Supplementary material: SLR estimation using CEEMDAN near the Korean Peninsula

Young Jin Kim[1,*], Okyu Kwon[2], Hark-Soo Song[1], Jongho Kim[1], Hyuk Kang[1,+]

[1]Public Data Analysis Team, National Institute for Mathematical Sciences, Daejeon 34037, Korea

[2]Research team for transmission dynamics of infectious disease, National Institute for Mathematical Sciences, Daejeon 34037, Korea

[*]kimyoungjin06@nims.re.kr

[+]kh9395@nims.re.kr

S1: Interpretation of Sea Level Rise estimation

The Sea Level Rise (SLR) rates estimated by LR, EEMD, and CEEMDAN show a similar trend as we have seen in the main text. However, LR estimated the rate of SLR relatively large, EEMD estimated the low rate of SLR, and CEEMDAN estimated somewhere in between. Among them, the difference in estimation rates in the three methods is particularly evident in Pohang and Wido. We point to two kinds of issues in EMD-based analysis with real data.

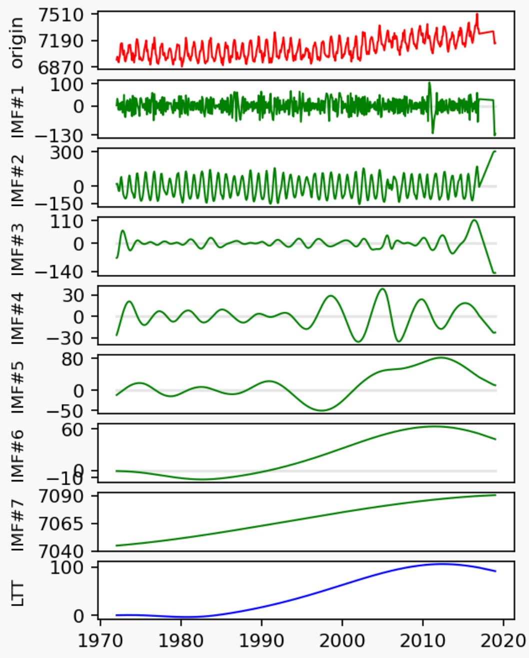

Figure S1.1 **(a)** Original and IMFs at Pohang with CEEMDAN. $ensemble = 1024$. The IMF#7 is estimated SLR. The bottom blue line is sum of long term trend (LTT; #6 and #7).

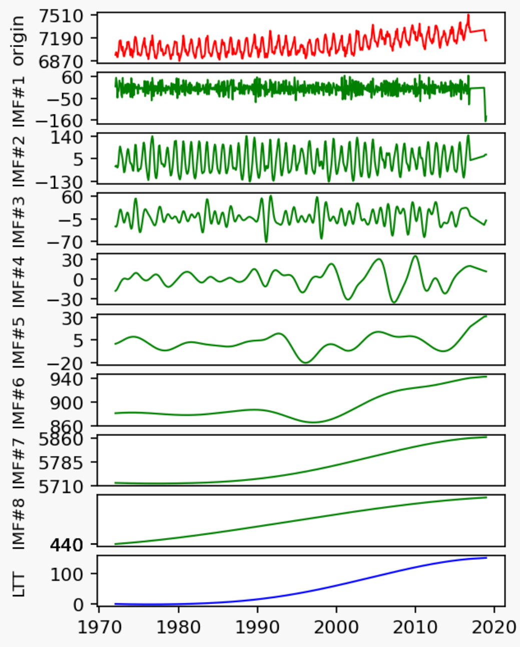

Figure S1.1 **(b)** Original and IMFs at Pohang with EEMD. $ensemble = 1024$. The IMF#7 is estimated SLR. The bottom blue line is sum of long term trend (LTT; #7, and #8).

In the case of Pohang, the SLR started around 1980~1990 (Fig. S1.1, blue line). The Long Term Trend (LTT) is a summation of IMF #6 and #7 with CEEMDAN in Fig. S1.1 **(a)**. In the case of EEMD, the LTT is a summation of IMF #7 and #8. The LTTs are similar except for the last decade. The difference in the last decade originated from the vulnerability of data. The last decade has sparse data, and it can cause a critical difference between the two results in the Pohang case. Additionally, according to the 2020 report of the National Oceanic and Atmospheric Research Institute, it confirmed that the SLR was caused by the erosion of the ground as measured by the Global Navigation Satellite System (GNSS). It seems to have appeared due to the construction instability and strong tide. This part is a limitation of the tide data itself, and they should be considered ground subsidence and thermal expansion according to the purpose. Figure S1.2 shows that LTTs match in the front part except for the end part.

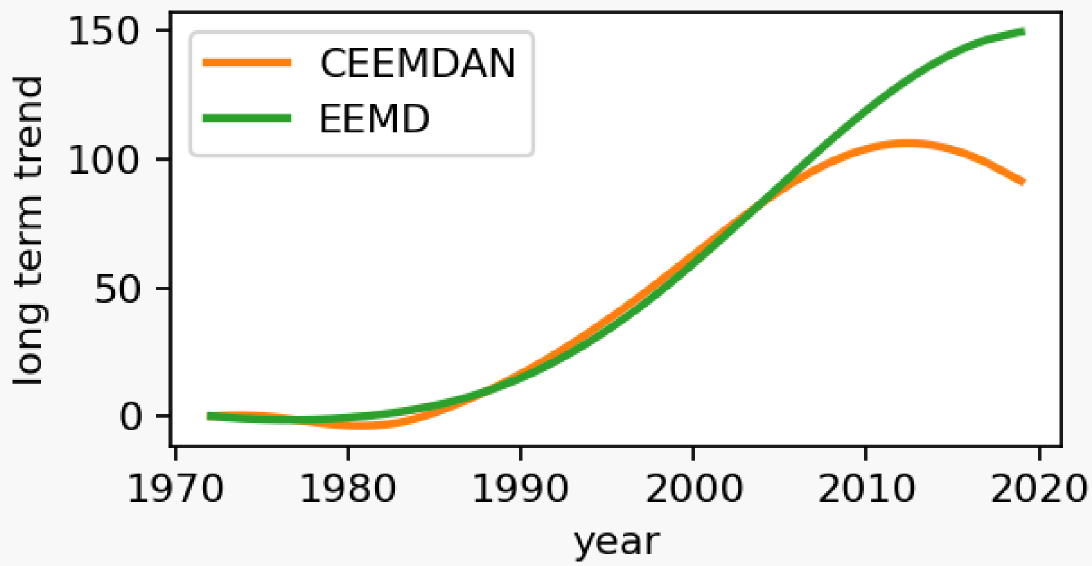

Figure S1.2 Long Term Trend (LTT) of tidal level in Pohang. The LTT of CEEMDAN is summed from #6. The LTT of EEMD is summed from #7.

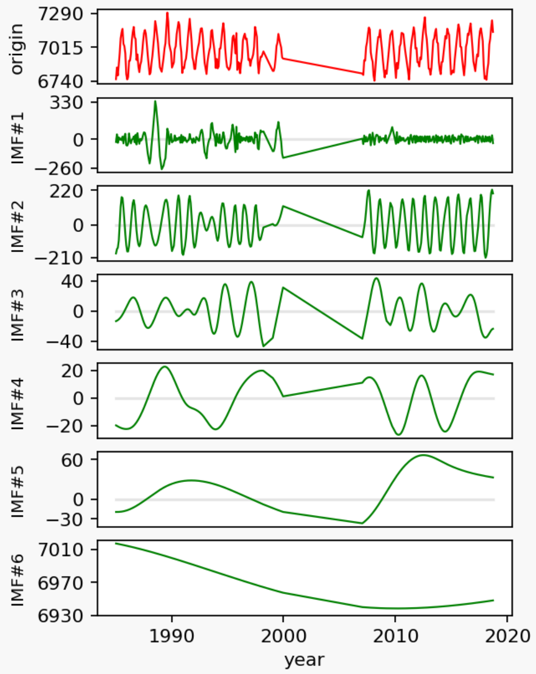

Figure S1.3 **(a)** Original and IMFs at Wido with CEEMDAN. $ensemble = 1024$. The IMF#6 is estimated SLR.

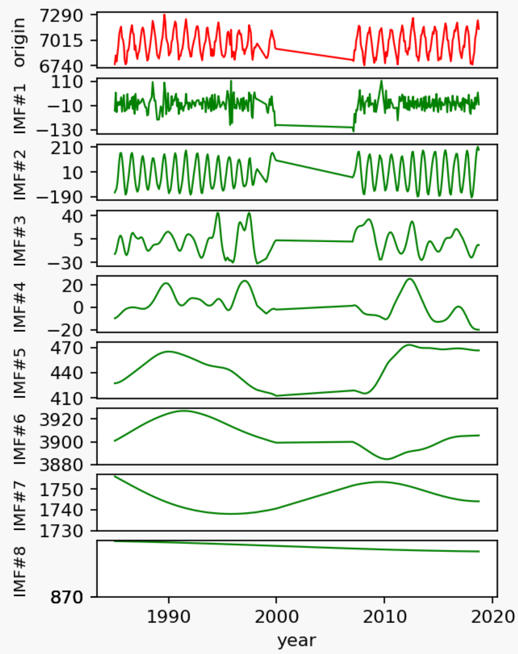

Figure S1.3 **(b)** Original and IMFs at Wido with EEMD. $ensemble = 1024$. The IMF#7 is estimated SLR.

The problem is more remarkable in the case of Wido. Figure S1.3 shows a long-term gap in the middle of the data. Such long gaps can reduce the reliability of the results. In addition, Pohang's IMF #6 in Fig. S1.1 **(a)** and Wido's IMF#5 in Fig. S1.3 **(a)**, both are a confused mode whether as an oscillation term in a long period or an LTT can be interpreted differently depending on the purpose of analysis and its scale in its data length.

S2: Boundary-dependency of the CEEMDAN with overlapped sub-sections

To confirm the boundary-dependency of the CEEMDAN, we apply the CEEMDAN at the overlapped sub-sections with half of the length. The first 50% part (0~50%) is the head part, its mid part (25~75%) is the body part, and the last part (50~100%) is the tail part. Figure S2 shows the sea-level change (SLC) and its derivative of each part. We anticipated the tendency of each part is similar to each other in overlapped sub-sections, and the results show a similar tendency that even does not equivalent exactly. For example, in the Jeju and Mokpo case, the relative magnitude of SLC in Fig. S2 is similar to the original SLC with total length in Fig. 2. In Wido and Incheon, the magnitude of decreasing part is over-estimated than the original SLC with total length in Fig. 2.

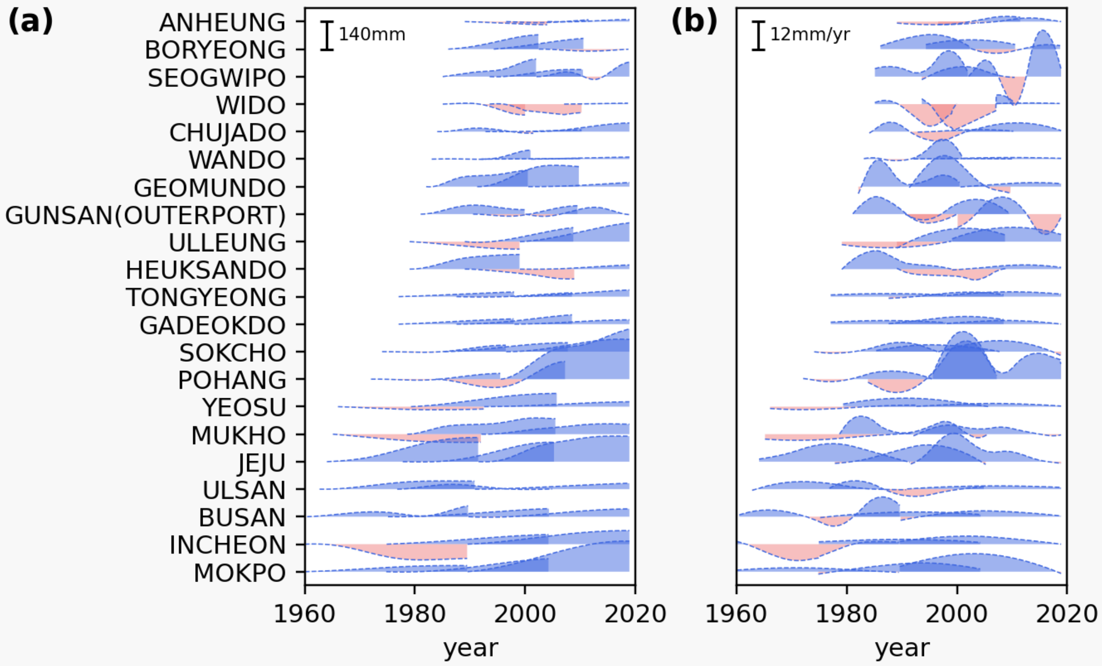

Figure S2. Sea Level Rise(SLR) comparison between tidal stations with head, body, and tail part (half of the original data). **(a)** Cumulative sea-level change(SLC) and **(b)** its derivative: SLR. By the definition of the method, the boundary values of each IMF are close to 0. Although the result depends on data length, the tendency of three parts - head, body, and tail - is similar with each other in the last IMF at least.

S3: Correlation between tidal stations with derivative Sea Level Change

We also look at the correlation map with the derivative of Sea Level Change (SLC) - Figure 4 uses the original tidal level. The initial purpose is to look further into the correlation between the last IMFs. The last IMF of tidal level correlated close to 1 because most of them were monotonic increasing. To examine the monotonic increase in more detail, we looked at the correlation between derivatives. However, in the correlation map using derivatives, only the geographical correlation of the south sea area could be confirmed more clearly as in Figure S3 **(a)** and **(b)**. No significant pattern could be found in the expected last IMF correlation map.

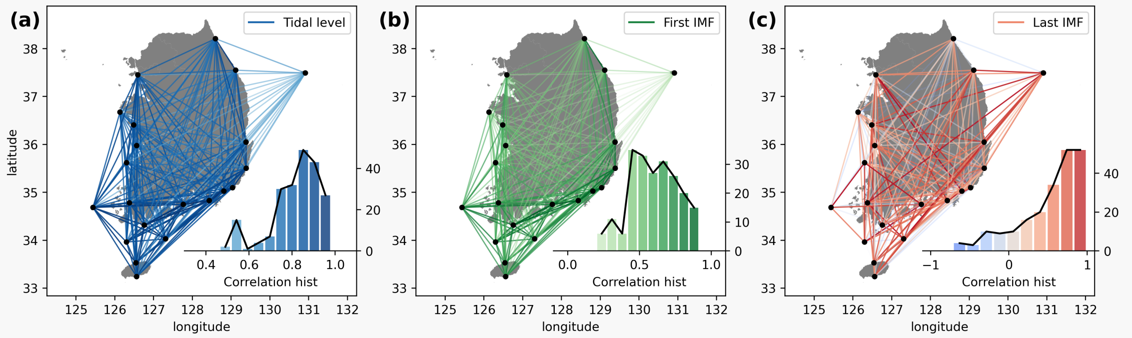

Figure S3. Correlations between tidal stations with derivative of Sea Level Change(SLC). The correlations are obtained with derivative of **(a)** the original monthly tidal level, **(b)** the first IMF, and **(c)** the last IMF. The correlation is calculated using Pearson correlation. **(a, b)** The derivative of original data and the first IMF shows similar results with non-derivative (Figure 4). **(c)** The last IMFs shows non-interpretable pattern. There is an more obvious geographical correlation between close stations in the south sea area with derivative of SLC.

S4: Correlation between SLR and ENSO indices

The most traditional El Niño-Southern Oscillation (ENSO) index is the Southern Oscillation Index (SOI). It is a standardized anomaly of the sea level pressure (SLP) difference between Tahiti and Darwin. And it strongly correlates with Sea Surface Temperature (SST) anomaly indices. We consider two more ENSO indices - Oceanic Niño Index (ONI) and Multivariate ENSO Index (MEI). The ONI is based on SST, and the MEI considers multivariate.

The Oceanic Niño Index (ONI) is one of the ENSO indices developed by the National Oceanic and Atmospheric Administration (NOAA). The ONI is a 3-month running means of Extended Reconstructed Sea Surface Temperature (ERSST) v5 anomalies in the Niño 3.4 region (5N-5S, 120-170W)], based on centered thirty-year base periods updated every five years. If the ONI exceeds a threshold of 0.5C, it is a warm/cold period. And MEI considers weighted anomaly average of 6 meteorological variables in the tropical Pacific including SST and SLP. Figure S4.1 shows the correlation map with ONI and MEI. There is no tendency, but some tidal stations in the Middlewestern area have a higher correlation with all ENSO indices.

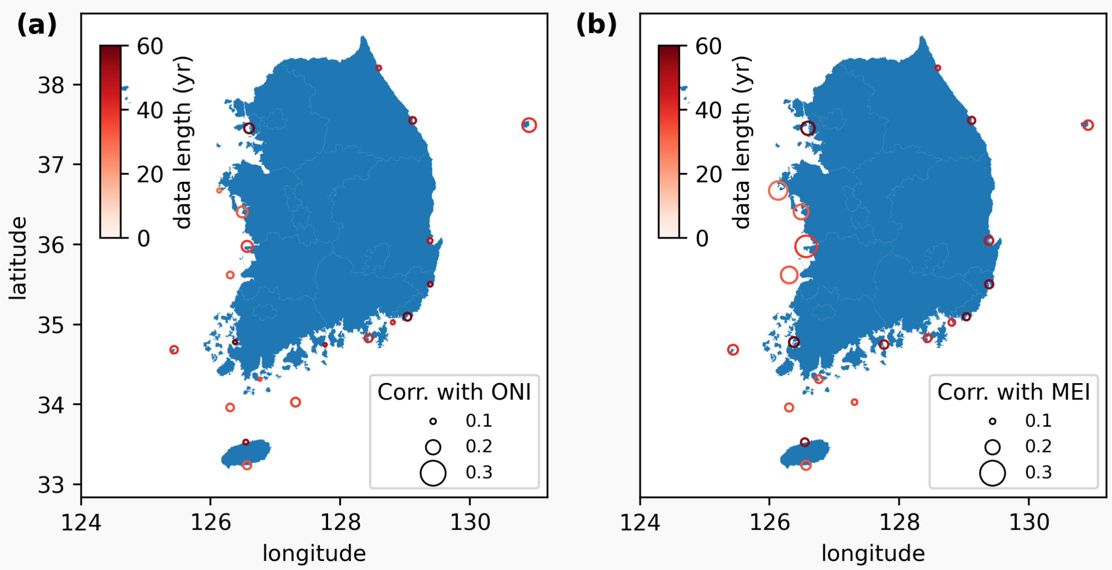

Figure S4.1. Correlation map with ENSO indices. **(a)** a correlation map with Oceanic Niño Index (ONI) and **(b)** Multivariate ENSO Index (MEI). The correlation with ONI is small totally. The correlations with three ENSO indices don't have a common tendency. Nevertheless the Middlewestern area has a higher correlation with ENSO indices.

Figure S4.2 shows the correlations between SLRs and ENSO indices. All SLRs have a strong correlation, but all ENSO indices don't have. Nevertheless, the Middlewestern area most correlates with ENSO indices.

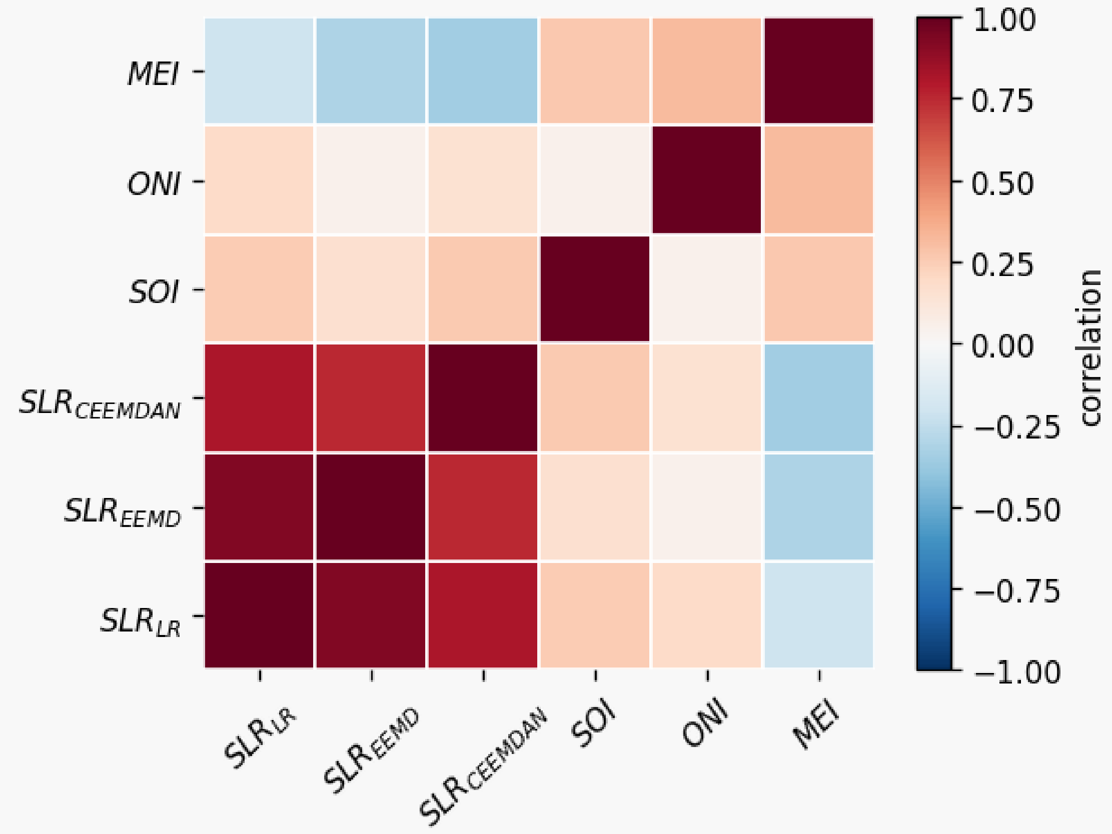

Figure S4.2. Correlation matrix of SLRs and with ENSO indices. The correlations with three ENSO indices don't have a common tendency.

S5: Method of SLR estimation

We estimated the SLR through three methods: linear regression (LR), EEMD, and CEEMDAN. The LR estimated the rate of SLR with fitting to minimize Root Mean Square Error (RMSE) in SLC versus time. There are two reasons why the cubic term was not used. First, the estimated average rate of SLR in a given data interval is not significantly different from linear or cubic. However, the SLR prediction in the future with the cubic term can be inappropriate in some data sets rather than the LR. Therefore, there is no benefit to be gained by using cubic. The EMD-based method has several advantages over regression. First, EMD doesn't require a target function, and it is data-driven. Also, the EMD-based method obtains the SLR in the form of a time series. The time-dependent form is useful to understand an acceleration/deceleration trend of SLR suitable for each data.

EEMD borrows the concept of an ensemble to apply to EMD. The ensemble has the benefit of removing the noise inherent in the original signal by averaging the results of several EMDs performed by adding Gaussian noise to each original signal. However, in this case, the number of total modes may vary with additional noise. To solve this, CEEMDAN shares the averaged residual obtained by removing the averaged IMF in each step. To maintain the effect of added noise, EMD is separately performed on the noise, and the residuals of noise are added on the residual of the matched stage. So, CEEMDAN has the same total number of modes, and it is reasonable that the last mode obtained by CEEMDAN can be SLR. However, the SLR obtained by EEMD uses the second-last mode. It is related to the total number of modes and incomplete decomposition. In SLR estimation with EEMD, some ensembles decompose to one more mode than other ensembles. In this case, the last mode doesn't have enough information about the SLR, and the last mode doesn't have any information about the SLR. Therefore, in the case of EEMD, the second-last mode is equivalent to the SLR.

It can be considered as a type of mode mixing problem. The mode mixing can be found in Fig. S1.1 and S1.3. In the EMD, the upper/lower envelope is found using the local extrema of the raw signal (or residual), and the mean of upper and lower envelopes is set to 0. Therefore, each mode obtained through EMD should have an average of 0 except for the last mode in the methodology. However, Fig. S1.1 **(b)** and S1.3 **(b)** show some modes with non-zero means. The reason is that, as mentioned in the text, CEEMDAN uses a common residual, so the number of modes is almost the same in methodology. However, since EEMD does not share residuals, the number of modes in each ensemble may vary depending on the type of noise. Therefore, since the number of each mode is not the same when averaging, the average is non-zero in the IMF in which the last mode and the other modes are mixed. In addition, since it depends on the number of ensembles, it is important to set it appropriately. We can confirm that the second-last IMF is suitable for SLR at $ens = 1024$ in EEMD statistically from the given data. For smaller ens, even in EEMD, the last IMF may be more suitable for SLR.