

This change has been addressed.

Line 400 — are the results are in-line with that from the NOTA strainmeters?

Some of the nota BSMs likewise have unequal tidal amplitudes and phases for gauge combinations that should be equal in the isotropic case. This was found in both Hodgkinson et al. (2013) and Roeloffs (2010). I have added the sentence: "Similar deviation from the isotropic case was found for several NOTA stations (Roeloffs, 2010; Hodgkinson et al., 2013)."

Figure S5 — I don't like the plot of phases for gauge consistency. If the amplitudes of the differences between the two qualities are small, then the uncertainty in phases becomes large and that uncertainty isn't shown. I suggest removing the phase-consistency plot.

This change has been addressed. Thanks for the suggestion, the amplitudes reflect the desired information sufficiently.

Line 409 — It looks like table S8 shows the calibration matrix when it is assumed that the measured orientation is indeed correct. However, the discussion centers on the coupling equation and specifically the estimated value of shear coupling.

The presented values are the coefficients. The table caption describes how the coefficients are organized: "Numbers correspond to individual areal, differential, and shear coupling coefficients (in rows) for each gauge (in columns), consistent with the structure of the calibration matrices presented in table 1."

Line 470 — doesn't TSM-1, like TSM-2 also show a positive correlation with air pressure? (And, hence an overall, positive, areal strain coupling?). This is consistent with the caption for Figure S8.

Excellent catch – this has been fixed to include TSM1.

Figure S8 — looks to me that TSM-4 areal coupling could be positive for barometric pressure; ie, the tidal calibration areal strain is in phase with the manufactures' areal strain….

The pattern for TSM4 from Figure S8 matches those of TSM3-6, indicating a negative areal coupling. No installation orientation is available for TSM4, so the manufacturer's calibration is not provided.

Figure s7 and s7 — From a discussion point of view, it seems that inclusion of the shear strain time series clutters the figures — I suggest removing the shear strain plots.

I think you are referring to Figure 6 and/or Figures S8 and S9? I think leaving the full time series may be useful, especially in the supplement, as reference for anyone wanting to look closer at the full tensor strain response to various reference signals.

Section 3.3 dynamic strain…. Why is TSM-4 not included? I realize that the manufacturer calibration is not available, but for the unconstrained, calibration used for TSM-4 (like the others), the principle strains from TSM-4 can be shown for the 3 sets of dynamic strains.

TSM4 had a long sequence of outages and difficulties getting consistent power hooked up to the station, so it was not recording during the event. This is stated in section 2.2.

Line 536. Change "modeled magnitudes" to "predicted strains".

This has been corrected.

Lines 603 to 608 — the two sentences to me don't make sense (or are awkward). The model offsets and the strains for TSM-6 are in agreement for the Turkey EQ.

For the Turkey earthquake static strains, the principal strain axis directions are misoriented 22°. I have rephrased the sentence, hopefully clarifying my point: "For example, the observed static strains for TSM6 from the Mw 7.8 Kahramanmaras event are misoriented by 22°, but uncorrected response from environmental influences could easily bias the small (<3 nanostrain) offsets in addition to the 5% shear strain uncertainty from the calibrations (Table 1)."

Line 617 "Majority of the time"? That's perhaps a radical statement. Your recent paper using same model-type strainmeter to measure coseismic strain offset for the Ridgecrest EQ would seem to contradict the notion of "majority".

Fair point ☺. I softened the language and dropped that qualifier. In other papers detailing static strain response, including some from Barbour and Langbein that are cited, the static strains are determined unreliable.

Table S8 doesn't seem to correspond to the discussion in lines 409-420.... I was expecting to see values of estimated C and Ds in S8, but the numbers shown in S8 appear to the calibration matrices using equation 8 of KH.

These are the calculated coefficients. The figure caption reads: "Numbers correspond to individual areal, differential, and shear coupling coefficients (in rows) for each gauge (in columns), consistent with the structure of the calibration matrices presented in table 1."

Table S9 — the calibration numbers from the Mandler are much different (10x?) than those in S6 and T1. I believe that is due to Mandler using the non-linearized version of the gauge data. I see that the S9 says the calibration is derived from the raw data, but, perhaps one needs emphasize that the gauge data have not been linearized. (Parenthetically, why does Mandler choose to use the un-linearized data? Seems strange as it is easy to linearize those data....). Any possibility of redoing the Mandler calculations using linearized strain? Then, one can evaluate RSMI.

The original motivation for the Mandler calibrations was to provide a calibration in absence of instrument information (for example, the gap or diameter necessary for linearization). This was true for strainmeters he worked with in Taiwan. I feel it would not add much to the paper to redo the calibrations with linearized strain, but it does provide an interesting additional point of comparison for the separate methods that any interested reader can discern more from in the supplement. I could see, for example, a comparison of these calibrations being used to determine what bias the Taiwan calibrations may have without linearization.

It is a good idea to make the point absolutely clear that the Mandler calibrations are not linearized. I added this to the caption.