Response to the comments (Referee 1)

(a) The sea level change may be associated with many factors such as ocean temperature (including SST), salinity, currents, and surface winds etc. Therefore, the multi-regression between SST and SL PCTs may not include all aspects of SL changes. I am wondering whether the reconstruction could further be improved if more physical variables are considered.

To apply multi-variables to current reconstruction scheme, there are several problems. First, when we applied multi-variable's PCTs as predictor, the over-fitting occurred because as the mode goes higher, the possibility of overfitting increases. Second, some data that has less relation with SLA ruined the right signals. Third, the reanalysis process can increase the uncertainties of reconstruction. Actually, we tried to reconstruct SLA using wind and SST data simultaneously, but the result is poorer than each individual reconstruction case. To solve these problems, we need to input lots of efforts and we thought that is beyond our study boundary. However, the reconstruction applying multi-variables is valuable topic for the future study.

(b) The SL reconstruction does not include TG observations, but have a clear improvement over a similar reconstruction that includes TG observations. I am wondering whether the SL reconstruction could further be improved if all available TG observations are included.

To include TG data, there are two problems. First, using TG data we cannot conduct CSEOF analysis because they have lots of discontinuous points and their spatial coverage are too poor. So, we cannot establish the proper regression relationship between TG and SST. Second, the TG data's quality is not good. The vertical land motions cannot be calibrated and a lot of Japanese TG could be suffered by earthquakes or volcanic activities. To use TG data, the additional researches are necessary to correct the vertical land motions.

(c) How to validate the SL reconstruction in the early period over 1900-30 when no TG observations are available. It might be a little risky to include the reconstruction in this period.

As we can see in Fig. 11, even we have TG data for vilification, prior to 1970, the agreement between MSLA from TG and ReSLA-KP is very poor. But we thought that it does not mean our reconstruction is not good because the TG data is not enough to verify. Even we cannot verify the reconstruction results, we think the result is still valuable.

(d) Writing and presentation may need improving. There are too many abbreviations such as SL, MSL, GMSL, SL-KP. For example, MSL and GMSL could be explained in figure captions. KP is unnecessary because the study focuses on KP region only.

We have modified the abbreviations. But we cannot omit every KP, because, after removing KP, we found the unnecessary global terms must be necessary.

(e) Figure captions should identify the data source and average region etc.

We have applied the commend.

Detailed comments

P1L11, revise: extend the spatial resolution ..into the past

We have revised.

P3L5, CSEOF is not defined

We have defined in Page 2.

P3L14, "KP" could be deleted throughout the manuscript since the study has been limited over the KP region anyway, which will greatly improve the readability. "KP" could be noted in the figure caption when necessary.
We have deleted 'KP' and add explanation about the default domain.

P3L21, revise: looking at the regional level will lead to
We have fixed.

P4L11, annual signal=> seasonal signal?
The terms, annual signal and seasonal signal, are having same meaning. But to prevent confusions, we change seasonal signal to annual signal.

P4L20, include data => included data?
We have fixed it.

P4L21, over => from?
We have fixed it.

P6L5-7, revise the sentence
We have fixed.

P6L11, delete "in this case", "really" P6L12, independent of => independent from?
We have erased this sentence.

P9L3, How does "summing" actually do, arithmetic or squareroot?
We mean Root Sum Square. And this part have deleted.

P10L10, this is an indication that SL is not merely dependent on SST.
We have deleted this part. And we gave up to explain the physical reasons for the extreme SLR values.

P11L22, delete "then"
We have deleted it.

P11L25, delete "cases of"
We have deleted them.

P11L28-29, delete "considering the available number of TG data"
We have deleted them.

P12L4, It is not clear how MSLA-KP is defined (assuming every ocean grid in reconstruction). How MSLA-KP can be compared with TG-KP (only in TG grids).
We have added more explanations as follow.
To check the reconstruction results, we calculated MSLA of TG-KP, ReSLA-H, and ReSLA-KP. Spatial mean was calculated for the two grid datasets. For TG-KPs, we calculated mean differences between each time steps and we integrated the differences. The integrated mean differences became the MSLA of TG-KP.

P12L12, revise "was edited to have the same time span data gaps"
We have erased 'data gaps'.

P12L14-15, revise the sentence: ReSLA-KP show a better agreement of AVISO-KP than ReSLAH.
We have revised.

P12L17-18, how many modes are used in Hamlington?
We have added detailed number.
Hamlington et al. (2011) used a limited number (< 90% of total variance) of CSEOF modes to avoid over-fitting issues, but in this study, nineteen CSEOF modes are used which explain 98% of total variance of SLA-KP.

P12L23, thousand => a thousand
We have corrected.

P13L17-18, authors should extend the conclusion of a better current SL reconstruction. there is no way from Figures 16-17 to tell the current study is better.
We have deleted this part.

It is not clear in Figure 13 either. It may be necessary to point to Figure 14a. A better way is to calculate the RMSE.
We have added more figure.

Fig. 1, digital quality should be improved.
I think it has a high resolution, 600 ppi.

Fig. 2, coastal line should be consistent with those in other figures.
We have changed the figure.

Fig. 3, I could caption the figure as "Mean SLA in KP (gray) and global (black) regions from AVISO" so that I can get rid of some abbreviations.
We have changed the caption and figure.

Fig. 4, add "AVISO" in caption
We have modified.

Fig. 5, add "AVISO" in caption
We have modified caption.

Fig. 6, revise: trends (shapes) and correlation (color), change the red color of triangle into black so that the color will not be confused with correlation.
We have modified the caption and figure.

Fig. 7, NRMSE, I don't know the advantage of using normalized RMSE instead of RMSE.
'NRMSE' and 'RMSE' very similar, but when NRMSE has 'zero' value this means the regression is same with some constant value cases and if the value are negative that the compared data is less agreed its mean value. So I NRMSE gives some intuitive interpretation.

Figs. 8-9, I am confused how the 3-month averaged mode is plotted. I assume there is only one CSEOF associated with one PCT for a particular mode.
Yes, you are right. But the evolution is small through the months, therefore we represent the results as seasonal mean values to save some space.

We have modified the caption.

We have modified the caption.

Because we have 6 cases but ReSLA-H is just one case.

I think the color problem is related to the resolution, I provide 600 ppi image and that figure has no problem to recognize yellows.
And the time period is 1970-2008 for the both cases.

Yes, that's right. Nevertheless, ReSLA-H has very poor agreement. We just want to show the limit of global reconstruction, as you can see the below figure. Over 1993-2015, the correlation coefficient is values are pretty high, this means if we applied Hamlington et al. (2011)'s method in local scale, the correlation coefficients must be higher than current Fig.
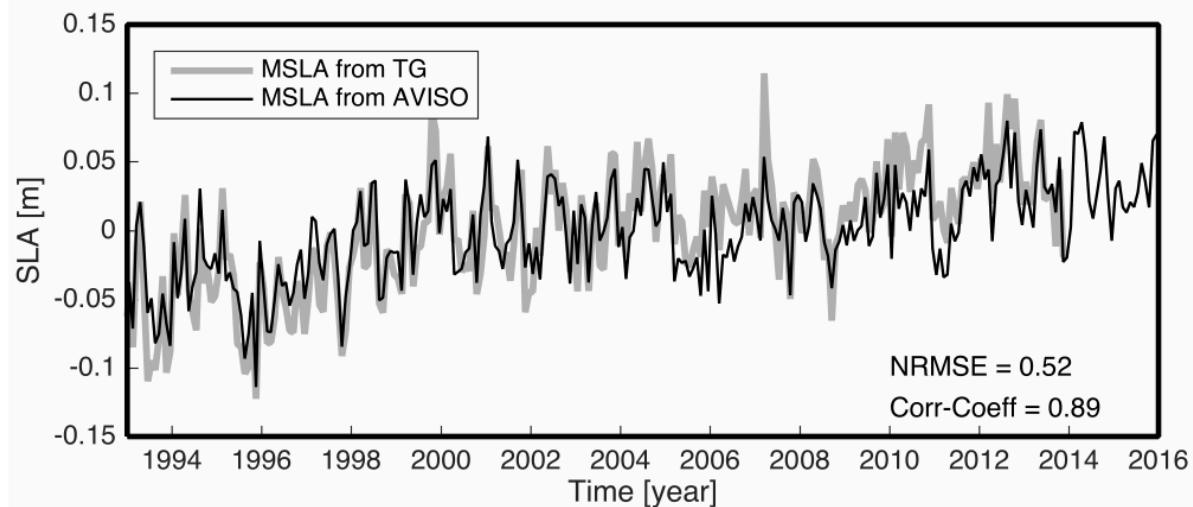
Response to the comments (Referee 2)

We admitted the title was not proper and I omitted the 'projection' from the title.

Actually, if we can secure the reasonable number of TG data, the reconstruction using TG data is the best case. Unfortunately, TG data around the KP is less than 10 prior to mid 1960s. Therefore, we used SST data instead of TG data. And to supplement the sparse observations, we made several cases: different datasets, different areas.

I still believe the wind stress and local surface currents are dynamically important for sea level variations, like many studies have shown. There is no direct link between coastal sea level and SST in open ocean. How possible to include other dynamical factors?
SST and sea level has strong relationship when we analyze both of data through CSEOF analysis (Hamlington et. al., 2011, 2012a). Actually, many CSEOF modes shows great agreement, e.g., ENSO and PDO (see, below figures. LHS and RHS are ENSO modes of AVISO and OISST, respectively; c.f. we just introduce these to show the example of their relationship)
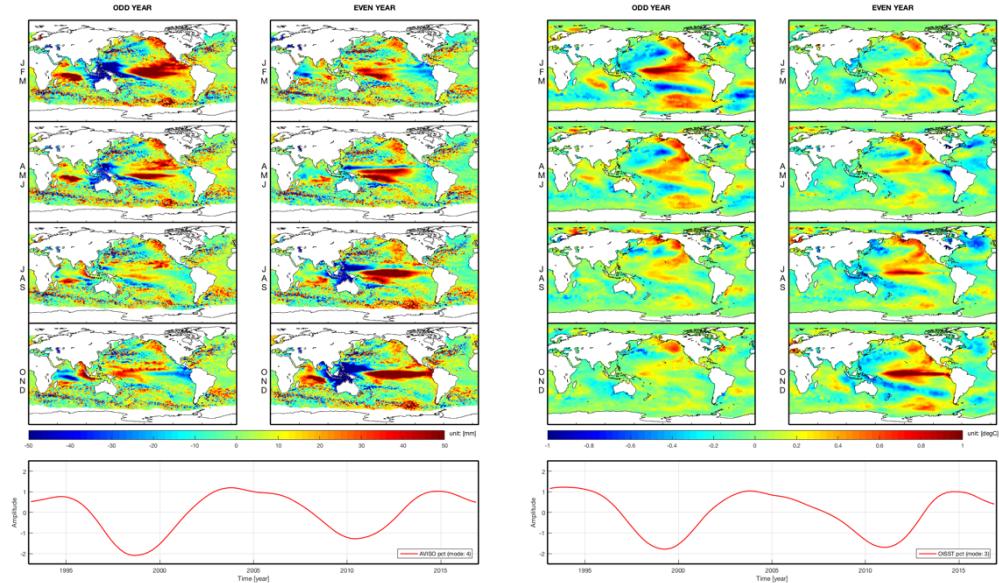
As explained in 2.2.1 and 2.2.2, we tried to find the lagged relationship between PCts (SST and sea level). Actually, only a few modes can be interpreted. Only we can understand that each mode is mathematically orthogonal. This means these modes can be the best prediction variables. Therefore, even we cannot explain the exact physical background of each modes and their relationship, the reconstruction results are still valuable. And this kind of situation is same with other reconstruction studies.
And I had a reconstruction case that used wind data, but the result is not good as much as SST. (see below figure). Because we cannot introduce every result and some cases were skipped.
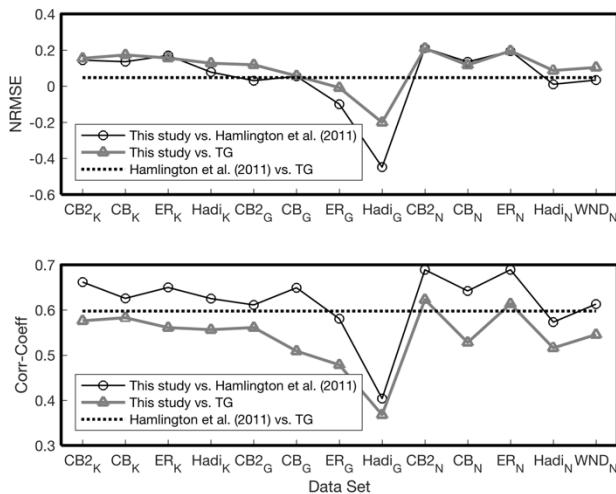
3. Section 2: this part reads loosen and tediously long, and many parts are unnecessarily mentioned with many times. I would suggest shortening this section with concise contents to avoid readers losing interests.

We have trimmed out lots of repeating parts in sections.

4. Section 3: This section again is not properly presented.

We have modified this section.

4.1 Essential questions: 3.1: I do not think the following key question is answered. 'To reason whether the extreme trends patterns was related to the local mass distribution caused by various sources such as vortex and river discharge or was an independent. . ..' The extreme trends on China coasts in Fig 2 are proposed as a result of increasing river discharge by the authors. However, there is no convincing evidence supporting this. (one would not expect that river discharge can cause sea level increase on north Chinese coasts, because it is drying over recent years in this region). Same for the ocean current impacts. Can authors provide evidences supporting this (P10 Lines 28-31)?

We agree with your comment. There are no studies that explain the relation between the sea level rise trend and ocean current (or river discharge). To explain the relationship, we need more research that is beyond this paper's boundary. So, we removed this part.

Also, I cannot see any point of separating the regions with local correlations </>0.5. Because the two regions are both located in Yellow Sea and Japan Sea, the regional averages are supposed to not contain local information, and they instead reflect the large-scale variations. This might be reason why the two series in Fig5 are always highly correlated.

This study's basis assumption is the SLA-KP can be represent as the difference with the GMSL. And we worried about the extreme trend zones because if the extreme zones had significant differences with the other zone than the separated reconstructions were necessary. So to pull out the extreme zones we calculated the averaged correlation coefficient.

For the correlation map e.g. Fig4 (and Fig 6), is the annual cycle removed? Removing the seasonal cycle is critical. Otherwise, they are always statistically correlated but it does not make any sense. Need to clarify.

Yes, we removed the annual cycle. We have clarified it on the figure captions.

How can the sea level records between TG and AVISO be correlated e.g. Fig 6 when also having linear trends? If linear trends exist, they are always correlated. Correlation is for assessing the similarity between detrended variability/anomalies but cannot be used for assessing the trends. The basic concept I think is wrong.

We removed linear trends during the calculation of correlation coefficient. In my opinion, for sea level data the linear trend is very small variance than the data fluctuation so the trend has very less effect on the correlation coefficient values (see below figure). I checked the correlation coefficients of Fig 6 after removing linear trend of each time series. But it only made less than 1% changes.
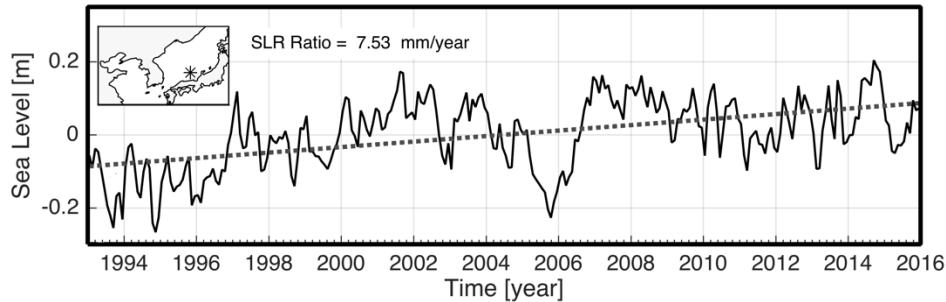
The trends and correlation coefficients in Fig. 6, actually, calculated separately, there were several figures before. To reduce the figures, we combined the information and put the information in one figure. See the below figures.
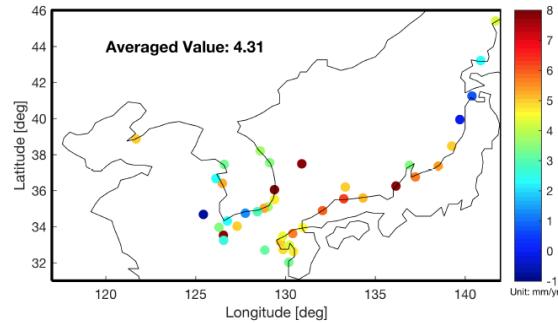
Figure 3. SLA linear trends over 1993-2013 using TGs_KP
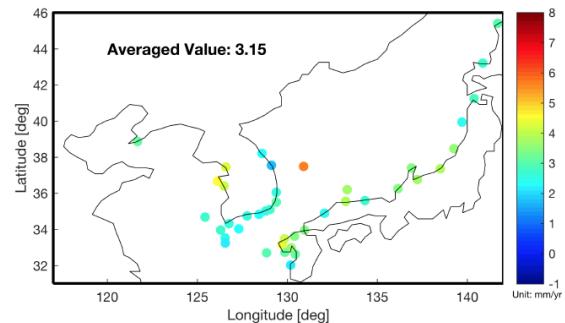

Figure 4. SLA linear trends over 1993-2013 using AVISO_sla
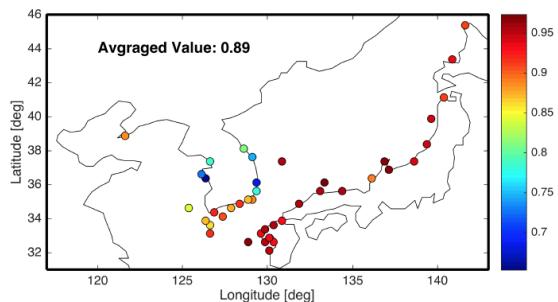
Figure 6. Correlation Coefficients between TGs_KP and
AVISO_sla points with seasonal signals

Please clarify. Fig 6 & 7: how far are the AVISO sites from TG stations?
I clarified the maximum distance (about 12 km) at the caption.
In Fig. 7, but, to calculate MSLA of AVISO, I just used entire area not the closest point.

Fig8 & 9: I cannot see there is a trend in the PC series of Fig9.
We have changed the Fig.8 and 9 to show bigger PCT. We think you can see the trend.

What are the trend value and its significance level?
To help readers' understanding, we have added one more figure. Fig 12 shows the linear trend values of each mode and their confidence intervals.

Does it agree with the values based on the local estimation i.e. Fig 3.
Fig 3 doesn't have annual signal. If you see Fig 11, the wiggled signals can achieve by the summation of each CSEOF modes. So, our answer is Yes it does, it agree with MSLA-KP in terms of low-frequency signal.

Because there is no annual cycle signal in Fig9, there is no need of presenting it with 4 seasons.
The spatial pattern of one CSEOF analysis is not a single map, so they need to represent through their nested period. Actually, most of the CSEOF mode do not have similar spatial patterns though the time evolution. The reason why we can determine the 2nd mode as the trend mode, is these spatial patterns are pretty similar though the nested period.

4. 2 Section 3.2: what are the reasons for COBESST2-NWP having best correlations with sea level? Do author have interpretations?
We though the COBESST-2 data was made by Japanese researchers, and this means there is high possibility that the calibration can be focused on their near boundaries. So, this datasets have more accurate results for NWP area. But there are no similar studies before we cannot support this interpretations with reference. We just provide the best cases.

Why does not the local SST do better job than others?
We thought that SLA-KP is influenced by ocean current too. But the small domain's SST data is enough to interpret this ocean current effect. And the global SST contains too much information.

Also, the short names e.g ReSLA-NWP are not used in figures, which however use the long name. Authors need to be careful for the presentation throughout the whole paper.
I have checked the shorten variables and corrected them.

Again in Fig 14 & 15, are the linear trend and annual cycle both removed before calculating correlation?
Yes, we removed trend before calculating correlation coefficient.

Are the trends in Fig 14b statistically significant?
Most of the trends are statistically significant. But some of them are not. We have added the p-test result.
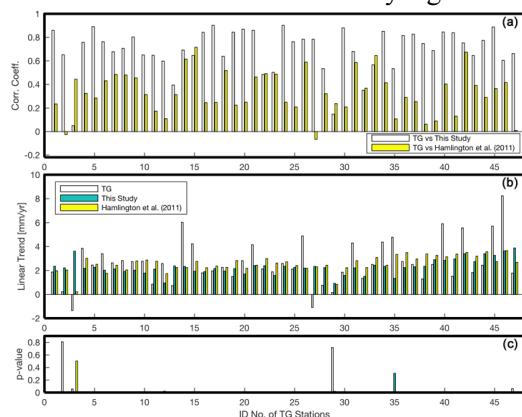
'these detailed fluctuations are closer to the actual sea level variability': what is the actually sea level variability?
'the actual sea level variability' means AVISO-KP. We have changed this.

Authors seem to insist that the SST-based reconstruction shows better results. What are the reasons for that?
As we mentioned in the paper, the other reconstructions were not focused on the SLA-KP and they didn't used entire decomposed mode for the reconstruction process. This means that the global reconstruction has high possibility to omit some important modes in certain local scale reconstruction. And also, for we applied lagged regression, we can include the lagged relationship between each basis functions.

In the marginal seas of NWP, many studies have shown that the local ocean surface currents and wind tress determine the sea level, and the open ocean in far-field has less impacts. However, this paper finds the (far-field) NWP SST can 'statistically' better capture the sea level in marginal seas of NWP i.e. KP. What is the science behind it? Please keep in mind that the sea level variations between the two sides of western boundary currents (Kuroshio/Oyashio) are very differently forced e.g. by the thermalsteric height and open ocean currents via geostrophic balance and by local wind/surface currents.
This reconstruction was conducted by extending the PCTs of AVISO-KP CSEOF. And, we cannot explain most of the CSEOF modes. The other reconstructions have the same problem. At beginning stage, we tried to understand the background mechanism and the relation between the factors which you mentioned above. But we figured out that our trials were beyond our research boundary. We thought that SLA-KP related with many factors: ocean currents, thermal expansion, global sea level rise, wind and so on, and SST-KP is not big enough to cover these factors. Even though SST-KP contains every effect, but the problem is whether CSEOF can decompose these factors well. And global SST contained too much information which can lead a over-fitting issue.

More essentially, this paper is focusing on 'reconstruction capability', but it spent a lot space in section 3.1 comparing TG and AVISO. Authors should work properly to make the presentation and structure of this paper concise and focused.
We have omitted unnecessary parts.

Conclusion: What is the linear trend map of reconstructed SLA-KP over satellite era? Are they comparable with Fig 2?
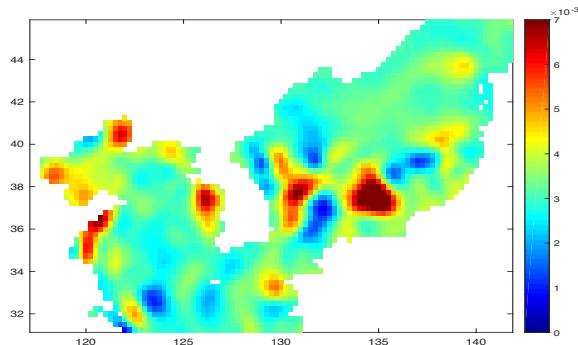
We calculated the linear trend map of ReSLA-KP over the satellite era (1993-2014). Even a time period is not exactly same but the result agrees with the AVISO-KP's trend map.

How are the SST variations looking like over this region/NWP? Does SST follow the sea level changes very well?

I don't know how I can explain the SST variations of SST-NWP. But, as I showed you above, the relationship between SST and SLA is close. And we want to explain our reconstruction process again. We used SST's CSEOF PCTs as predictors for the multivariable regression. And using the past SST PCT and the regression relation we extended the AVISO-KP's PCTs to the past. Therefore, the evolutions between SLA and SST are not necessary to be similar.
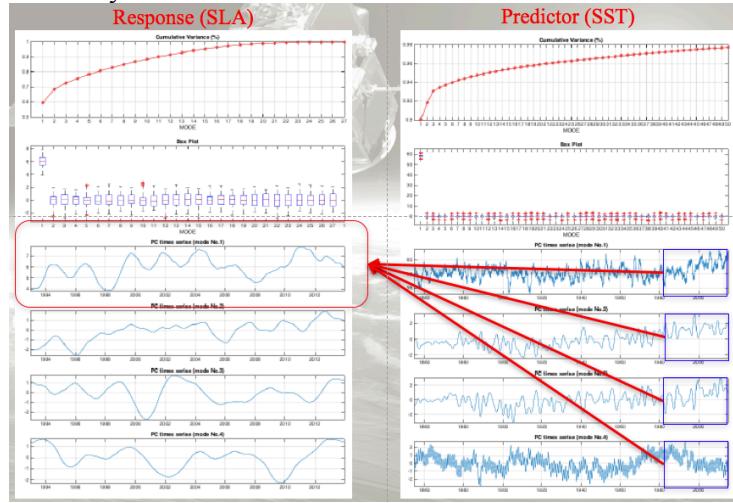
Minor comments

P2, lines 6-7: do not understand. What does bias mean?
It means that most of the TG stations are located on the Northern Hemisphere.

Also, references are needed to support this statement. P2, lines 29-31: references?
Instead of reference, we have added global linear trend map.

P2 lines: 31-32: do not understand. P2 line 33: this needs to reword
We rewrite this part.

P3 lines 14-15: what reconstructions?
Global reconstructions by Hamlington et al (2011) and Church and White (2011)
We have rewrite this part.

P3 lines 11-22: references?
We have put the reference.

P3 lines 11-27: the focus/motivations are loosen and not concise.
We have rewritten this part.

P4 lines 9-10: do not understand
We have erased this sentence.

Figure 1 is not readable Figure 1 seems to have 3 TGs on China coasts, while there is only one appearing in Figure 3. Any flags applied?
The two TG's time spans did not cover 1993-2014.

Response to the comments (Topic editor)

Topic Editor Decision: Reconsider after major revisions (14 Nov 2017) by John M. Huthnance
Comments to the Author:
Dear Authors
Thank-you for your revised manuscript. In due course I shall be sending it back to both referees who asked to see it again after "major revision". However, there are a few things which I would ask you to consider before that.

You have responded to (both) referees' question about using SST rather than other variables (e.g. wind, runoff, tide-gauge records). However, I think that a few sentences on this question should be included in the final manuscript so that it is "self-contained" and eventual readers do not have to search the discussion to find answers.
We inserted follows.
One of the unique characteristics of the current study is that we only used SST as a proxy of former SLA; other studies, however, used TG data or combined data (TG and SST). There are multiple reasons why we chose not to use TG data for the current reconstruction. The first reason is due to both the poor data coverage and the poor data quality. There are relatively few tide gauges extending into the past in our study area, and even fewer that are of high quality (i.e., unaffected by vertical land motion, with few gaps, free of non-physical jumps). The second reason, and related to the first, is that due to a methodological characteristic of the CSEOF analysis, a dataset that is free of gaps (temporally continuous) is needed. To satisfy this requirement, we are led to other gridded reconstruction or reanalysis products. There are many types of data that could potentially be used in our scheme (e.g. wind, ocean current, precipitation, atmospheric pressure). We used only SST for the following reasons. 1) SST and SLA have a distinct relationship when we analyze both of data through CSEOF analysis (Hamlington et al., 2011; Hamlington et al., 2012a; Hamlington et al., 2016) and Hamlington et al. (2012b) showed that SST could be a good proxy of SLA in this part of the ocean. 2) Limiting our analysis to SST reduces the possibility of overfitting in the regression scheme we use to reconstruct. As a final benefit of using SST, we can check against the available tide gauge data to provide an independent comparison to our reconstruction.

I also wonder whether referee 1 will be happy with the large number of remaining abbreviations.
We reduced abbreviations.

Here are a few more editorial-type details.
Page 3 lines 16-17. This sentence about the number of early tide gauges is not clear; are you saying that the first was in 1930 and this was the only one until 1950?

Original	Changed
Second, the temporal coverage of the TG around the KP (TG-KP) started around 1930 when the only TG had been available by 1950;	Second, the temporal coverage of the TG around the KP (TG-KP) started around 1930 and only one TG was available until 1950;

Page 3 line 19. Better ". . is proposing for the KP a new scheme . ."
Sorry, we want to keep the original sentence because our scheme can be applied for the other regions.

Page 6 line after (4). ". . and the ε is random error. . ."
We have corrected it.

Page 8 line 2 Not "boundary". Maybe simply ". . two years maximum lag. Using . ."
We have fixed it.

Page 8 line 27. Omit "of the linear trend"?
We have fixed it.

Page 8 line 29 "separate" (spelling)
We have fixed it.
Page 9 lines 10-11. "five TGs showed acceptable accuracies". This seems to assume that all error is attributable to TGs. But the TG provide real and important data in the right place. It is the AVISO data that is in the wrong place.
We agree with your comment and we modified our expressions little bit.
The comparison showed that only five TGs having less than 30\% of differences with the AVISO-KP's linear trend. Eleven TGs showed more than 30\% of underestimation and twenty-one TGs had more than 30\% of over estimation. While there was disagreement between TG locations and AVISO grids, over than thirty percent of differences were significant.

Page 9 line 30 What are the "Six" reconstructions – please say clearly what they are.
We explained the six reconstructions in Sec. 2.1.2 and Figure 14. We have fixed little bit.
We made six reconstructions (Sec. 2.1.2 and Fig. 14), and the mean SLAs of six reconstructions showed a reasonable agreement with the mean SLA of TG-KP over 1965-2014.

Page 9 line 32. ". . there were only a few . ."
We have fixed.

Page 10 lines 19-24 (Monte-Carlo) This is much the same as page 8 lines 5-9 except including sea-level trend. Probably these lines could be merged in one place.
We have fixed.

Figure 2 caption. "gauge" (spelling, twice)
We have fixed.

Figure 3, 4, 5, 7, 8, 15 captions. Replace "w/o" – "without"?
We have fixed.

Figure 9 caption. "Cumulative variance of CSEOF modes . ."
We have fixed.

Reconstruction ~~and Projection~~ of Sea Level Around the Korean Peninsula Using Cyclostationary Empirical Orthogonal Functions

Se-Hyeon Cheon[1], Benjamin D. Hamlington[1], and Kyung-Duck Suh[2]

1Department of Ocean, Earth and Atmospheric Sciences, Old Dominion University, Norfolk, VA 23529, USA

2Institute of Construction and Environmental Engineering, Seoul National University, Seoul 08826, Republic of Korea

Correspondence to: Benjamin D. Hamlington (bhamling@odu.edu)

Abstract. Since the advent of the modern satellite altimeter era, the understanding of the sea level has increased dramatically. The satellite altimeter record, however, dates back only to the 1990s. The tide gauge record, on the other hand, extends through the 20th century, but with poor spatial coverage when compared to the satellites. Many studies have been conducted to ~~extend the spatial resolution~~ create a dataset with the spatial coverage of the satellite ~~data into the past~~ datasets and the temporal length of the tide gauge records by finding novel ways to combine the satellite data and tide gauge data in what ~~are~~ is known as sea level ~~reconstructions~~reconstruction. However, most of the reconstructions of sea level were conducted on a global scale, leading to reduced accuracy on regional levels, ~~particularly where~~ especially when there are relatively few tide gauges. The sea around the Korean Peninsula is one such area with few tide gauges ~~prior to~~ before 1960. In this study, new methods are proposed to reconstruct ~~the~~ past sea level ~~and project the future sea level~~ around the Korean Peninsula. Using spatial patterns obtained from a ~~cyclo-stationary~~ cyclostationary empirical orthogonal function decomposition of satellite data, we reconstruct sea level over the ~~time~~ period from 1900 to 2014. Sea surface temperature data and altimeter data are used simultaneously in the reconstruction process, leading to an elimination of reliance on tide gauge data. Although we did not use the tide gauge data ~~was not used~~ in the reconstruction process, the reconstructed results showed better agreement with the tide gauge observations in the region than previous studies that incorporated the tide gauge data. This study demonstrates a reconstruction technique that can ~~be used on~~ potentially be used at regional levels, with particular emphasis on areas with poor tide gauge coverage.

1   Introduction

Although sea level rise is a global phenomenon, the impacts are local ~~,~~ and are happening now. Changes in sea level are ~~impacting~~ affecting communities across the globe on an almost daily basis through increased erosion, greater saltwater intrusion, more frequent ~~?nuisance ?~~ nuisance flooding, and higher storm surge causing severe damages on the coastal structures (e.g., **???**). Planning for, adapting to, and mitigating current and future sea level has necessarily begun in many threatened areas. Expensive decisions - both in economic and societal terms - are already being made. Examples can be found throughout the world, with coastal communities making difficult decisions on how to address concerns associated with future sea level rise (e.g., **?**). The present and near-term threat of sea level rise across the globe ~~and the subsequent decisions to address the~~

~~problem highlight the~~ highlights the immediate need for actionable regional sea level projections~~across a range of actionable timescales~~. In order to improve future projections of sea level, understanding past sea level change is an important first step.

Before the satellite altimeter era, the only available sea level observations came from tide gauge (after this TG) records. The ~~TG data provide~~ TGs provide the records of local sea level variations, covering ~~a time period of~~ nearly two hundred years in some locations around the globe. Using TG data, scientists ~~can study~~ have studied past sea level changes ~~at specific locations across the globe~~ around the world. However, TGs ~~do not provide good global~~ provide poor spatial coverage as they are necessarily only located at coastal sites and ~~have a bias~~ weighted towards the Northern Hemisphere. ~~Satellite altimeters, on~~ On the other hand, ~~have been~~ the satellite altimeters collecting data since ~~1992. The satellite altimetry data~~ 1992 have near-global coverage of sea level but a relatively short observation period compared to TG observations, which is a severe handicap to analyzing long-term changes in sea level. This disadvantage is particularly true given the presence of sea level variability with decadal and longer timescales.

? ~~attempted~~ was one of the first to reconstruct sea level anomalies (SLA) by combining TG data and satellite altimeter data. In their research, they studied low-frequency variability in global mean sea level (~~or global mean sea level anomaly; hereafter GMSLor GMSLA~~ GMSL) from 1950 to 2000. They interpolated sparse TG data into a global gridded SLA pattern applying EOFs (Empirical Orthogonal Functions) of SLA using data from the TOPEX/Poseidon satellite altimeter to capture the interannual-scale signals, e.g., ENSO (El Nino-Southern Oscillation~~(hereafter ENSO) and the~~) and PDO (Pacific Decadal Oscillation~~(from now on PDO)~~). Building on previous studies (**???**), **?** created a reconstruction from 1950 to 2001 using EOFs of SLA data measured from satellite ~~altimeter~~ altimeters and a reduced space optimal interpolation scheme. This research was subsequently updated to increase temporal coverage from 1870 to the present (**??**) and the reconstructions have been made available to the public through the website (http://www.cmar.csiro.au/sealevel/sl_data_cmar.html). In these studies, GMSL was found to rise approximately 210 mm from 1880 to 2009, with a linear trend from 1900 to 2009 of 1.7 ± 0.2 mm~~per year~~/yr. The resulting SLA is one of the most comprehensive and widely cited reconstructions. While these studies focused largely on the reconstruction of GMSL, **?** applied cyclostationary empirical orthogonal functions (CSEOF) as basis functions for the reconstruction of SLA in an attempt to improve the representation of variability about the long-term trends. This approach ~~has been shown to provide~~ provided an advantage for describing local variations such as ENSO and PDO. After that, **?** proposed an improved scheme of their reconstruction using sea surface temperature (hereafter SST). Given the limited TG data in the past, ~~reconstructions of sea level~~ the reconstruction of SLA relying only on TGs ~~are poor~~ were inaccurate, particularly before 1950. Leveraging other ocean observations (e.g.~~SST) to reconstruct sea level leads to an improved sea level~~, SST) led to an improved SLA reconstruction further into the past. In addition, this approach provides an advantage for describing local variations such as ENSO and PDO because the SST data gave an information on deep oceans where only few TGs are available.

While sea level is a global phenomenon, the extent of sea level change ~~can vary~~ varies dramatically across the globe. During the ~~24-year~~ 24-years satellite altimeter record, regional trends have been measured to be four times greater than the global average in some areas ~~. Properly planning for future sea level change requires an assessment of sea level on local or regional levels, as future sea level for one location could be quite different than~~ (Fig. 1). Therefore, sea level assessment on a regional level is necessary to plan for future sea level ~~in another location. Rather than using a global reconstruction, several studies have~~

~~instead~~ accurately. Several studies focused on regional reconstructions ~~of sea level, targeting a specific area of focus~~ targeting a particular area of interest. As an example, using an optimal interpolation method, **?** reconstructed the distribution of SLA in the Mediterranean Sea over 1945-2000. They used EOFs of satellite altimeter data spanning from 1993 to 2005 as basis functions and interpolated the TG data using these spatial patterns. A spatial distribution of sea level rise trends for the Mediterranean for the period of 1945-2000 indicated a positive trend in most areas. **?** performed a regional ~~sea level reconstruction based on the scheme applying CSEOFs~~ SLA reconstruction using CSEOFs as basis functions (**?**) with a domain covering only the Pacific Ocean. They found that a choice of basis functions had a significant effect on the spatial pattern of the sea level rise and the ability to capture internal variability signals. Global basis functions, either ~~CSEOF or EOF~~ CSEOFs or EOFs, are typically dominated by large-scale variability in the Pacific Ocean associated with ENSO or the PDO. As a result, global reconstructions are poorer in some ocean basins (e.g., Indian Ocean, Atlantic Ocean) than others (Pacific Ocean). This issue is likely exacerbated even further when looking at even smaller regions.

In this paper, we focus on one such region: the Korean Peninsula ~~where over seventy-five million people live~~. In South Korea, over twenty-seven percent of its forty-five million people live in coastal city areas, and nearly thirty-six percent of ~~GRDP (~~Gross Regional Domestic Product ~~)~~ is produced by coastal city regions (**?**). As a result, policymakers have a keen interest in a sea level rise around the Korean Peninsula (hereafter KP; a suffix, '-KP' means the spatial domain of the data or variable is around the Korean Peninsula) to establish proper remedies to ~~future~~ sea level rise. ~~Using global reconstructions around the KP,~~ Studying SLA-KP, researchers have primarily relied upon globally reconstructed SLAs (**??**). However, extracting SLA-KP (or more generally any small ~~region, is a problem~~ regions) from a grobally reconstructed SLA have some problems. First, global scale reconstructions use a limited number of basis functions to prevent the interpolation from over-fitting and creating spurious sea level fluctuations. There is a difference between the ~~major modes for~~ dominant modes of variability at the global scale and ~~the major modes for~~ local scale; e.g., there is a high possibility that the globally selected basis functions, which represent 90 % of the total variance in the global level ~~, for example,~~ will not represent 90 % of the total variance in local scale. Second, the ~~temporal~~ coverage of the TG ~~data~~ around the KP (TG-KP) started around 1930 ~~when less than 10 TGs were available~~ and only one TG was avaiable until 1950; it is too little to secure accuracy on these local scales. As mentioned above, ~~TG~~ TG-KP coverage is poor extending back into the 20th century, and ~~looking at the regional level will lead to relatively few gauges to analyse in most areas .~~ relatively few TGs are available to analyze in some areas (Fig. 2). Hence, the goal of this study is ~~proposing~~ to propose a new scheme that builds off of **?** that applies CSEOFs to reconstruct local SLA where the TG data is not enough to ensure a quality of reconstruction through the 20th century. We focus on the KP both due to its exposure to risk from impending sea level rise and also as a test case to demonstrate how this technique could be applied at other locations across the globe. ~~In brief, the primary goals of this study can be summarized as follows: 1) Broaden our understanding of the SLA around the KP both in the past and present and 2) Suggest a new reconstruction scheme for local areas where have insufficient tide gauge coverage in spatial and temporal domain.~~

2 Data and Methods

2.1 Data

2.1.1 Sea level ~~anomaly~~anomalies

The basis functions ~~for this reconstruction~~ of this study's reconstructions are the CSEOFs ~~monthly mean gridded SLA covering the time period from 1993 to present. This monthly data has a $0.25° \times 0.25°$ grid resolution and it is available via the~~ of a gridded satellite data of SLA provided by AVISO (the Archiving, Validation, and Interpretation of Satellite Oceanographic~~); this data opens in public (~~; ftp://ftp.aviso.altimetry.fr/global/delayed-time/grids/climatology/monthly_mean/) from 1993 to 2015. This monthly data has a $0.25° \times 0.25°$ resolution and hereafter this ~~data set~~dataset is written as ~~AVISO-SLA. The data is based on satellite altimeter measurements over 1993-2015; Topex/Poseidon, ERS-1&2, Geosat Follow-On, Envisat, Jason-1&2, and OSTM satellites collected the SLA. The delayed time Ssalto/DUACS multi-mission altimeter data processing system has created this product.~~ AVISO. Before conducting the CSEOF decomposition, mean values for each grid point were removed to center the data. The annual signal has not been removed as it is accounted for by the CSEOF analysis (see more details in section 2.2.1~~below~~). The data was trimmed to contain only the ~~ocean~~ seas around the KP (31°-46°N and 117°-142°E~~; hereafter AVISO-KP~~) and it was multiplied by the square root of the cosine of latitude to consider the actual area of each grid.

2.1.2 Sea surface temperature

In this study, two SST reconstruction ~~data sets~~datasets were used: ERSST ~~(Extended Reconstructed Sea Surface Temperature) (???)~~ (Extended Reconstructed Sea Surface Temperature; ???) and COBESST2 (Centennial in situ Observation-Based Estimates; ?). The ERSST dataset is a global monthly SST dataset based on the observation of ICOADS (International Comprehensive Ocean-Atmosphere Dataset). This monthly analysis has a $2° \times 2°$ grid resolution and its time coverage is from 1854 to the present, and the included data are anomalies based on a monthly climatology computed from 1971-2000. The COBESST2 dataset is a monthly interpolated $1° \times 1°$ SST product ~~over~~ from 1850 to the present. It integrates several SST observations: ICOADS 2.5, satellite SST, and satellite sea ice. ~~The bucket correction process was applied to the data up to 1941.~~ In addition to OI (Optimal Interpolation) scheme, this ~~data set~~dataset used an EOF reconstruction.

Each data was trimmed as three different regions: a global domain (no trim), the Northwest Pacific ~~(NWP)~~ domain (25°-55°N and 110°-160°E), and around the KP area~~; to indicate the domains of dataset we put '-NWP' and '-KP' behind the name of dataset~~. Before conducting the CSEOF decomposition, these ~~data sets~~datasets were treated as follows. 1) The mean values for each grid point were removed~~to prevent those values to have a significant influence on CSEOFs~~. 2) The data were weighted by the square root of the cosine of latitude to consider the actual area of each grid. 3) Any grid points that were not continuous in time were removed. Like the satellite altimeter dataset, ~~the annual signal of SST data~~ an annual signal was not removed.

4

2.1.3 Tide gauge data

Monthly mean ~~sea level~~ records of 47 TGs ~~for the KP~~ were obtained from the Permanent Service for Mean Sea Level (~~hereafter PSMSL, see~~ PSMSL, Fig. 2) ~~over 1930-2013. The Revised Local Reference (RLR) data were selected; the RLR data are measured sea levels at each site about a constant local datum over the complete record. The~~ from 1930 to 2013. The earliest
data of ~~TG-KP~~ the TGs is traced back to 1930 at Wajima Station (~~see~~ Fig. 2). Before 1965, the number of available TG datasets is fewer than 10, with only one TG (Wajima Station) providing data ~~before~~ unitl 1950.

    An ongoing GIA (~~glacial isostatic adjustment~~ Glacial Isostatic Adjustment) correction was applied to the TG data using ICE-5G VM2 model (**?**). Since an IB (Inverted Barometer) correction was applied to the satellite altimetry data, the ~~TGs-KP~~ TG data are IB-corrected based on the pressure fields from 20th Century Reanalysis V2c data (**???**). The ~~TG-KP~~ TG data in this study ~~is~~ are modified with further editing criteria. The techniques for editing are similar to those of **?**, with ~~TG-KP~~ TGs that have shorter record length than 5 years and unphysical trends (greater than 7 mm/yr) likely owing to uncorrectable vertical land motions being removed prior to analysis. After calculating a month-to-month change, jumps greater than 250 mm were also removed.

2.1.4 Reconstructed sea level anomalies of previous ~~study~~ studies

~~(??) created a reconstruction~~ **??** created the reconstruction of a global SLA from 1870 to 2009 using EOFs of SLA from satellite altimeter over 1993-2009. They applied ~~the Reduced Space Optimal Interpolation~~ a reduced space optimal interpolation technique. According to their research, the GMSL rose about 210 mm over 1880-2009, and the linear trend through 1900-2009 was 1.7 ± 0.2 mm~~per year~~/yr. The resulting SLA is one of the most comprehensive and widely cited reconstruction results. This ~~data set~~ dataset was employed for long-term background trend for this study (see ~~more detail below~~ section 2.2.3). The GMSL ~~portion of this reconstruction~~ timeseries (**??**) has been extended and made publicly available (http://www.cmar.csiro.au/sealevel/GMSL_SG_2011_up.html). To ~~create the reconstructed sea level anomaly (hereafter ReSLA)~~ reconstruct the past SLA, **?** combined the CSEOFs of the satellite altimetry and historical TG record. This ~~approach provides an advantage for describing local variations such as ENSO and PDO. This~~ weekly analysis has a $0.5° \times 0.5°$ grid resolution and its time coverage is over 1950-2009. This ~~data set~~ dataset was used for the comparison with the reconstruction of this study (see section 2.2.3). This reconstruction dataset ~~?~~ (**?**) can be downloaded from a NASA JPL/PO.DAAC (ftp://podaac.jpl.nasa.gov/allData/recon_sea_level/preview/L4/tg_recon_sea_level/).

2.2 Methods

~~Most of the studies on the reconstruction of sea level have been done on a global scale ??????. In some parts of the world with sparse observations, however, the quality of the reconstruction is poor. Hence to get more accurate results, a local scale study is necessary to produce the level of quality that is necessary for planning and policy-making purposes. To date, this has been an understudied area, however, with relatively few studies on the subject ???.~~

~~The main difficulties are the lack of historical observations and poor spatial distributions of the TG data. The regional reconstruction of sea level around the KP suffers from these problems. The longest TG record extends back only to 1930, and most of the TG data is available only after the mid-1960s with relatively few available in the northern area of the KP. If previous reconstruction schemes are applied that rely only on sea level, then it is likely only possible to obtain reliable results after 1970. A~~ We propose a modified reconstruction method ~~is proposed~~ for an area such as the ~~KP~~ seas around the Korean Peninsula having poor TG coverage ~~. The approach is~~ based on the CSEOF decomposition and multivariate regression ~~while~~ taking into account a time lag. This ~~approach~~ method is a progression from the technique described in **?**. In ~~that study, given the relatively large region of reconstruction (Pacific Ocean basin), tide gauge observations were available for the entirety of the reconstructed record. In this case, suitable tide gauge coverage around the KP is only really available after the mid-1960s, necessitating an approach that is independent of the tide gauge observations. In this section,~~ this section, we show the procedure of the proposed scheme and fundamental theories~~are shown~~.

2.2.1 Cyclostationary empirical orthogonal functions

To understand the complex response of a physical system, the decomposition of data into a set of basis functions is frequently applied. The decomposed basis functions have the potential to give a better understanding of complex variability of the fundamental phenomenon. The simplest and most common computational basis functions are EOFs, which have often served as the basis for climate reconstructions. When a reconstruction selects the EOFs as basis functions, one basis function is defined as a single spatial map accompanied by a time series representing the amplitude modulation of this spatial pattern over time. The EOF decomposition of the spatio-temporal system, $T(r,t)$, is defined by the Eq. (1):

$$T(r,t) = \sum_i LV_i(r)PCT_i(t), \tag{1}$$

where $LV(r)$ is a physical process (or loading vector) modulated by a time series $PCT(t)$ (principal component time series or PC time series). Combining each LV and PCT pair, a signal of single EOF mode can be produced.

The assumption underlying EOF-based reconstruction is the stationarity of the spatial pattern represented by the EOF over the entire period. However, the fact that many geophysical phenomena are cyclostationary is well known ~~?~~(**?**). That is, ~~these some~~ processes are periodic over a certain inherent timescale, with ~~the amplitude of this periodic process~~ corresponding amplitudes varying over time. Even though EOFs represent cyclostationary signals through a superposition of multiple modes, as stated in **?**, representing the cyclostationary signals with stationary EOFs can lead to an erroneous and ambiguous interpretation of the data. It also requires many EOFs to explain a relatively small amount of variability in a dataset.

To remedy some of these issues, **?** introduced CSEOFs as the basis for ~~SLA?s reconstruction~~ the global reconstruction of SLA instead of EOFs. The CSEOF analysis ~~has been~~ proposed to capture the ~~cyclo-stationary~~ cyclostationary patterns and longer scale fluctuations in geophysical data (**?????**). The CSEOF analysis can capture the time varying signals as a single mode by giving a time dependency to the loading vectors.

The system is defined as Eq.(2) and (3).

$$T(r,t) = \sum_i CSLV_i(r,t)PCT_i(t) \tag{2}$$

$$CSLV(r,t) = CSLV(r,t+d) \tag{3}$$

5 where ~~is a cyclo-stationary LV~~$CSLV(r,t)$ is a Cyclostationary Loading Vector (for convenience, we call this as LV) and it is time dependent and periodic with a particular period (called a "nested period"~~and more details in the following sections~~). Previous studies (**???**) provide more detailed walk-through for the CSEOF computation and properties. CSEOFs ~~provide significant advantages~~ have a significant advantage over EOFs since CSEOFs can explain cyclostationary signals in one mode; ~~this means the opportunity of separating physical signals into a single, easy-to-interpret mode~~ that is, CSEOFs of periodic

10 processes are much easier to interpret than EOFs (**????**). **???** demonstrated that CSEOFs provided significant benefits dealing with repeating signals such as ENSO ~~(El Niño-Southern Oscillation) and MAC (~~and Modulated Annual Cycle ~~)~~ signals.

2.2.2 Multivariate regression using CSEOFs

When considering the complete Earth climate system, one variable is often directly connected to another variable. In some cases, they are impacted by a common physical process, or in other cases, one variable may directly influence another. To take

15 advantage of these relationships and establish links, we can perform a multivariate linear regression as following Eq. (4).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{4}$$

where $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ are regression coefficients and the ϵ is random error. In this study, the response variables are each PCT of ~~AVISO-KP~~AVISO's CSEOF and the predictor variables are all PCT of each SST dataset's CSEOF. Eq. (4) can be re-written as follows:

20 $$PCT_{SLA}^m = \beta_0^m + \beta_1^m PCT_{SST}^1 + \beta_2^m PCT_{SST}^2 + \cdots \beta_k^m PCT_{SST}^k + \epsilon^m \tag{5}$$

where PCT_{SLA}^m is the m-th PCT of ~~SLA-KP~~AVISO's CSEOF and β_k^m are regression coefficients for the m-th target and k-th PCT of SST ($m = 1, 2, \ldots, M$; M is total number of target's modes), PCT_{SST}^k ~~and~~ is the k-th PCT of SST's CSEOF. The matrix form of the Eq. (5) is:

$$
\begin{bmatrix} T_1^m \\ T_2^m \\ \vdots \\ T_n^m \end{bmatrix} = \begin{bmatrix} 1 & P_1^1 & P_1^2 & \cdots & P_1^k \\ 1 & P_2^1 & P_2^2 & \cdots & P_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & P_n^1 & P_n^2 & \cdots & P_n^k \end{bmatrix} \times \begin{bmatrix} \beta_0^m \\ \beta_1^m \\ \vdots \\ \beta_k^m \end{bmatrix} \tag{6}
$$

5　where T_n^m is the n-th component of PCT_{SLA}^m, P_n^k is the n-th component of PCT_{SST}^k.

Additionally, many geophysical signals have lagged relations with other geophysical signals (**???????**). Hence~~we think that the~~, by assuming that the each mode of CSEOF represents an independent physical event, we can conclude the PCTs which are mathematically independent of each other also can have a lagged relationship. If we consider the lagged relationships between the target and predictor variables and use the predictors having a higher correlation, we can reduce the number of

10　predictors in the regression; generally, the more predictors applied for the regression, the more noise is likely to appear in the ~~simulation~~result. Before performing the multivariate linear regression system as in (5), we calculated the cross-correlation between the target PCT of ~~SLA-KP~~ AVISO and predictor PCTs of SST. The predictors were selected based on their cross-correlation values. The threshold cross-correlation value did not have a sensitive effect on the regression if the value can select more than ten predictors; in ~~the~~ this study, we used 0.3 as the threshold. By assuming the lag of the i-th mode having maximum

15　cross-correlation at lag ρ_i, the m-th mode's PCT of ~~AVISO-KP~~ AVISO can be given as follow based on the Eq. (5).

$$
PCT_{SLA}^m = \beta_0^m + \sum_{i=1}^{k} \beta_i^m PCT_{SST}^i(t - \rho_i) + \epsilon^m \tag{7}
$$

2.2.3　Reconstruction of ~~the past~~ SLA-KP

~~By extending the PCT of AVISO-KP's CSEOFs, we can reconstruct the past SLA-KP. A unique characteristic of this reconstruction in contrast with others is the non-use of the local TG data sets. As mentioned above, the main motivation for this is the poor~~

20　~~coverage of TGs around the KP. After removing GMSLA from the AVISO-KP at each grid point, the CSEOF decomposition was conducted. This means that if we conduct the reconstruction using AVISO-KP that has no GMSLA (hereafter AVISO-KP0), then the reconstructed SLA-KP0 (hereafter ReSLA-KP0) similarly includes no GMSLA signal.~~

~~Using the regression coefficients and lagged relationship between the PCTs of each SST dataset and AVISO-KP0, we can extend the PCTs of AVISO-KP0 through Eq. (7). By combining the LVs of AVISO-KP0 and extended PCTs, we can rebuild~~

25　~~the past SLA-KP albeit with no GMSLA. Finally, after adding the GMSLA to the ReSLA-KP with no GMSLA, the SLA-KP can be reconstructed with a regional mean sea level change.~~

~~To estimate the confidence intervals of the reconstructions in this study, both AVISO-KP and the SST reanalysis data are assumed as correct values. Based on the assumption, the multiple linear regression provides confidence intervals for each regression coefficient. A MC (Monte Carlo) simulation was carried out using the confidence intervals of multiple linear regression coefficients and GMSL. The MC simulation created 1000 sample-sets for ReSLA-KP with no GMSLA (hereafter ReSLA-KP0). By analysing 1000 sample-sets, we estimate the confidence interval of ReSLA-KP0. However, to ReSLA-KP0 we need to add the GMSLA which has their own uncertainties. We used the GMSLA of ? which played the role of the long-term background change of the SLA-KP and this data-set provided their confidence intervals. Consequently, the overall confidence intervals of the current reconstruction can be estimated by summing the two confidence intervals.~~

~~A procedure of the current reconstruction can be summarized as following. Every SST dataset was~~

As a starting point, every SST dataset was trimmed to have the time span of 1891-2014. The ~~AVISO-SLA was trimmed to contain only the data~~ AVISO was trimmed around the KP ~~(31°-46°N and 117°-142°E). The~~ and the southeast sea of the Japanese islands was removed. Every SST ~~dataset~~ data was cut into three regions: around the KP~~(same box with AVISO-KP; hereafter add '-KP')~~, the Northwest Pacific Ocean~~(25°-55°N and 110°-160°E; hereafter add '-NWP')~~, and global (no trimming). All grid points that were not continuous in time were removed for every dataset. In total, we tested six different SST ~~data combinations. GMSLA~~ combinations. GMSL and mean values were removed from ~~AVISO-KP~~ AVISO at each grid point. Each ~~dataset~~ data point was weighted by the square root of the cosine of latitude to consider the actual area of each grid. ~~The CSEOF decomposition was applied to all data sets (AVISO-KP0 and SST datasets~~We conducted the CSEOF decomposition for all data (AVISO without GMSL and six SST combinations) with twelve ~~months~~ month nested period. The lagged relation between PCTs of ~~AVISO-KP0~~ AVISO without GMSL and PCTs of each SST dataset were estimated with two years maximum ~~lagging boundary~~lag. Using the PCTs of each dataset's CSEOF, we built the multiple linear regression systems ~~were built~~ based on Eq. (7) over 1993-2014. In this regression, the target variables were each PCT of ~~AVISO-KP0~~ AVISO and the predictors are PCTs of each SST ~~dataset~~combination. The regression coefficients and their confidence intervals were estimated to extend the target variables. Applying ~~MC~~ Monte Carlo simulation that used the confidence intervals of regression coefficients, we randomly generated a thousand sample-sets of each extended PCT of ~~AVISO-KP0~~AVISO without GMSL. By combining the extended PCTs to the LVs of ~~AVISO-KP0~~AVISO without GMSL, we produced a thousand ~~ReSLA-KP0s~~SLAs without GMSL. By adding ~~the GMSLA~~ randomly generated GMSLs (?) to the ~~ReSLA-KP0~~reconstructed SLAs without GMSL, a thousand of ~~ReSLA-KPs~~ SLAs were generated. Finally, by statistical analysis of each time step of the random samples, we estimated the mean variation and their confidence intervals of each reconstruction.

For comparison, in addition to the ~~TGs-KP~~TG, we used the reconstructed dataset of ~~?; hereafter ReSLA-H. Their reconstruction was based on the TG records and satellite altimetry's CSEOF~~. We trimmed the dataset to have same domain with this study. The reconstruction results over 1970-2009 are quite reliable, because, after 1970, the number of available TG record around the world is enough to guarantee the reconstruction results. The correlation coefficient (ρ) and NRMSE (Normalized Root Mean Square Error; we obtain this value through dividing RMSE by the standard deviation of the reference dataset; ~~see~~ Eq. (8)) values for the entire domain and each TG location were calculated. By using these two values, we decided the best reconstruction

case among the six reconstructions which are introduced in section 3.2.

$$NRMSE = 1 - \frac{\|x_{ref}(i) - x(i)\|}{\|x_{ref}(i) - \mu_{x_{ref}}\|} \tag{8}$$

where $\|\blacksquare\|$ indicates the 2-norm of a vector, x_{ref} and x are reference data and tested data respectively.

3 Results and Discussions

3.1 Sea Level ~~Anomaly~~ Anomalies around the KP

Using ~~AVISO-KP~~ the AVISO over 1993-2015, a linear trend map was estimated as shown in Fig. 3. The mean trend was found to be 3.1 ± 0.5 mm/yr. The linear trend of mean SLA-KP ~~(hereafter MSLA-KP)~~ agrees closely with the ~~global SLA~~ GMSL trend, 3.0 ± ~~0.0~~ 0.4 mm/yr (~~see~~ Fig. 4). Due to the similarity between the long-term trends of ~~MSLA-KP and GMSLA~~ mean SLA-KP and GMSL (Fig 4), it is reasonable that ~~the MSLA-KP~~ we assume the SLA-KP can be described as the combination between background signals (~~GMSLA) and variabilities from the background signals (see Fig. 4).~~ GMSL) and the residuals which contain local characteristics of SLA-KP. Most of the SLA-KP trends were close to the mean, but some parts of the East/Japan Sea, and of the Yellow Sea close to land, exhibited extreme patterns. Some areas showed trends over 7 mm/yr, while in other regions there were trends less than 1 mm/yr of the linear trend (~~see~~ Fig. 3). To ~~reason~~ check whether the extreme ~~trends patterns was related to the local mass distribution caused by various sources such as vortex and river discharge or was an independent phenomenon, we~~ linear trends patterns had a significant influence on mean SLA-KP, we compared the mean SLA of the area having the extreme linear trends and the other area. We calculated the mean correlation (hereafter $\bar{\rho}$) of each ~~AVISO-KP?s grid point~~ grid point of AVISO to separate the two areas. For example, $\bar{\rho}$ at a single grid point P was calculated by taking mean of ρ values that had been estimated between P and all other points. By repeating these calculations at all the points of AVISO, we obtained Fig. 5. We deemed that the ~~SLA~~ SLAs of the regions having relatively high $\bar{\rho}$ ~~fluctuates with each other~~ fluctuate together, on the other hand, the ~~SLA~~ SLAs of the low $\bar{\rho}$ regions ~~did not change with each other~~ oscillate separately. The regions that had the relatively low correlation coefficient agreed with the regions that had the extreme linear trends (~~see~~ Fig. 3 and 5). We divided the SLA-KP into two regions according to the mean correlation coefficient; we roughly selected the threshold value as 0.5, which can separate the area having extreme trend and the remaining area. The ~~MSLA~~ mean SLA of each region shows a good agreement each other (~~see~~ Fig. 6). This demonstrates that the small-scale extreme features tend to cancel out and do not significantly impact ~~MSLA-KP~~ on mean SLA-KP. This also suggests that the entire region can be treated as local variability fluctuating about some background long-term mean, an important feature for this reconstruction procedure.

The ~~purpose of the study of SLA-KP during the satellite era is to increase our understanding of SLA-KP before conducting the reconstruction of SLA-KP. To achieve this goal, an agreement between TG-KP and AVISO-KP was estimated in terms of correlation coefficient and linear trend by using averaged time series and individual time series~~ linear trend at each TG location

~~. These uneven patterns originated from two sources; one is river discharge in the Yellow Sea, and the other is a vortex induced upwelling and downwelling effect in the East/Japan Sea area. The Dayang, Huli, Yingna, Zhuang, and Xiaosi Rivers flow into the Yellow Sea from China, and Yalu (Amnokgang), Taeryong, Taedong, Han, Geum, Mangyeong, Dongjin, and Yeongsan Rivers discharge into the Yellow Sea. The extreme patterns near the land seem to relate to the variation of river discharge. In the East/Japan Sea, both warm currents and cold currents exist simultaneously and the borderline repeatedly oscillates north and south. Near the borderline, the warm current and cold current make small gyres, and the gyres make the uneven surface variations. These kinds of large variability sea level features make the assessments of the linear trend poor.~~

The linear trend at each TG location was estimated and it was compared with the nearest point in ~~AVISO-KP (Fig. 7). The~~ AVISO; using the same data, the ρ values ~~between TG-KP and AVISO-KP~~ were estimated and the mean $\bar{\rho}$ value of the ρ was about 0.72 (Fig. 7). ~~The comparison showed that only five TGs showed acceptable accuracies having less than 30% of difference with the AVISO-KP's linear trend.~~ Eleven TGs showed ~~more~~ greater than 30% of underestimation and twenty-one TGs had ~~more~~ greater than 30% of over estimation. To figure out the effect of these disagreements, the ~~MSLA-KP~~ mean SLA of AVISO was compared with the ~~MSLA of TGs-KP, and these time series~~ TG's mean SLA, and they showed $\bar{\rho} = 0.89$ and NRMSE = 0.52 (~~see~~ Fig. 8). The ~~MSLA rise of combined~~ linear trend of mean SLA of the TGs was estimated as 4.31 mm/yr and this value is about 40% higher than the ~~MSLA-KP~~ mean SLA of AVISO. This disagreement ~~originated from the~~ likely results from the mismatching between locations of TG stations and AVISO grid points, the short time period, and a lack of TGs. Unresolved vertical land motion at the ~~TG-KP~~ TGs could also lead to such disagreements.

3.2 ~~Sea Level Reconstruction around the KP~~

~~To begin the process of reconstructing sea level around KP, CSEOF decompositions (???) with twelve-months nested period were performed on both the AVISO- KP and the SST datasets as described above. The datasets were decomposed into Loading Vectors (LVs)and corresponding time series of Principal Components (PCTs).~~

~~To reconstruct~~ CSEOF decomposition was conducted to investigate the variability of SLA-KP with twelve month nested period after removing mean values at each grid point. The first mode represents an annual variation considering the spatial patterns and PCT of the CSEOF (Fig. 10). Nearly 60% of SLA-KP ~~over 1900-2014, we then applied the multivariate regression accounting for lagged relationships, relying on CSEOF?s modesof SST and AVISO-KP. For these reconstructions, two SST reanalysis datasets (ERSST and COBESST2)were used. Each SST data was divided into three cases: global, NWP, and~~ variations can be presented by the first mode (Fig. 9). The second mode shows similar spatial patterns having positive value for all months, and the PCT shows clear positive trend (Fig. 11). This mode can be interpreted as representing the rising sea levels, explaining 10% of variations of SLA-KP roughly. The third and fourth modes were not obviously related to specific modes of variability, explaining only 5% and 3% respectively. Using the four modes, we can explain about 70% of SLA-KP. The first and second modes have the linear trend, but the linear trend in the first mode is negligibly small compared with the signal itself (Fig. **??** and **??**). Hence, we can say that the second mode is the most important key to estimating SLA-KP.

3.2 Reconstruction of SLA around the Koran Peninsula

One of the unique characteristics of the current study is that we only used SST as a proxy of former SLA; other studies, however, used TG data or combined data (TG and SST). There are multiple reasons why we chose not to use TG data for the current reconstruction. The first reason is due to both the poor data coverage and the poor data quality. There are relatively few tide gauges extending into the past in our study area, and even fewer that are of high quality (i.e., unaffected by vertical land motion, with few gaps, free of non-physical jumps). The second reason, and related to the first, is that due to a methodological characteristic of the CSEOF analysis, a dataset that is free of gaps (temporally continuous) is needed. To satisfy this requirement, we are led to other gridded reconstruction or reanalysis products. There are many types of data that could potentially be used in our scheme (e.g. wind, ocean current, precipitation, atmospheric pressure). We used only SST for the following reasons. 1) SST and SLA have a distinct relationship when we analyze both of data through CSEOF analysis (**???**) and **?** showed that SST could be a good proxy of SLA in this part of the ocean. 2) Limiting our analysis to SST reduces the possibility of overfitting in the regression scheme we use to reconstruct. As a final benefit of using SST, we can check against the available tide gauge data to provide an independent comparison to our reconstruction.

We made six reconstructions (Sec. 2.1.2 and Fig. **??**), and compared the ~~entirety of the KP region shown in the figures. As a result, six cases of reconstructions were conducted and the six reconstructions showed a reasonable agreement with MSLA-TG~~ six reconstructions with **?** and TG over ~~1965-2014. For the period prior to 1965, however, the results showed considerable diversity (see Fig. **??**). The mean reconstructed SLA-KPs (hereafter ReSLA-KPs) were compared with the mean reconstructed SLA of previous study (**?, ;** ReSLA-H) and the MSLA-TG from 1970-2009 considering the available number of TG data~~ 1970-2008; we could not use complete TG coverage for the comparison because there were only a few TG data available before 1970. Both a correlation coefficient and ~~normalized root mean squared error (NRMSE~~ an NRMSE (Normalized Root Mean Square Error; Eq. 8) were applied for the quantified comparison ~~. The comparison result is given in~~ (Fig. **??**). Considering the NRMSE, ~~we can see that the SST of NWP and KP provided better reconstructions than ReSLA-H because the NRMSEs of these cases are greater than ReSLA-H. However, considering the~~ all reconstructions except the global ERSST case provided better agreement than **?**; the best reconstruction was the case of COBESST2 of the Northwest Pacific. Regarding correlation coefficient, ~~only SST of NWP datasets~~ two reconstructions (COBESST2 of the Northwest Pacific and ERSST of the Northwest Pacific) showed better results than ~~ReSLA-H. Finally~~**?**; the reconstruction from COBESST2 of the Northwest Pacific provided the best result. Consequently, we selected the reconstruction ~~using COBESST2-NWP (hereafter ReSLA-NWP)~~ from COBESST2 of the Northwest Pacific as the best reconstruction ~~considering~~ regarding both NRMSE and correlation coefficient. And the mean SLA of the best case showed a reasonable agreement with the mean SLA of TG over 1965-2014. For the period before 1965, however, the result showed considerable disagreement (Fig. **??**).

Most of the reconstructions show better agreement than ~~the reconstruction of **?** in terms of correlation coefficients despite~~ **?** concerning correlation coefficients although we did not use TG data during the reconstruction process. ~~We compared MSLA-KPs from TG-KP, ReSLA-H, and the results of current study to check the reconstruction result. The mean ReSLA-KPs show~~ The mean of reconstructed SLA shows good agreement with the ~~mean ReSLA-H~~**?**, but poor agreement with the ~~MSLA-TG (see~~

12

TG (Fig. **??**). This disagreement, however, is likely caused by lack of high-quality TGs before 1970. We further calculated ~~the correlation coefficient, ρ, and linear trend using ReSLA-KP, ReSLA-H, and TG-KP~~correlation coefficients and linear trends using TGs and reconstructions (current study and **?**) at the each TG location; for the reconstructed data, we calculated the linear trends at the nearest grid points. We made two correlation comparisons: one between ~~ReSLA-KP and TG-KP~~this study

5 and TG, and the other between ~~ReSLA-H and TG-KP to check if ReSLA-KP showed better representation of each TG-KP. The ReSLA-KP showed higher ρ values than ReSLA-H (see Fig. ??a)~~ **?** and TG. This study's reconstruction showed higher correlation coefficients than **?** demonstrating the better agreement between the current reconstruction and ~~TG-KP.~~ TG (Fig. **??**a). The linear trends of ~~TG-KP, ReSLA-KP, and ReSLA-H~~ TG, current reconstruction, and **?** were estimated at the TG location over 1970 to the present; for the calculation, each time series was edited to have the same time span~~data gaps~~. The es-

10 timated linear trends are given in Fig. **??**b. ~~Fig. ?? indicates that the ReSLA-KP~~The current study has similar linear trends with ~~ReSLA-H~~ **?** at the TG location, and the variance of the trends are smaller than ~~TG-KP. ReSLA-KP comparing to ReSLA-H~~ TG (Fig. **??**b); we conducted t-test to check statistical significances of the trend values, and the p-values read in Fig. **??**c. The current study shows better agreement with the ~~AVISO-KP~~ AVISO than **?** over satellite era (~~see~~ Fig. **??**)~~; it~~. It also has more fluctuations (~~see~~ Fig. **??**), ~~which are important to apply this results for engineering purposes. These~~ and these detailed

15 fluctuations are closer to the ~~actual sea level variability~~AVISO, and this is likely a result of the applied number of modes for the reconstruction process. **?** used a limited number ~~of~~ (< 90% of total variance) of CSEOF modes to avoid over-fitting issues, but in this study, we used nineteen CSEOF modes ~~are used~~ which explain 98% of total variance of SLA-KP.

Using ~~MC simulation, a 95% confidence interval was estimated based on~~ Monte Carlo simulation, the means and standard deviations of reconstructed SLAs were estimated for the best reconstruction case (~~COBESST2-NWP~~COBESST2 of the Northwest

20 Pacific). By applying the ~~regression coefficients? mean and standard deviation~~means and standard deviations of regression coefficients (Eq. 7), each mode~~?~~'s PCT was randomly ~~generated~~extended into the past, and the process was repeated by ~~thousand times and these~~ a thousand times. The extended PCTs were combined with ~~CSLV's of AVISO-KP~~corresponding LVs of AVISO. Through this process, ~~thousand of SLA-KP reconstructions~~ a thousand of SLA were generated, and the mean and standard deviation were estimated ~~using these. This means that the reconstructed data has their mean and standard deviation values~~at each time step and grid point. The resulting ~~MSLA-KP~~mean SLA and 95% confidence interval are shown in Fig. **??**.

The linear trend in ~~SLA-KP~~ the reconstructed SLA over 1900-2014 is estimated as 1.71 $\pm$ 0.04 mm/yr, and this value is similar to the linear trend of **?** as 1.70 $\pm$ 0.02 mm/yr. A linear trend ~~at each grid point of AVISO sea level anomaly data~~ map of the reconstructed SLA was calculated, and the maximum and minimum linear trends are about 2.1 mm/yr and 1.4 mm/yr, respectively (Fig. **??**). The difference ~~on the linear trends map~~ between two extreme values of the reconstructed ~~SLA-KP~~

30 SLA is much less than the ~~AVISO-KP?s linear trends~~AVISO over 1993-2015~~. This~~, particularly in the Yellow Sea, (Fig. 3 and **??**). This alleviation means that the ~~long time period reduced~~extended reconstruction period can reduce the effect of ~~large amplitude signals. This is particularly true for the high-trend areas in the Yellow Sea where trends were weakened significantly.~~ the internal variability having a large amplitude.

4 Summary

There were two primary goals of the work presented in this study: 1) Improve the understanding of the sea level around the KP both in the past and present and 2) Present a new reconstruction scheme for local areas with insufficient tide gauge coverage. To meet these goals, we used the satellite altimeter data ~~from AVISO~~ and the TG data ~~from PSMSL~~ to investigate the characteristics of SLA-KP. The linear trend of ~~MSLA-KP~~ SLA-KP was estimated as 3.1 ± 0.5 mm/yr from the satellite altimeter data (~~see~~ Fig. 4). However, when we looked into the trend map, some areas (such as near the river mouth in the Yellow Sea and in the middle of the East/Japan Sea) showed significant departures from the mean (~~see~~ trend (Fig. 3). ~~Understanding this spatial variability has important implications for future planning efforts around the KP.~~

To investigate this further, the reconstruction was performed using ~~AVISO-KP~~ AVISO and two SST reanalysis datasets. Each SST dataset was divided into three cases (global, ~~North-west Pacific, around the Korean Peninsula~~ the Northwest Pacific and KP). The six datasets were decomposed by CSEOF analysis; the ~~AVISO-KP~~ AVISO was decomposed into CSEOF modes after removing the GMSL. The decomposed ~~CSEOF modes' CSLV~~ LVs played a role of basis functions for the reconstruction, and the main process of reconstruction was extending the PCTs of each mode into the past. ~~The six reconstructed SLA-KPs~~ Six reconstructions were generated by this study over 1900-2014. Using ~~the~~ a correlation coefficient and ~~the normal root mean squared error~~ an NRMSE, the best reconstruction was selected. The best reconstruction was produced by COBESST2 ~~data of the North-west Pacific~~ area of the Northwest Pacific. Through the best reconstruction results, the linear trend of ~~SLA-KP~~ SLA was estimated as 1.71 ± 0.04 mm/yr ~~.~~ over 1900-2014 (Fig. **??**). The extreme linear trends shown in Fig. 3 did not appear in the reconstructed SLA-KP (Fig. 3 and **??**). ~~This reconstruction showed better agreement than the previous study's result (?, ; see Fig. ?? and ??).~~

While we focus here on a specific example (the KP), this study can be used to inform other efforts in studying past ~~, present and future~~ and present sea level in areas with poor tide gauge coverage~~and significant future risks to impending sea level rise~~. Our interest was on the KP, specifically, but it was found that including information from the Northwest Pacific improved the localized representation of sea level. Consequently, considering large-scale ocean variability and teleconnections between different parts of the ocean is important when selecting the reconstruction domain. This study also demonstrates that ~~tide gauges~~ TG data may not even be necessary to understand sea level in the past. Using only satellite-based sea level information and SST, we found dramatic improvements between the current reconstruction and past efforts, particularly when comparing to the ~~tide gauge~~ TG variability. Many ~~tide gauges~~ TGs are influenced by vertical land motion that cannot easily be corrected for. Relying on SST alleviates concerns associated with non-ocean related trends. It should be noted that this reconstruction may not work as well in other parts of the ocean, especially those with a less pronounced agreement between sea level and SST. This study does, however, demonstrate the extended efforts that must be made to obtain accurate information about past sea level. As planning efforts get underway in more parts of the world, such comparisons between past and present sea level will become more important, and alternative approaches to simply using ~~tide gauge~~ TG information are going to be needed.
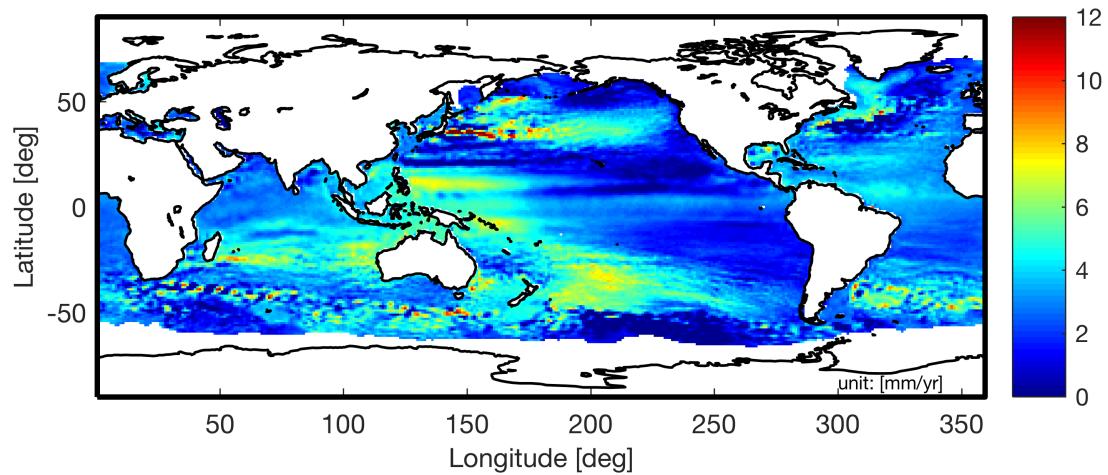
Figure 1. ~~TG locations around the Korean Peninsula. The black square is Wajima TG station which has the longest record length (1930-present)~~Global linear trend map of sea level anomalies using AVISO from 1993 to 2015
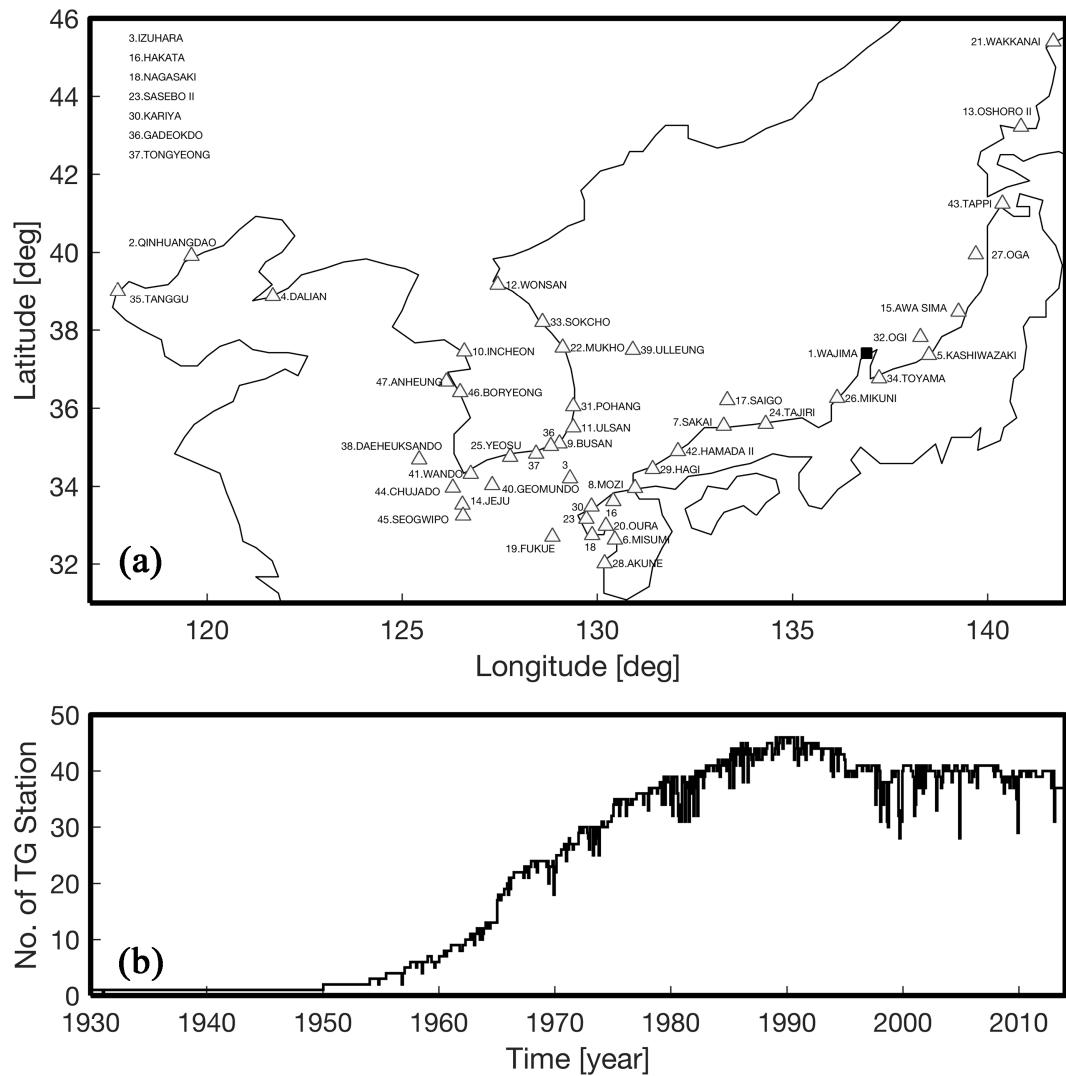
Figure 2. (a) The locations of tide gauge station used in this study around the Korean Peninsula. The black square is Wajima station which has the longest record length (1930-present); (b) The number of tide gauge stations provided by PSMSL around the Korean Peninsula
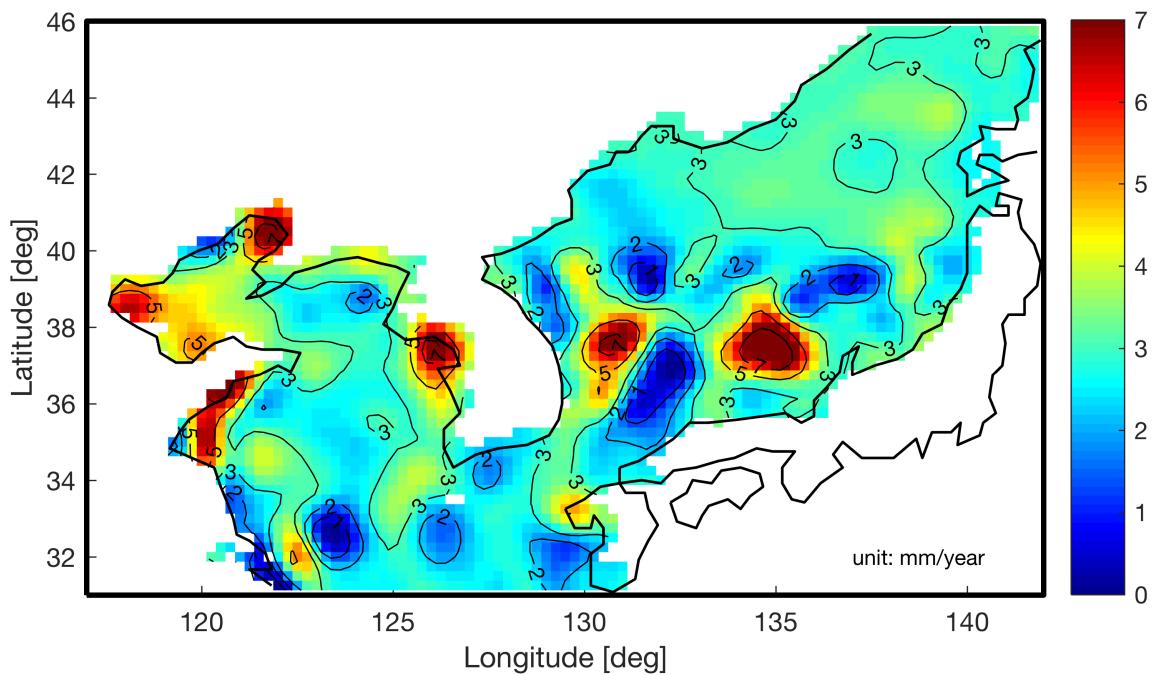
Figure 3. Linear trend map of ~~AVISO-KP (1993-2015)~~ sea level anomalies around the Korean Peninsula from AVISO without annual signal from 1993 to 2015
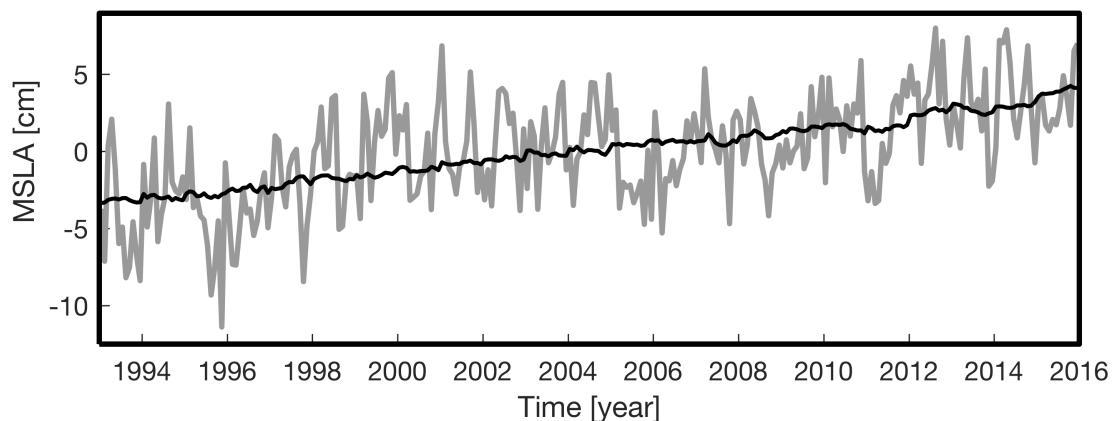
Figure 4. ~~Comparison between MSLA-KP and global MSLA~~ Spatial mean time series of sea level anomalies around the Korean Peninsula (~~1993-2015~~gray) and global (black) from AVISO without annual signal
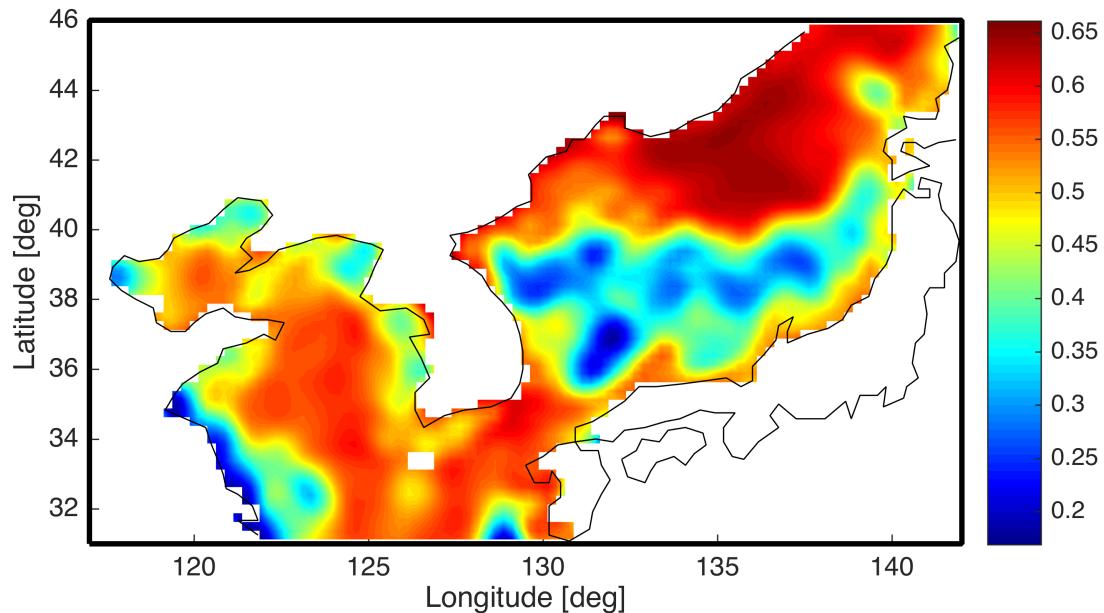
Figure 5. Mean correlation ~~coefficients between each grid?s SLA and other grid?s values (1993-2015)~~ coefficient map of sea level anomalies around the Korean Peninsula from AVISO without annual signal from 1993 to 2015
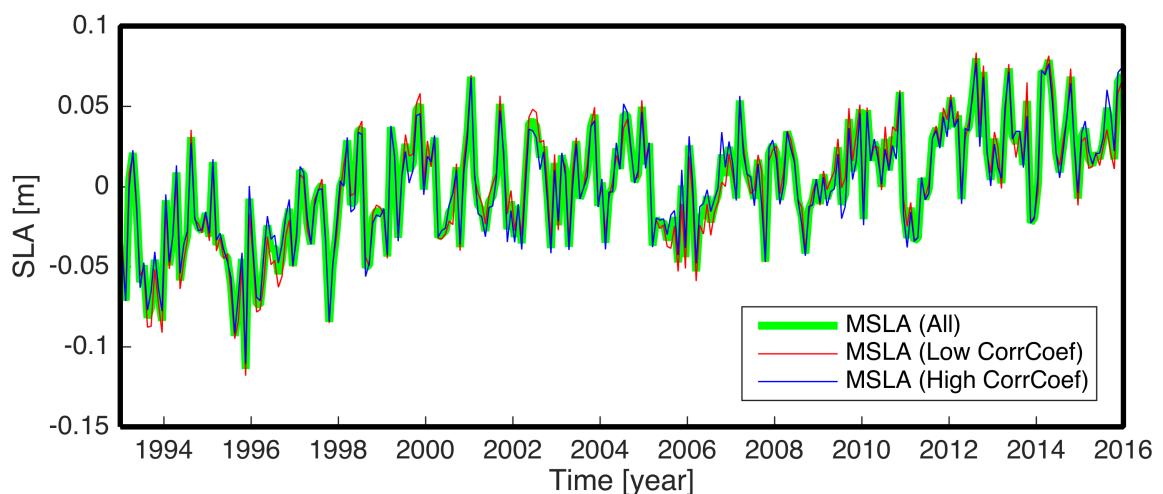
Figure 6. ~~Comparison of the~~ Spatial mean ~~SLA divided into~~ time series of sea level anomalies from two regions based on the correlation coefficients in Fig. 5
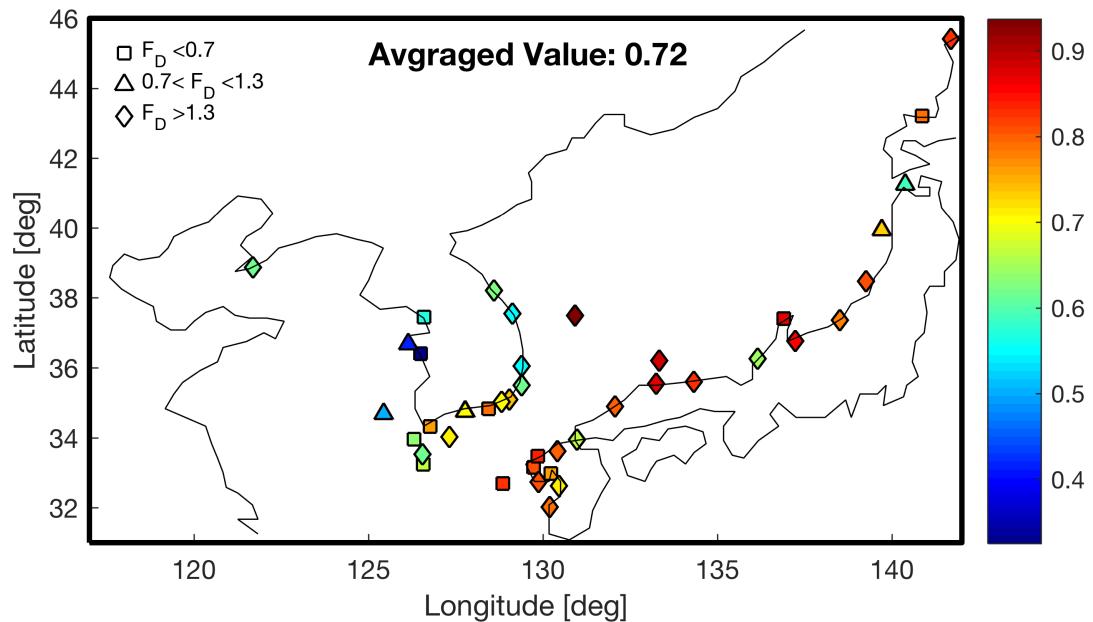
Figure 7. Linear Trends comparison (shapes) and correlation coefficients (colors) between ~~TG-KP~~ tide gauge and ~~AVISO-KP over 1993-2014~~the closest AVISO grid point (< 12 km) from 1993 to 2014, where $F_D = SLR_{TG} / SLR_{AVISO}$ (without annual signals)
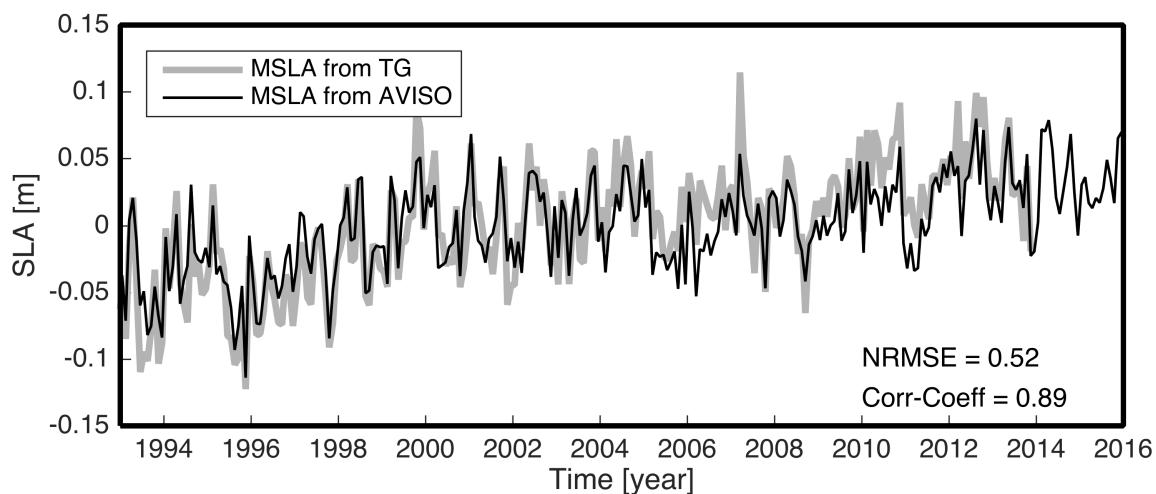
Figure 8. ~~MSLA-KP~~ Spatial mean time series of ~~AVISO~~ sea level anomalies of tide gauge and ~~TG~~ AVISO around the Korean Peninsula without annual signal
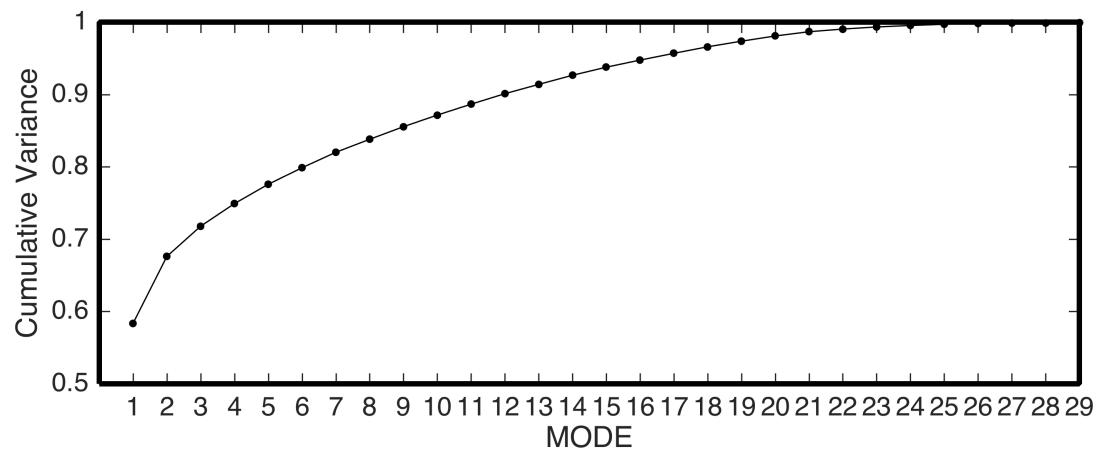
Figure 9. Cumulative variance of CSEOF modes of the AVISO around the Korean Peninsula
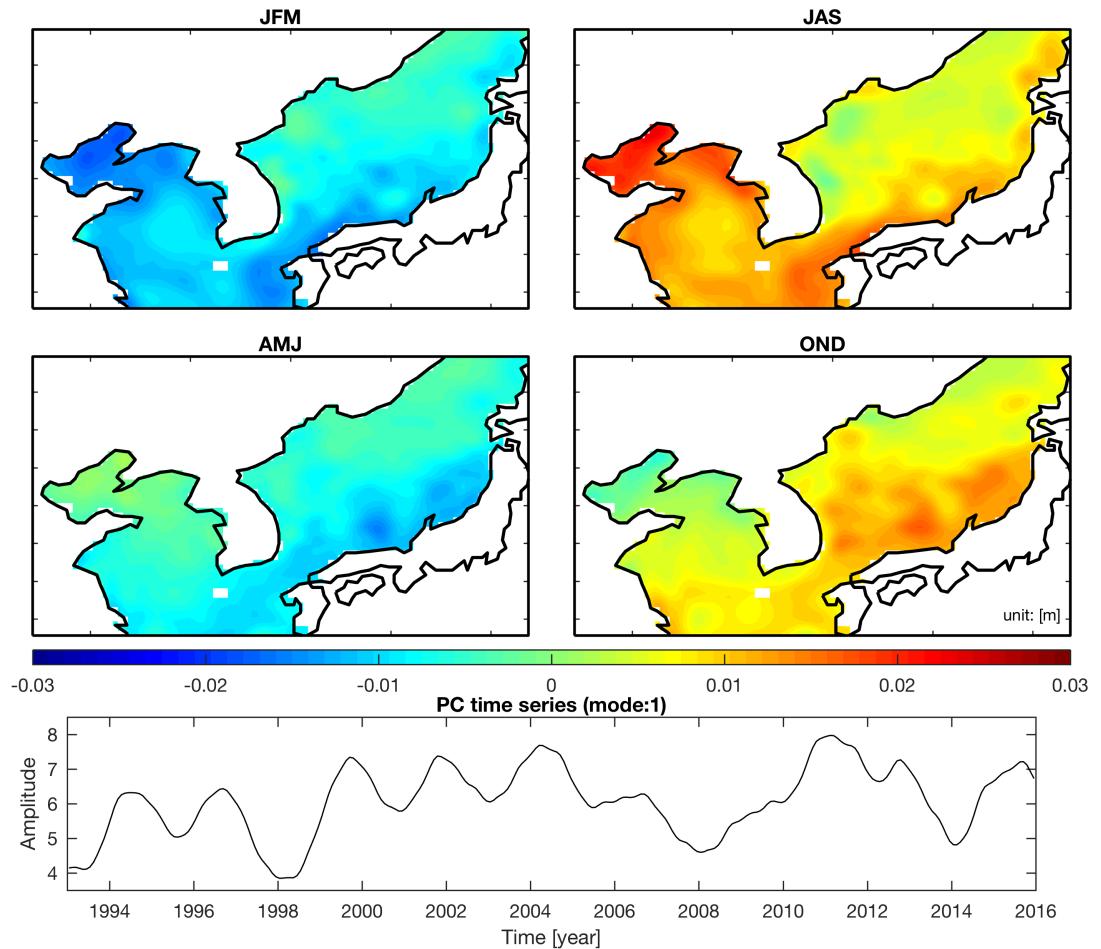
Figure 10. The first CSEOF mode of ~~AVISO-KP~~AVISO around the Korean Peninsula
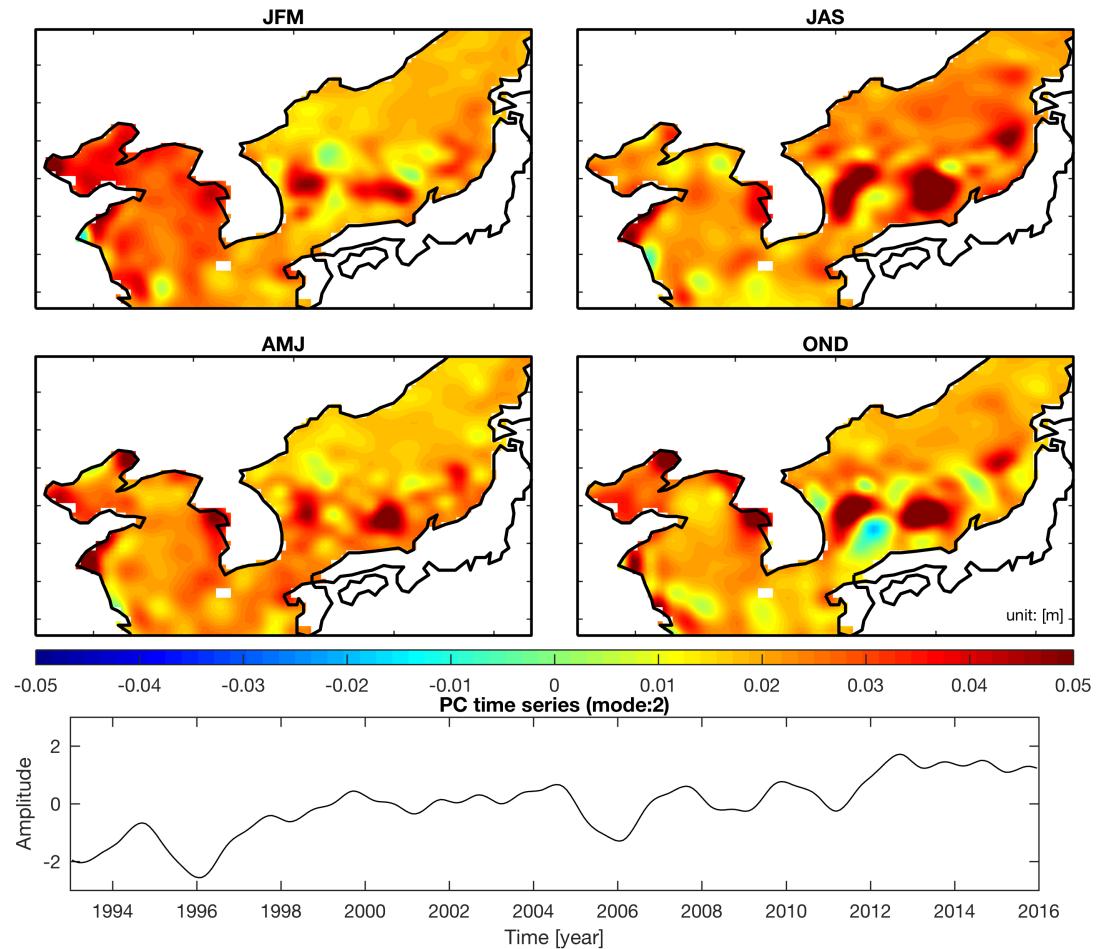
Figure 11. The second CSEOF mode of ~~AVISO-KP~~<u>AVISO around the Korean Peninsula</u>
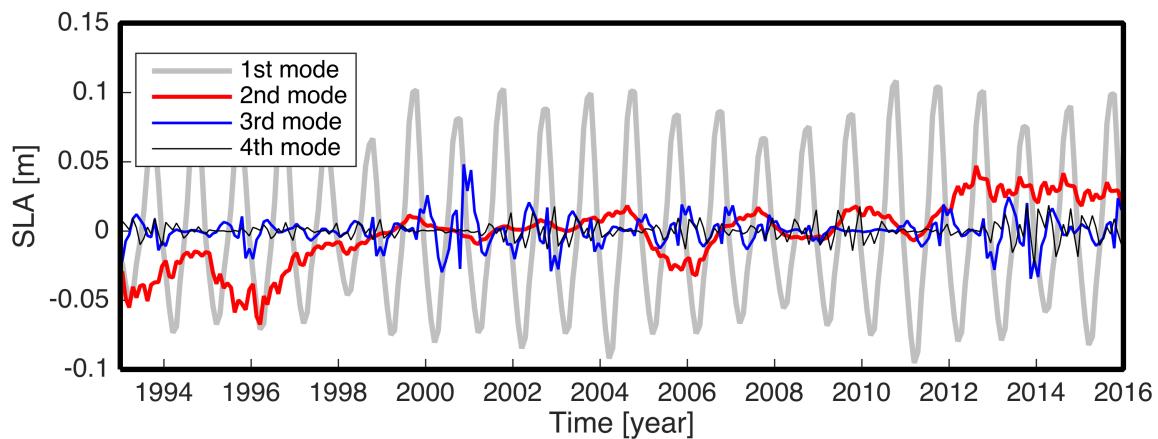
Figure 12. ~~MSLA~~ Mean SLA of the four biggest modes of CSEOF decomposition of AVISO around the Korean Peninsula
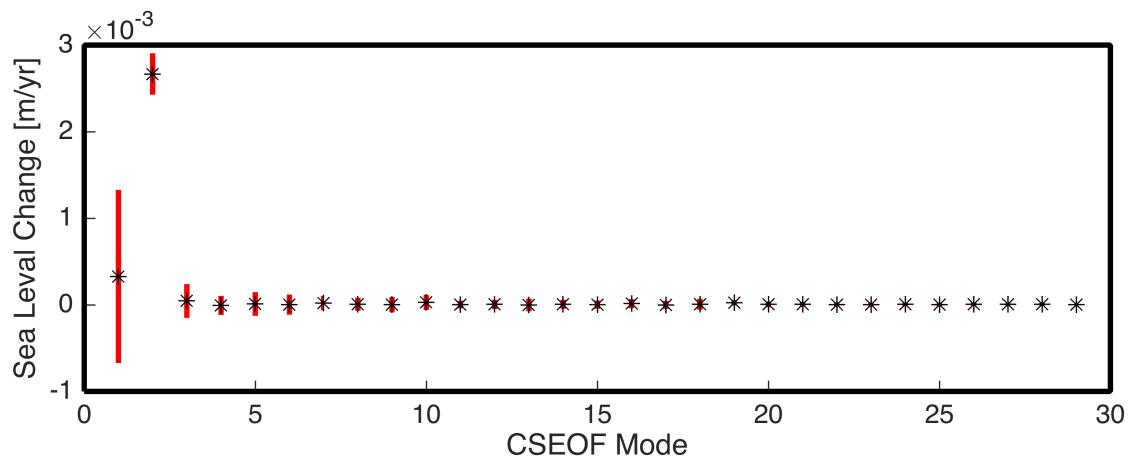
Figure 13. ~~Comparison between reconstructed MSLA~~ Linear trends (black '*') and ~~the TG MSLA~~ 95% confidence intervals (~~ERSST~~ red line) of the ~~North-West Pacific)~~ spatially averaged CSEOF mode of AVISO around the Korean Peninsula
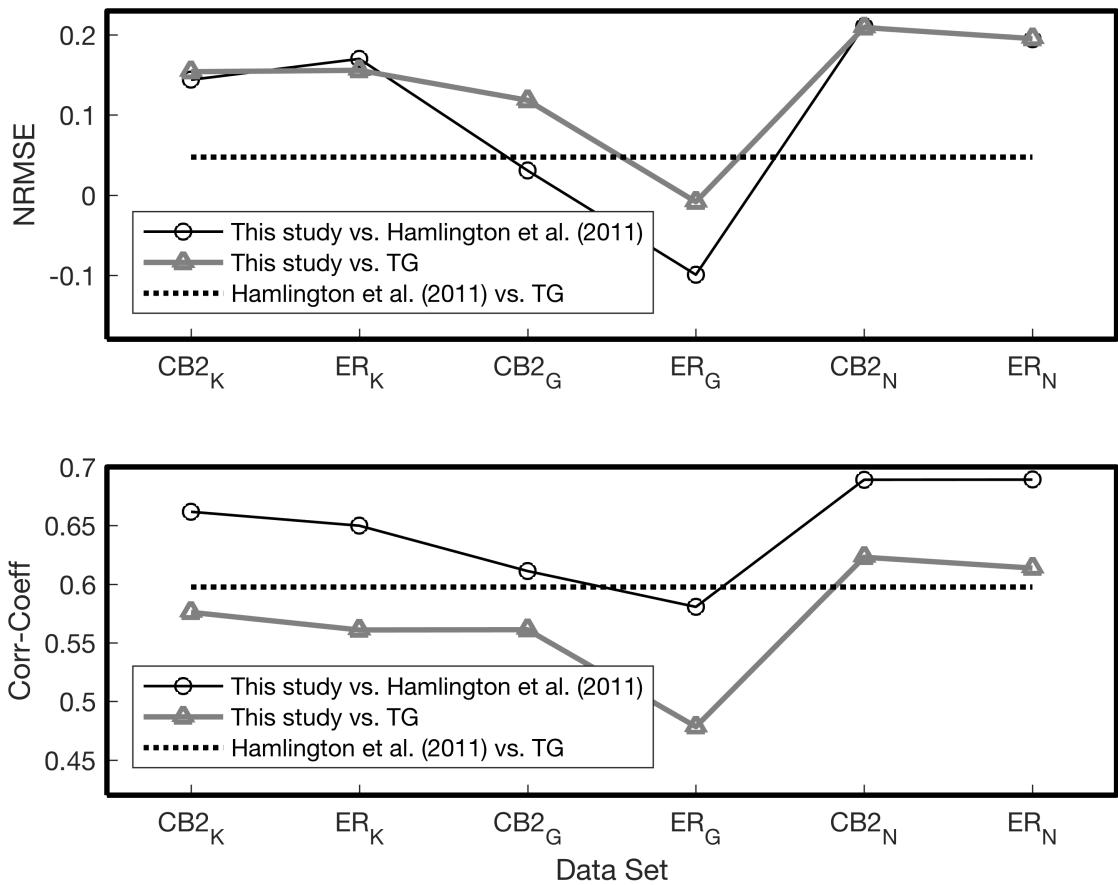
Figure 14. Results of goodness of fit test for Reconstructed ~~MSLA~~ Mean SLA according to **?** and TG ~~MSLA~~Mean SLA; the top figure include normalized root mean squared error and the other include the correlation coefficients; here subscripts K, G, and N represent around the Korean Peninsula, Global, and the North-West Pacific, respectively and CB2 and ER represent COBESST2 and ERSST
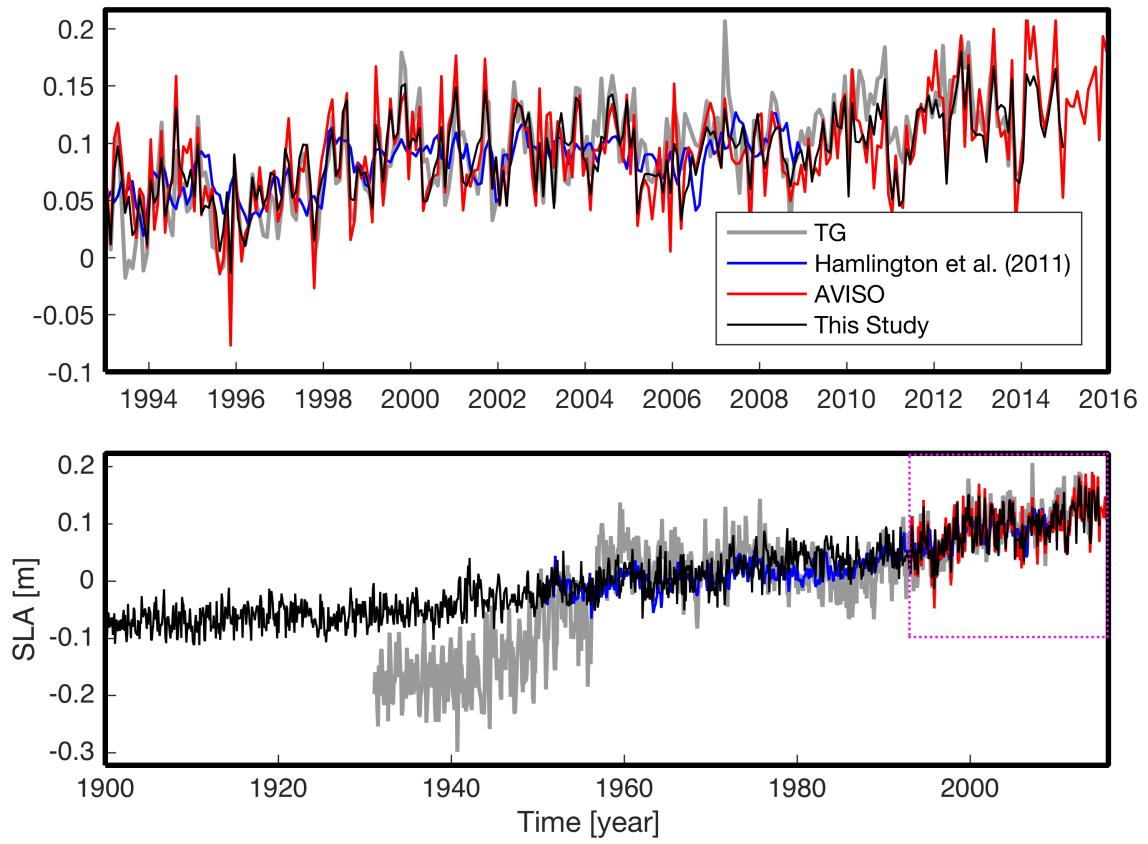
Figure 15. Comparison of ~~MSLA-KP~~ spatial mean time series of sea level anomalies around the Korean Peninsula without annual signal; the top figure is the expansion of a box in bottom figure.
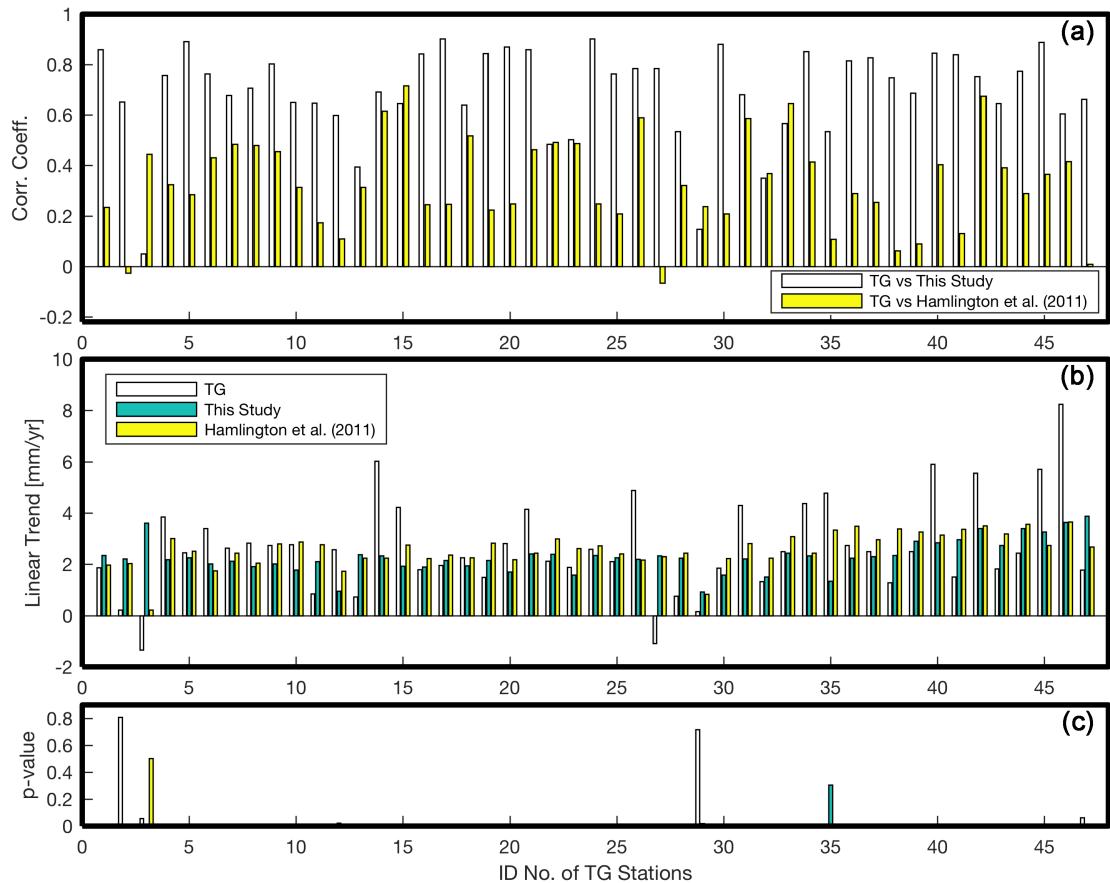
Figure 16. (a) Comparison of correlation coefficients between ~~TG-KP~~ TGs and the ~~reconstruction results~~ reconstructions over ~~1993-2008~~1970-2008; (b) Comparison of linear trends over 1970-2008; (c) t-test result of (b)
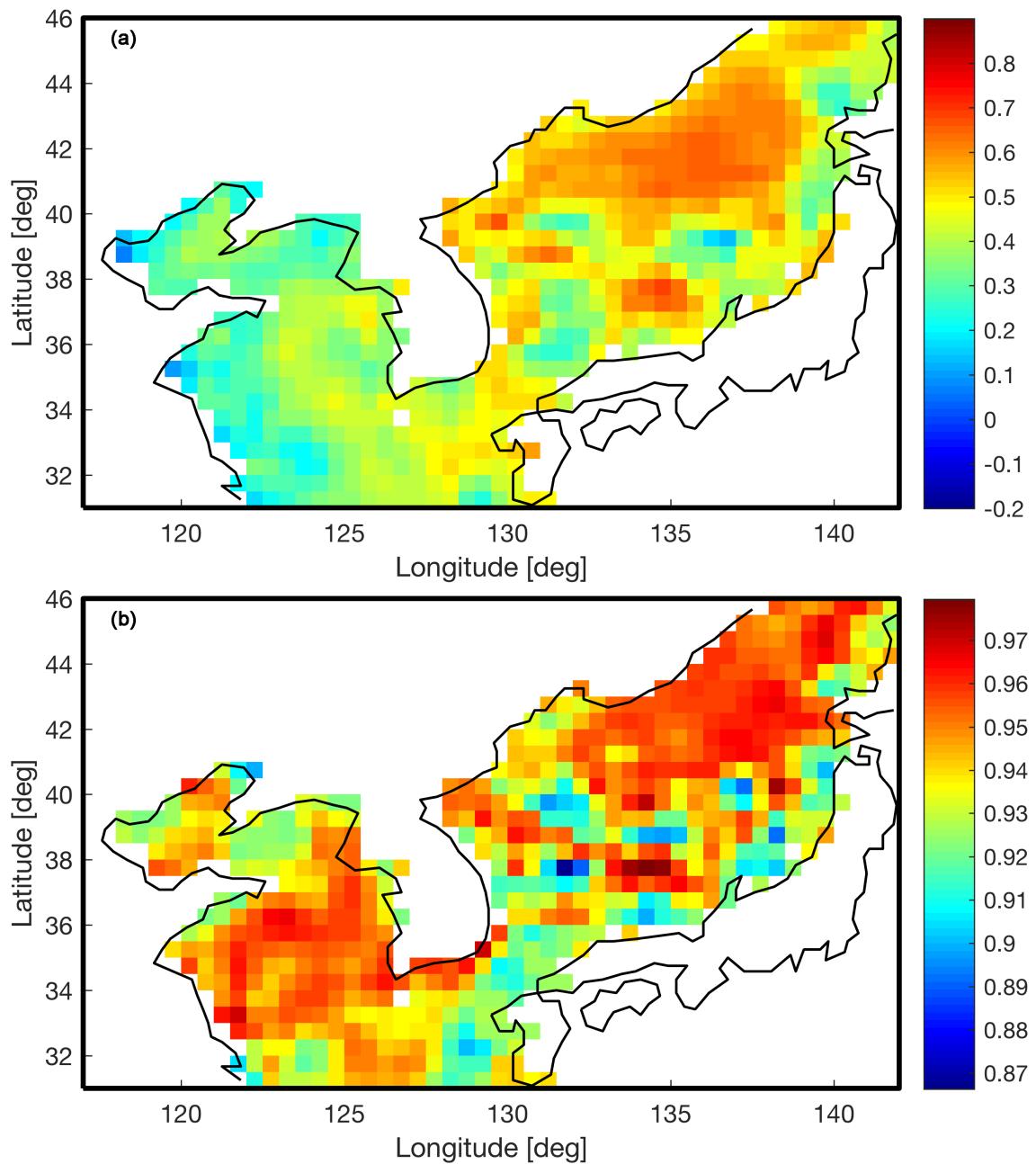
Figure 17. (a) Correlation coefficient map between **?** and ~~AVISO-KP~~ AVISO over 1993-2008; (b) Correlation coefficient map between this study and ~~AVISO-KP~~ AVISO over 1993-2008
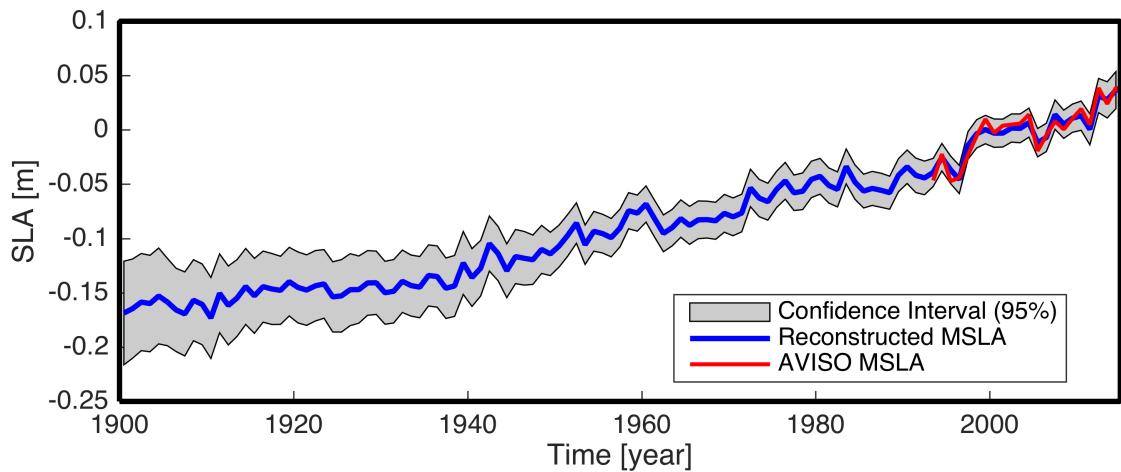
Figure 18. ~~The Best reconstructed~~ Spatial mean time series of sea level anomalies (MSLA) of the best recostruction case (COBESST2 of ~~the North-West~~ Northwest Pacific~~Ocean~~) and 95% confidence interval.
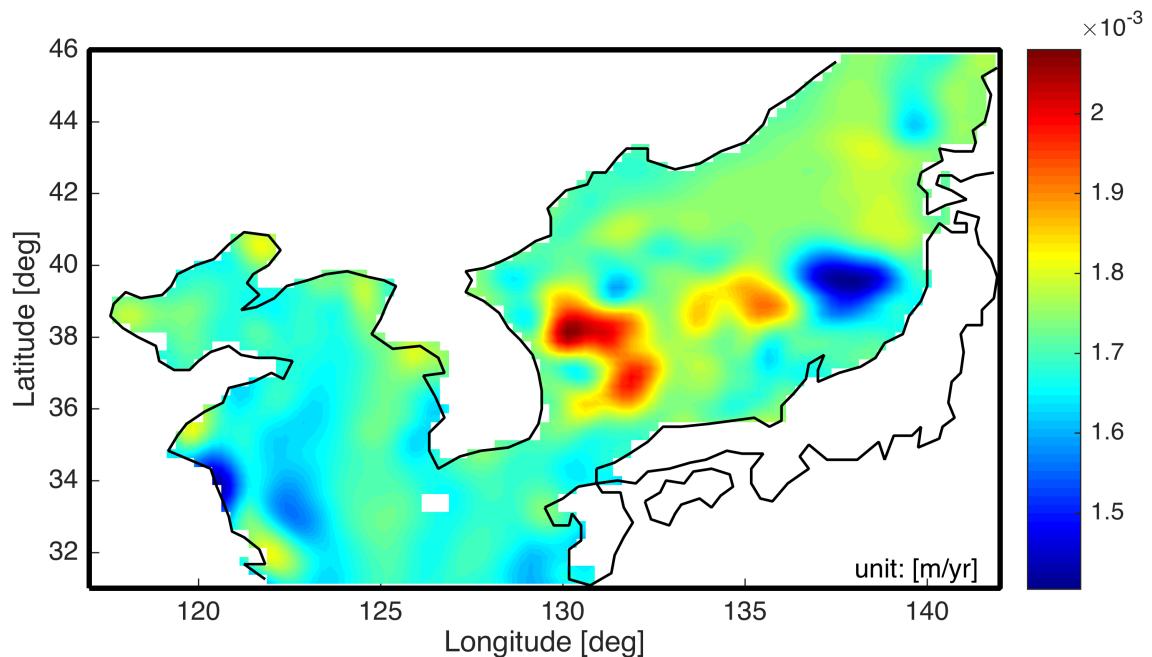

Figure 19. Linear trend map of the ~~reconstructed SLA KP over 1900-2014~~ best reconstruction of current study from 1900 to 2014