Dear reviewers:

We would like to thank the two reviewers for constructive comments and suggestions, which helped enhance the content of our manuscript. We have tried our best to carefully answer all your comments and questions. We provide response to the reviewers' comments in a question-answer format.

1. General

A novel tsunami arrival time detection system is proposed in the manuscript. The system consists of three separate modules: outlier detection algorithm, gap filling algorithm, and tsunami detection algorithm. System is tested on sea level time series measured at Ulleung-do (an island off shore of Korea) tide gauge instrument during March 2011. Described tsunami detection method is interesting and worthy of further development and testing. However, before manuscript can be published I have two major comments:

(1) Detection system is calibrated on too short (and too few) time series, namely only one month of sea level data including only one tsunami event (regardless of strength of this event), and this is simply not enough to test or calibrate such a system. Authors are aware of this short coming and say that it might be overcome by either using longer measured time series which include more tsunami events or by using synthesized time series. I believe authors should follow their own suggestion. Certainly, there are a number of tide gauge stations in the World which have longer records and have more tsunami events in them, perhaps even in a nearby Japan. Or if authors are not able to obtain this data, then they should do their tests on synthesized time series, following some of the papers they quote.

We agree to the reviewers' suggestion that the TADS should be inspected for a long dataset. We added a new section called 'performance of TADS' based on not only longer time series but also several synthesized time series. We increased the dataset by one year to investigate the performance of an outlier removal algorithm and a gap-filling algorithm. Also, we added 100 synthesized tsunamis to investigate the performance of a tsunami detection algorithm. Please see section 4 in the revised manuscript for detailed description of the performance of TADS.

(2) Manuscript is not very clearly written. With all the abbreviations, flow charts, tables without explanations, it is hard to understand some of suggested algorithms. Authors should make their text more clear. More specific comments on this will follow.

We agree to the reviewers' opinion. We reorganized the manuscript and expanded the analysis to highlight the performance of TADS (Tsunami Arrival time Detection System) as follows:

1 Introduction
2 TADS (Tsunami Arrival time Detection System)
  2.1 Outlier removal algorithm

2.1.1 Start mode
2.1.2 Keep mode
2.1.3 End mode
2.2 Gap filling algorithm
2.2.1 SGFA (Short Gap Filling Algorithm)
2.2.2 LGFA (Long Gap Filling Algorithm)
2.3 Tsunami detection algorithm
2.3.1 DART
2.3.2 SLOPE
2.3.3 TIDE
3 Calibration of TADS
3.1 Calibration of outlier removal algorithm
3.2 Calibration of gap filling algorithm
3.3 Calibration of tsunami detection algorithm
4 Performance of TADS
4.1 2011 Tohoku tsunami
4.2 Performance of outlier removal and gap filling algorithms
4.3 Performance of tsunami detection algorithm
5 Conclusion

The figures and tables were also reorganized accordingly.

2. Specific

(1) A number of terms are not explained when first introduced but only afterwards (and some never), e.g. event period is mentioned in the abstract, and several times after it, but not explained until results section.

We agree to the reviewers' opinion. We moved the purpose and the definition of the event period to section 2, where the event period mentioned for the first time. Also, we added the explanation of each term in the manuscript whenever it appears first.

(2) In introduction what are: short-time outliers, long-term outliers, short gaps, long gaps? I understood after reading the manuscript that these values (length of time steps) are later defined through calibration for specific herein described tsunami detection system. However, in Introduction, when dealing with values from literature, you should write a range of these values used by other authors. It would also be good to say that for this particular tsunami detection system, values will be determined through calibration.

We agree to the reviewers' opinion. The length of time steps is not an important issue in outlier removal but an important issue in gap-filling. Thus, we deleted ambiguous comments such as short-time outliers, long-term outliers in the introduction, whereas, we retained the long gaps that had an additional explanation which confined the range as follows:

“For long gaps that are expected to include complex patterns ~”

We added an additional explanation for short gaps to limit the range as follows:

“~ short gaps where a linear change is expected ~”

The criteria of short gaps and long gaps are introduced in section 2.2 in the revised manuscript. And how each criterion is determined is explained in section 3.2 in the revised manuscript.

(3) Also in Introduction, what are soft computing techniques?

Soft computing is a generic term for Fuzzy Logic, Machine Learning, Artificial Neuron Network, Chaos theory and so on. Please see the reference below.

https://en.wikipedia.org/wiki/Soft_computing

(4) Still in Introduction, there are two contradictory statements: (1) "A low probability exists for tsunamis to occur in the East Sea" (2) "This can be used to detect weak tsunami signals that are common in the Ulleung-do surge data" What is correct then? If there are more tsunami signals in the Ulleung-do surge data, why not incorporate longer time series with these signals into your analysis?

The first statement that says that there is a low probability of tsunami in the East Sea is correct. The word ″common″ in the second statement leads to misunderstanding. It was meant to explain not the frequency of tsunami but the characteristic of tsunami. Thus, we removed the second statement.

(5) Figure 1. I suggest adding bathymetry contours (perhaps coloring the figure?) and also pointing to Yamato rise mentioned in introduction.

We thank the reviewer for suggesting two points in Fig. 1. We changed the background of the figure to bathymetry contour and pointed out not only Yamato rise but also the location of 1983 and 1993 earthquakes which were mentioned in the introduction. Also, the locations of tide stations which were used in section 4 were added.

(6) Figure 2. Resolution should be increased. Why are sea levels showed with dots? I think it’s better to just use line. Also, it would be nice to add a zoomed in window showing Tokohu Earthquake period.

We kept the line with dots so as to show the time interval of sea level data. We think that it is better not to add zoom-in the shot of Tohoku Earthquake period because it already exists in Fig. 10 or Fig. 11 in the revised manuscript.

(7) In methods, you again refer to long and short gaps without defining them or saying that they would be defined later.

Please see the answer in (2).

(8) In general, idea behind your process should be more clearly presented. Why do you remove outliers and fill in gaps when these algorithms are not used during the event? I assume to be able to compare event period time series to time series from previous time steps - but then this should be clearly stated. Also, what exactly do you do with outliers, remove them and then fill the gaps? I believe so, but this is not clearly written.

Since tsunami detection algorithms require several hours of past data, outliers could cause false alarms and gaps could stop the tsunami detection algorithm even after the recording is restarted. We added the background and purpose of each algorithm not only in the abstract but also in the introduction to make it easy to understand the concept of each algorithm.

In order to state the meaning of each algorithm clearly, we changed the name of the outlier detection algorithm to outlier removal algorithm because the algorithm not only detects the outliers but also removes them.

(9) In 2.1. Outlier detection algorithm, entire chapter is highly difficult to follow. Try to simplify while still keeping the most important points. Likewise, Figure 4 is also very difficult to comprehend. I understand that it is a code flow chart - but perhaps here you could put a more simple version, and put this one (alongside with a code) to supplementary material? I.e. if your code is not described in some other paper. If yes, I do not see a need for a complicated figure and code.

We agree to the reviewers' opinion. We deleted the complex flow chart and added a simpler figure easy to understand the concept of the outlier removal algorithm. Since the new figure does not represent all the conditions of the algorithm, we listed all kinds of conditions in Table 1 in the revised manuscript.

(10) In 2.2. Gap-filling algorithm, I have similar comments as for 2.1., although it is written a bit more clearly, and figure is more understandable and helps follow text (but still not good enough!). There are also a number of abbreviations - so it is easy to get lost. Some of this abbrevations are, I believe, not explained: what is SW, what is EPFM, what is SWEP? what is h1, h2, h3, h4, t1, t2.

We agree to the reviewers' opinion. In order to explain the gap filling algorithm clearly, we divided the section into two sub-sections: 2.2.1 SGFA and 2.2.2 LGFA. Also, we simplified the previous complex schemes by explaining them though two figures, Fig. 5 and Fig. 6 in the revised manuscript. We added the full name of each abbreviation whenever it appears first.

(11) As I understand, you basically copy search data to gap window - but before that you make sure that you fit first and last point of search data to first-1 and last+1 point of the gap? If so, this can be clearly written.

The search window consists of search data and SW data (Please see the Fig. 6 in the revised manuscript). The search data are used to find the most suitable search window to fill the gaps.

And the SW data are fit to first-1 and last+1 point of the gap (which is called EP1 and EP2 in manuscript). Please see section 2.2.2 in the revised manuscript for detailed description of LGFA.

(12) In sentence "A predefined length of points (N_inter) from the last poing... to create the linear interpolation..." does this mean that you linearly interpolate data by using the least square method? Or something else?

Yes. We used predefined number of points before short gaps, N_inter, to linearly interpolate the data by using the least square method. We deleted ambiguous comments and corrected a text. Please see section 2.2.1 in the revised manuscript for detailed description of N_inter.

(13) In 2.3. Tsunami detection algorithm you present Table 1. This table is completely unclear. I guess that first two columns are related to DART, second two to SLOPE, and last two to TIDE algorithm. This is nowhere written (should be in the table caption). Some of abbreviations in Table 1. are defined in text but most of them not. What is tIS, tG, tGTide, tTide, and so on?... I assume some parameters related to DART, SLOPE and TIDE equations. But if so, these equations should be written and explained in the manuscript. Also, are all of these calibrated values or general values related to method or a mix?

We reorganized the table to make it easy to understand (Please see Table 3 in the revised manuscript). First, we categorized the parameters depending on the algorithm. Also, we added a description for each parameter.

(14) How is Tsunami Detection Index divided into five levels if you have only three tsunami detection algorithms? What are these five levels?

There are a total of 4 thresholds: TH_{DART} for DART, TH_{IS} and TH_{CF} for SLOPE, TH_{TIDE} for TIDE. If any kind of threshold is triggered, the TDI (Tsunami Detection Index) increments by one. Thus, The TDI ranges from zero (all thresholds are not triggered) to four (all thresholds are triggered) which means that the TDI is divided into five levels. Please see section 2.3 in the revised manuscript for detailed description of the tsunami detection index.

(15) Figure 6. is also really difficult to follow. I'd say if all of algorithm you use (including outlier and gap filling algorithm) are from previous papers, there is no need in including such a complicated version of Figures 4, 5 and 6. Something simpler would do, or even omitting figures.

Since the algorithms are newly developed or modified to be applicable to the Ulleung-do surge gauge, we should include figures which explain the algorithm. Therefore, instead of deleting the figures, the complex figures have been modified completely to make it simple to understand the concept of each algorithm. Please see Figs. 4 – 7 in the revised manuscript.

(16) Related to Tables, none of them are very clear or fully explained. In Table 3, what does it mean that windowsize is 2 (two of what? Points, hours?...), or that npastdata is

100 (100 of what?).

We agree to the reviewers' opinion. We added a description for each parameter in tables (Please see Tables. 2 – 3 in the revised manuscript.). The window size is a factor that determines the size of the window and the npastdata is also a factor that determines the size of the m_{search}, the past dataset to find a suitable data to fill the gaps. Thus, both parameters are dimensionless. Please see section 2.2.2 in the revised manuscript for detailed description of both windowsize and npastdata.

(17) In Table 4, why is search window located 3-14 days before actual gap?

Since the search window moves back over the length of m_{search} from the EP1, the final selected search window could be located in several days before gaps started. Please see section 2.2.2 in the revised manuscript for detailed description of how the search window is selected.

(18) In Results, you say that yellow alarm is triggered outside of the event period. How is this alarm triggered if you are not in the event period? And thus (from Figure 1) no tsunami detection algorithm should be activated?

The tsunami could not be alarmed outside the event period. We corrected the figure (Fig 11 in the revised manuscript) and the text accordingly (section 3.3 in the revised manuscript).

(19) In Discussion, I believe your method which is triggered only when there is an event, would be extremely difficult to use during events which have not-easily detectable sources (like meteotsunamis, landslide, ...). You can elaborate further.

It is true that the TADS could not cope with events like meteo-tsunamis. We described the limitation of the event period in the performance section. Please see section 4.1 for detailed description of the limitation of an event period.

(20) Page 5. line 22. "start and end points" instead of "end points".

We changed the term 'end points (EP1 and EP2)' to 'EP1, EP2'. And we added an additional description as follows:

"EP1 is the end point just before where gaps start, while EP2 is the end point just after where gaps end."