A description of the inversion algorithm used to compute the initial ocean surface deformation from sensor-obtained time series should be provided. A great deal of the error and uncertainties the authors are trying to quantify have as much to do with limitations in the inversion scheme as with location of the sensors. Some of the observed errors may have little to do with sensor location, but a lot with other inversion parameters.

1. Since the proposed sensor location is in the seismic rupture area (pre-seismic sensor depth will be different from post-seismic resulting in a BPR time series drift) an explanation of how the inversion process handles the drift in the sensor signal due to co-seismic displacement of the sea-floor should be provided.

R:> It is true that inversion algorithms may be subject to inaccuracies arising from a wide range of origins and limitations. However, in the context of the present work, it is assumed that these inaccuracies will weigh equally for all tested inversions since the process is the same and, consequently, the difference in result can be attributed to sensor placement. In that sense, it is an underlying hypothesis that the methodology as proposed can be used with other inversion schemes as well.

In particular, the algorithm used in this article was proposed by Tsushima et al (2012) (op. cit. in the manuscript). They proposed a method of tsunami waveform inversion to accurately estimate a tsunami source by incorporating the effect of permanent seafloor deformation recorded by ocean-bottom pressure gauges (OBPGs) within the source region. In order to take the drift into consideration, it is assumed that coseismic rupture propagates with infinite velocity. The general expression is reduced to an equation relating observed OBPG waveforms to initial sea-surface displacement at the source by using a Green’s function consisting of two terms: The Green’s function used in regular tsunami inversion and a correction term to account for water-depth change in response to permanent seafloor deformation.

We have modified the text to highlight the hypothesis mentioned above and that the use of this particular inversion scheme is aimed to be a proof-of-concept rather than relying specifically on this method. (See the expanded Page 4, line 27-33).

Page 4, line 27-33: “The method builds on the premise that, in order to determine the tsunami source, an inversion procedure must be implemented. While an inversion algorithm can be subject to a wide range of errors and uncertainties on its own, for the purpose of the present work it is hypothesized that using a single inversion procedure will weigh equally those errors allowing comparison 30 among different sensor configurations. Consequently, it is the procedure presented herein can be applied independently of the inversion method, and the nature of the sensors. Similarly, other sources of signal noise such as atmospheric variations, tides, etc.”
2. The authors should explain whether sensor signal noise (seismic (Raleigh waves), tidal cycle, atmospheric pressure variations, electronics noise, ...) has been included in the modeled sensor time series and justify why it was included or left out, and what the effects of inclusion or exclusion may be on inversion results.

R:> It is true that in the original version of the manuscript, these effects were not mentioned. We follow previous research in assuming that pressure variations due to non tsunami phenomena (ocean tides, atmospheric and oceanographic disturbances, vertical deformation of the seafloor, and seismic waves) have very different time scales from those of tsunamis and can therefore be easily separated, e.g. Tsushima et al (2009). For operational settings, Tsushima et al (2009, 2012) applied a 60-s moving average and a low-pass digital filter with a cutoff period of 60 s (Saito, 1978) to remove seismic waves (acoustic waves which are converted from body waves at the seafloor). The ocean tide components were removed by subtracting the sea level variation computed from a theoretical tide model developed by Matsumoto et al. (2000). Those subtraction was implemented during realtime waveform data processing. For our purposes, however, we did not consider these effects and even in the case of the Illapel earthquake and tsunami, the tides were removed during preprocessing.

We have mentioned our scope now in the text, Page 4, L31-33.

Page 4, L31-33: “Similarly, other sources of signal noise such as atmospheric variations, tides, etc. It is noted also that in actual operational conditions, incoming data from sensors could need to be pre-processed as to improve the signal to noise ratio. For instance, Tsushima et al. (2012) use a 60 s moving average to process data from ocean bottom pressure sensors.”

3. While the combination of three metrics into a single Accuracy Parameter is convenient for evaluation, it does not always result in the most accurate evaluation. For instance two identical time series with only a small phase shift originating from a small discrepancy in arrival time will result in very large error estimate in the “skill” (waveform) metric (as defined in Eq'n (5)), although both waveforms are basically identical, such waveforms should only show error in arrival time since H and ‘skill’ accuracy are 100%, but according to the metrics used, it will also sure large errors in the “skill”. The authors should provide some comments on that or at least acknowledge the limitations of the metric used.

R:> We think we understand the source of the reviewers’ concerns, but we think the example highlights the benefits of combining multiple metrics. Of the two time series alluded (note we understand these are two time series resulting from the inversion, not two time series that need to be compared one against the other directly as would be the case if one of them is the target and the other the observed time series), it is expected that only one of them would be closer to the true series. Hence, one of them will show better values in arrival time and skill, while sharing the same value for height. Hence, one of them will be evaluated as better among the two but probably would compare less favorably against another having better skill.

We have added a short introduction explaining the idea behind the combined metric, in page 6, L24 and page 7, L1-14.

4. The authors should explain the rationale for capping the error in all three metrics at 100%.

R:> The original intention was to provide a common base of comparison, and to minimize the possibility of one of the metric carrying too much weight in the evaluation. By saturating at 100%, all cases reaching this value are weighted the same thereby allowing discrimination by differences in the other metrics.

We had briefly mentioned this in the original manuscript, but we have now expanded the description in page 7, line 4-19

Page 7, line 4-10: “In assessing the accuracy, the error between observed and forecasted quantities is estimated for arrival times and maximum 5 tsunami amplitudes.
However, one possible difficulty in establishing a standard metric is that each of these parameters has its own scale with significantly different ranges of values. For example, while an error in arrival time of a few minutes can be considered reasonable (for instance, less than five), a variation in height of more than one meter can signify a large error. To provide a common comparison basis for all possible sensor configurations, each parameter error is adimensionalized by dividing by the reference provided by the observed dataset. In addition, it is possible that one parameter having a large error could bias the combined assessment to be implemented. Therefore, the error estimated is capped under the consideration that errors larger than the observed value will be treated as equally significant.

5. A brief description of the JAGURS tsunami modeling code (mentioning physical model, discretization approach, validation tests) should be included in addition to the reference, for the benefit of the unfamiliar reader.

R: The suggestion was considered, and is included in Page 9, Line 14-17.

Page 9, Line 14-17: “JAGURS is a numerical model for dispersive tsunami wave modeling. This is a parallel software which solves the nonlinear Boussinesq dispersive equations in spherical coordinates. To solve these equations, a leapfrog staggered-grid, finite-difference calculation scheme is used.”

6. pp8, l25: The authors state “30 arc sec GEBCO global bathymetry is used since that resolution is sufficient for subfault size larger than 40x40 km”, but in pp5, l10 they state the size of their finite faults is 700x700 arc sec (20x20 km), this seems to imply they will need higher resolution than 30 arc sec for the bathymetry. Perhaps, in their statement they mean “smaller than 40x40 km”, not larger?

R: We have decided to remove the sentence, to avoid confusion. Our minimum source size would be constrained by a single source and the grid used is 23 times smaller than the unit source size along one dimension.

7. pp5, l12-15: The authors state that due to their current locations outside the study area, DARTs will not be used as sensors. An explanation of why DARTs cannot be relocated if it is determined they would be more useful at different locations, would be appreciated.

R: As stated in previous responses, the goal of the article is to provide a methodology to compare and evaluate among different sensor configurations. Whether these configurations include or not DART sensors are not intrinsically relevant, although they could easily be included if needed. Moreover, the basis is to evaluate how to design a new sensor array, for instance, in places where a system was to be built from scratch. A byproduct of this work could be the relocation of a DART buoy, but it is not the intended goal to be site-specific.

We acknowledge that the original wording in the article may induce confusion from this intended meaning, therefore we have modified the text to clarify this. See page 5, Line 20-25.

P page 5, Line 20-25: “They cover an area spanning about 7\textdegree\text{U} \times 3\textdegree\text{U} (latitude and longitude), consistent with the so called Northern Chile Gap (Comte and Pardo, 1991; Metois et al., 2013). In what follows, although an inversion process can benefit by other data sources such as Deep ocean Assessment and Reporting of Tsunamis (DART) buoys, these are not considered in the analysis under the premise that the area of interest is developing a completely new system. On the other hand, for the particular case of Chile, the location of the existing DART buoys 25 in the area of interest is such that requires longer observation times than the ones studied here (Williamson and Newman, 2018).”

8. pp13, l19: One of the conclusions states that the highest accuracy results are achieved when, at least, one sensor is located along the main energy beam. This is an obvious a priori conclusion, however it also points at the possibility that the identified locations may not be optimal for events with slip distributions disimilar to the test
cases, particularly larger events with longer rupture lengths that may exhibit an energy beam not necessarily focused on the sensor. They authors should comment on that.

R:> It is true that the performance would degrade for actual cases where the slip distribution would be such that no sensors are located near the main energy beam. Acknowledging this, we evaluated scenarios where the target earthquake was located in between the original target earthquakes, and also tested the performance for an existing earthquake with its slip distribution. While some degradation of the results is present (see Figures 5-7), the overall performance is considered acceptable.

We have now updated the discussion on this regard in the updated manuscript. See P18, 2-19.

9. The manuscript is for the most part well-written, however, some stylistic improvements and typos should still be reviewed and corrected. For instance:

"...ar considered..."
"...against an historical..."
"...Hence, it highly desireable,..."
"...range os sources..."
"capable of constrain the source..."
"it benefits for the fixed..."
"...at leas in..."

R:> Thank you for catching these typos. The suggestions were considered. However, we noticed that these were not all instances. Therefore, these and some other modifications in the text aimed at improving style and grammar were introduced. The reviewer is referred to the version with tracked changes for details.

C7

Please also note the supplement to this comment: