Evidence for multidecadal variability in US extreme sea level records

Thomas Wahl¹,² and Don P. Chambers¹

¹College of Marine Science, University of South Florida, St. Petersburg, Florida, USA, ²Research Centre Siegen, University of Siegen, Siegen, Germany

Abstract

We analyze a set of 20 tide gauge records covering the contiguous United States (US) coastline and the period from 1929 to 2013 to identify long-term trends and multidecadal variations in extreme sea levels (ESLs) relative to changes in mean sea level (MSL). Different data sampling and analysis techniques are applied to test the robustness of the results against the selected methodology. Significant but small long-term trends in ESLs above/below MSL are found at individual sites along most coastline stretches, but are mostly confined to the southeast coast and the winter season when storm surges are primarily driven by extratropical cyclones. We identify six regions with broadly coherent and considerable multidecadal ESL variations unrelated to MSL changes. Using a quasi-nonstationary extreme value analysis, we show that the latter would have caused variations in design relevant return water levels (50–200 year return periods) ranging from ~10 cm to as much as 110 cm across the six regions. The results raise questions as to the applicability of the “MSL offset method,” assuming that ESL changes are primarily driven by changes in MSL without allowing for distinct long-term trends or low-frequency variations. Identifying the coherent multidecadal ESL variability is crucial in order to understand the physical driving factors. Ultimately, this information must be included into coastal design and adaptation processes.

1. Introduction

Mean sea level (MSL) research has mostly focused on quantifying and understanding long-term trends and acceleration patterns. The results of this research provide stakeholders and decision makers with scenarios to develop sustainable coastal adaptation strategies. In its 5th Assessment Report (AR5), the Intergovernmental Panel on Climate Change (IPCC) projected that globally MSL may rise by between 28 and 98 cm over the next century [Church et al., 2013]. Locally, the changes will likely deviate strongly from the global average for periods of several decades. These larger local fluctuations caused by vertical land motions [e.g., Wahl et al., 2013] and oceanographic/atmospheric processes are most relevant to coastal adaptation planning [e.g., Nicholls et al., 2014], especially over the next half-century when the rate of global MSL (GMSL) rise will likely still be low enough that local variations dominate [Church et al., 2013]. Several studies have identified significant multidecadal variations in MSL time series, which were attributed to both internal climate variability and anthropogenic interference—for the US coast [see e.g., Bromirski et al., 2011; Calafat and Chambers, 2013; Chambers et al., 2012; Ezer, 2013; Ezer et al., 2013; Hamlington et al., 2014; Merrifield and Maltrud, 2011]. Such variations superimposed onto long-term trends are important because they hamper the early detection of accelerated sea level rise, may facilitate the exceedance of certain critical thresholds much earlier, or can lead to larger changes toward the end of the century than currently expected [e.g., Dangendorf et al., 2014; Haigh et al., 2014a].

However, with the exception of small island states (e.g., in the tropical Pacific, Indian Ocean, or Caribbean), extreme sea level (ESL) events associated with storm surges primarily endanger coastal communities and fragile ecosystems rather than the immediate impacts of MSL rise. Therefore, any changes in ESL unrelated to MSL rise need to be accounted for in order to plan adaptation measures such as dikes, barriers, and sea walls in a way that high safety standards are guaranteed throughout the anticipated lifetime of the project (often 50–100 years or more).

Several studies have compared long-term trends in ESL and MSL, with the most comprehensive ones at global scale by Woodworth and Blackman [2004] and Menéndez and Woodworth [2010]. They used quasi-global tide
gauge (TG) data sets—which limited the majority of analyses to the last few decades—and concluded that changes in ESLs were, with some exemptions, generally coherent with and primarily driven by changes in MSL. Therefore, what we refer to here as the "MSL offset method" has been widely applied in the past; i.e., in order to infer information on future return water levels (RWLs) the expected MSL rise (using for example IPCC projections adjusted for local conditions) is added linearly to present-day RWLs [e.g., Hunter, 2012; Hunter et al., 2013; Pugh and Woodworth, 2014; Tebaldi et al., 2012; Obeysekera and Park, 2013]. In this approach multidecadal variability in both MSL and ESL is ignored, and while the observed MSL variability along the US Atlantic and Pacific coasts has been studied extensively and is relatively well understood (see references provided above), changes in ESL different to those in MSL (long-term trends and especially multidecadal variability) have received much less attention.

A number of authors have used long tide gauge records to explore changes in storminess in the US [e.g., Bromirski et al., 2003; Grinsted et al., 2012, 2013; Kennedy et al., 2007; Park et al., 2010; Sweet and Zervas, 2011; Thompson et al., 2013; Zhang et al., 2000] and the links to climate variability, but they usually focused on (i) particular sites or coastline stretches; (ii) either winter storms or tropical cyclone activity; (iii) different parameters such as storm counts, storm surge durations, daily sea level anomalies, etc., instead of ESL/RWL changes; and (iv) considered a wide range of different approaches for the (statistical) analysis leading to results which are hardly comparable across studies. Most recently and for New York City as an example, Talke et al. [2014] compiled a TG record consisting of high and low waters dating back to the mid-nineteenth century. They detected a significant trend in storm tides and—using extreme value analysis—associated changes in the 5 and 10 year RWLs related to fluctuations in the North Atlantic Oscillation (NAO).

To date, no rigorous assessment of the observed changes in ESLs (and associated RWLs) along the entire US coastline has been conducted to identify nonstationarity patterns with potentially significant implications for coastal adaptation planning. In this paper, we systematically analyze the longest TG records from the contiguous US coastline with the three main objectives of (i) searching for ESL long-term trends above/below relative MSL trends; (ii) identifying regions with coherent ESL variability; and (iii) quantifying multidecadal ESL fluctuations relevant to coastal design and adaptation.

Detecting long-term trends and variations in ESL time series is not as straightforward as identifying such signals from MSL time series. Therefore, throughout the paper we employ a wide range of different analysis techniques previously used for this purpose and perform sensitivity tests to assure that the presented results are robust against the applied methodology. The remainder of the paper is organized as follows: the

![Figure 1](image-url). Figure 1. (a) Data availability at the 20 TG sites for the nineteenth and twentieth century: years with less than 75% of hourly data are omitted; (b) locations of the TG sites and completeness of the records for the 1929–2013 period (after years with less than 75% of data were removed).
TG data sets are introduced in section 2, the applied methods are described in section 3, results are presented and discussed in section 4, and concluding remarks are summarized in section 5.

2. Tide Gauge Records and Data Processing

We use hourly water level observations from 20 TGs covering the contiguous US coastline (Figure 1a); two TGs, Victoria and Halifax, are located in southwest and southeast Canada, respectively. Records were chosen to cover the common time period from 1929 to 2013 (Figure 1b); this period was selected as a trade-off between temporal and spatial data availability—the covered common time period should be as long as possible and the tide gauges that meet the criteria should cover all US coastlines. When comparing linear trends across sites we use data from 1929 to 2013; earlier data are used where available for the remainder of analyses. The majority of data was downloaded from the University of Hawaii Sea Level Center (UHSLC) database (http://uhslc.soest.hawaii.edu/data/download/rq); several sites (Baltimore, Sewells Point, and Seattle) were not included in this database but long records were available from the website of the National Oceanic and Atmospheric Administration (NOAA; http://tidesandcurrents.noaa.gov/). For New York City, we also use NOAA data of the Battery TG as it was more complete than the UHSLC data. When the analysis was conducted, many of the UHSLC time series ended in 2012; they were extended to December 2013 with NOAA data after checking for datum consistency based on overlapping periods.

San Francisco provides the longest record starting in 1854 [Bromirski et al., 2003; Smith, 2002], but for quality reasons we only use data since 1858. The time series of Mayport started in 1928 and ended in 2000; it was filled with data from Fernandina Beach (located only 31 km north of Mayport) for the periods from 1897 to 1928 and 2001 to 2013. For the overlapping period from 1985 to 2000 the correlation between the hourly records is 0.99. Sensitivity tests showed that too many (or too large) data gaps within a year may distort the results from the tidal analysis (see below) and it also becomes more likely that extreme events (which is what we are most interested in) are missing. Therefore, years with less than 75% of data were omitted from the analysis.

Many of the west coast sites exhibited obvious outliers caused by tsunamis. Since we are interested in identifying ESL changes related only to storminess and climate change, we removed data from the following (and surrounding) days for all west coast TG records: 23 May 1960 (Great Chilean Earthquake), 28 March 1964 (Great Alaska earthquake), and 11/12 March 2011 (2011 Tohoku earthquake in Japan). We note that measured water levels were likely affected by tsunamis at other times throughout the last century, but did not always result in extraordinary extreme events and can therefore be neglected for the purpose of the present study.

Based on the hourly observations we compile different data subsets which are used to explore long-term trends and multidecadal variability in ESLs as described in section 3. Major parts of the US coastline are susceptible to storm surges driven by extratropical and tropical cyclones, which are different in their genesis. Our long-term goal (in a future study) is to explore the links between large-scale climate change/variability and the ESL trends/fluctuations identified in the present study. Hence, in order to keep variations that are associated with different driving mechanisms separated, we perform all analyses separately for summer and winter half-years that we call the summer and winter seasons. We follow Thompson et al. [2013] in defining the winter season where extratropical cyclones dominate the storm surge climate from November through April; accordingly, the summer season where tropical cyclones dominate ranges from May to October.

First, and similar to Woodworth and Blackman [2002, 2004], we use the hourly data to calculate the 99.5th quantile values separately for summer and winter of each year. Then we determine the summer and winter median values (the median is preferred over the mean as it is more robust against outliers) for each year and subtract those from the 99.5th quantiles to account for MSL rise and intra/interannual variability. Finally, we perform a year-by-year tidal analysis with the Matlab package T_Tide [Pawlowicz et al., 2002] (accounting for 67 constituents including annual and semiannual cycles) and derive the 99.5th quantiles of the astronomical tidal water levels for summer and winter; subtracting these from the 99.5th quantiles of the observed water levels leads to a data set that is free of MSL and tidal influences. We refer to this data set as T995, and then derive the same for the 99.9th quantile (T999).

Second, we calculate skew surge time series for each TG. Skew surge represents the difference between an observed high water and the closest predicted astronomical high water [e.g., de Vries et al., 1995; Horsburgh and Wilson, 2007]; we use the same tidal prediction as described above. The skew surge approach is preferred over using nontidal residuals, which are more susceptible to timing errors in the observed water.
levels and/or tidal prediction. We identify the highest skew surge values for each season in a year (SKmax) and calculate seasonal quantile time series of skew surge on a year-by-year basis. We use the 98.75th quantile of skew surge (corresponds to the level of the four highest skew surge values in a season; the data set is called SK9875). Sensitivity tests showed that the two data sets SK9875 and T995 have similar standard deviations, and hence comparing trends and/or variations derived from them (even if they still comprise very different parameters) is more meaningful then when using the 99.5th quantile of skew surge (which would only include the highest two values leading to a much larger standard deviation).

Third, we take the raw hourly data, remove a 30 day running median (i.e., the median derived from 30 day windows shifted 1 h per time step) to account for MSL rise and variability, and identify the highest values in the summer and winter seasons of each year (Tmax). This data set (as well as SKmax) can be assumed to consist of independent events, whereas in the quantile approach the 22 highest hourly values (or four highest skew surge values) in a season may theoretically stem from the same event. The five distinct data sets T995, T999, Tmax, SK9875, and SKmax, all represent changes in ESLs and are used at different analysis stages to assure that the presented results are not distorted by the decision of using a particular data set or definition of “extreme events.”

3. Methods

3.1. Long-Term Trends

We start with searching for ESL long-term trends above or below changes in relative MSL. Therefore, we calculate linear trends for the common time period from 1929 to 2013 (which is covered by all 20 TGs) of the seasonal 99.5th quantile time series derived from the hourly records—before and after MSL and tidal influences are removed (i.e., T995)—and from the SK9875 data set. The ordinary least squares (OLS) method is used to estimate the parameters of the linear regression model. We use a Durbin-Watson test to search for autocorrelation in the residuals [Durbin and Watson, 1950, 1951]. Results show that after MSL and tidal influence are removed the null hypothesis that residuals are uncorrelated cannot be rejected (90% confidence) at all 20 sites for both seasons; however, if time series are affected by MSL and tidal changes we find significant serial correlation at the majority of sites. In such cases OLS standard errors become invalid and need to be corrected. Therefore, confidence intervals are derived here with the Newey-West Estimator (NWE) [Newey and West, 1987]. The NWE does not only account for serial correlation but also heteroscedasticity in the data, i.e., changes in the variance over time. The lag length for the NWE is calculated here as $L = 4(N/100)^{2/9}$ [Newey and West, 1994], where N represents the number of observations.

Next, we want to make sure that the results of this analysis step are unaffected by the choice of using seasonal quantiles for each year and more importantly by focusing only on a specific quantile. For example, the 99.5th quantile time series may exhibit significant trends whereas other high quantile time series do not. Thus, we employ the concept of quantile regression to identify trends in high or low quantiles significantly different to the 50th quantile (representative of the MSL trend). For the theoretical background of the method see Koenker [2005] and Chandler and Scott [2011]; applications to sea level data can be found for example in Barbosa [2008], Barbosa and Madsen [2012], Donner et al. [2012], and Ribeiro et al. [2014]. Here we apply linear quantile regression to all TGs (considering the entire record set, see Figure 1a) where we find significant trends in either winter or summer from analyzing the T995 and SK9875 data sets as described in the previous paragraph.

Finally, we test how sensitive the results are to the chosen time period. To do so, we calculate linear trends from the T995 data sets of selected seasons and TGs using a combination of a running and growing window approach, i.e., trends are calculated for all window lengths >20 years (trends from shorter records are not relevant in this context) and for all start years [Haigh et al., 2014a; Jevrejeva et al., 2006; Scafetta, 2014]. This confirms whether the trends derived for 1929–2013 are confined to this particular period, or whether similar trends are obtained when shorter/longer time spans are analyzed. The results for shorter window lengths (e.g., 20–40 years) also provide first indication of whether the time series exhibit significant decadal variations or not. Significance of the trends is assessed again with the NWE.

3.2. Variability

In the introduction, we mentioned that multidecadal variations superimposed onto MSL or ESL long-term trends may have significant implications for coastal adaptation. Therefore, we focus next on identifying coastline stretches with coherent variability and on detecting the relevant low-frequency fluctuations.
We first calculate linear correlation coefficients between all TGs using the T995 and SK9875 data sets to identify regions with broadly coherent variability. Significance is assessed with a t test at the 90% confidence level. We then use an extreme value analysis approach to identify, for each region, the season that leads to the highest RWLs; if RWLs are higher in summer tropical events dominate, if RWLs are higher in winter extratropical events are more important. Since we are interested in revealing long-term trends and low-frequency fluctuations in ESLs we employ a nonstationary extreme value analysis that allows us to identify temporal changes in the distribution parameters (and associated RWLs). Traditionally, nonstationary extreme value approaches have been used where temporal changes in distribution parameters (first of all location and scale) were modeled by fitting parametric functions, i.e., (linear) trends to account for sea level rise, and harmonics to capture semiannual, annual, and longer-term (18.6 nodal and quasi 4.4 perigean) cycles. Climate indices were also used as covariates to account for the interannual variations due to large-scale climate variability [e.g., Menéndez and Woodworth, 2010]. Here we employ a nonparametric approach which we refer to as quasi-nonstationary extreme value analysis (nEVA) [e.g., Mudersbach and Jensen, 2010; Talke et al., 2014]: the Generalized Extreme Value (GEV) distribution [e.g., Coles, 2001; Mudelsee, 2014] is fitted to moving 37 year windows (i.e., twice the 18.6 year nodal cycle) of the Tmax data set (which is free of MSL trends and seasonal cycles); a one-sided Kolmogorov-Smirnov (KS) test [e.g., Massey, 1951] was employed as goodness of fit test and passed by the GEV (90% confidence) at all sites for both seasons. We use the 37 extreme values that we find in each window to estimate the GEV distribution parameters with the Maximum Likelihood method, and then calculate the 100 year RWL for each window (i.e., the water level that is exceeded with a 1% chance throughout a year, or here, a season). Note that inferring the 100 year RWL from 37 years of data is well within the accepted limits of extrapolation, usually quoted as 3–4 times the record length [e.g., Pugh, 2004]. In the analysis we allow the location and scale parameters to change over time, but follow for example, Coles [2001], Menéndez and Woodworth [2010], or Mudersbach and Jensen [2010] in assuming that the sensitive shape parameter is stationary over the entire record length. When the 100 year RWL values obtained this way are consistently (i.e., for all 37 year windows) higher in summer or winter, this is the relevant season. Otherwise, or when RWL values are consistently higher in one season but only barely, both seasons are deemed relevant and considered for further investigation. A sensitivity test (not shown here) using the Generalized Pareto distribution (GPD) fitted to high threshold exceedances [e.g., Coles, 2001] of running windows instead of the GEV fitted to seasonal maxima, revealed similar results in terms of trends/low-frequency variations. The GEV approach is preferred here as it avoids choosing (varying) thresholds and decluster times (to assure independency) at all TGs and for the two seasons individually.

Based on the outcome of this analysis step we can compare the nEVA results derived from all TGs within the previously identified regions and for the relevant seasons. This will confirm whether or not multidecadal ESL variations within the regions are also coherent; the correlation analysis described above mainly captures interannual variability.

Next, and similarly to the long-term trend analyses, we want to ensure that the results are robust against the selected methodology (i.e., fitting the GEV distribution to 37 year running windows of the Tmax data set). Thus, for selected sites (representative of the variability in a particular region) we perform the same nEVA with the SKmax data set and compare the results to low-pass filtered versions (37 year moving averages) of the T999 and SKmax data sets. Since the location and scale parameters are allowed to fluctuate in the nEVA, changes in both contribute to the overall changes in the 100 year RWL. We compare the two parameters derived for each of the 37 year windows in order to identify coherent and especially incoherent changes between them, which need to be accounted for in order to understand how climate affects the observed RWL changes.

Finally, after merely focusing on the 100 year RWL, which is often the most relevant variable for design purposes, we identify the magnitude of decadal variations across a wider range of RWLs. This is done by analyzing the entire distribution functions (extrapolated up to the 200 year events) derived for the Tmax data and all consecutive 37 year periods.

4. Results

4.1. Long-Term Trends

The seasonal 99.5th quantile time series before and after MSL and tidal influences are removed are shown for eight selected sites and seasons (the same that are used later in the study to represent particular regions
and relevant seasons) in Figure 2. Outliers are evident for sites and seasons where strong hurricanes (Galveston and Baltimore) or winter storm surges (Atlantic City) occurred. We find significant trends (90% confidence) in the 99.5th quantile time series of observed hourly values for the 1929–2013 period at virtually all study sites and for both seasons (Figures 3a and 3b); there are a few exceptions in the northwest where trends are insignificant (shown as black dots).

The spatial pattern reflects the prevailing long-term changes in relative MSL as a result of GMSL rise, regional deviations from it due to oceanographic and atmospheric processes, and vertical land motions (subsidence or uplift) due to glacial isotactic adjustment (GIA), tectonic activity, and/or sediment compaction (e.g., resulting from extraction of ground fluids or drainage of susceptible soils). The estimated trends vary greatly, between 0 and 7 mm yr\(^{-1}\), and are generally larger in both seasons along the Atlantic and northern Gulf of Mexico coastlines compared to the Pacific coast and the west coast of Florida. Along large coastline stretches, trends are in the order of 3–4 mm yr\(^{-1}\) (in both seasons) which is higher than what is considered the best estimate of GMSL rise during the twentieth century (~1.7 mm yr\(^{-1}\)) [Church et al., 2013]. This highlights the importance of including regional and local effects into design guidelines to account for potential future ESL and MSL changes.
When the seasonal medians, and hence the MSL influences, are removed from the time series the trends become much smaller and insignificant at the majority of sites (Figures 3c and 3d). However, several TGs on the west and east coasts still exhibit significant trends of up to 1.5 mm yr$^{-1}$ which are unrelated to relative MSL changes; for Mayport we find a significant negative trend of $-0.5$ mm yr$^{-1}$ for the summer season, when extreme water levels are primarily governed by tropical cyclones. Removing the tidal influence (Figures 3e and 3f; i.e., data set T995) again changes the spatial distribution of significant trends, suggesting that changes in the tidal constituents (e.g., due to changes in the tidal potential, sea level rise, or natural/anthropogenic morphological changes) and associated effects on extreme events play an important role when assessing ESL changes.

However, it has to be considered that whether or not a trend is deemed statistically significant does not only depend on the slope, but also on the variability exhibited by the time series under investigation. Here the variability is reduced considerably by removing the tidal influence. A closer look reveals that the differences at TGs San Francisco and Eastport in summer (i.e., comparing Figures 3c and 3e) and at TGs on the west and northeast coasts in winter (i.e., comparing Figures 3d and 3f) can indeed be attributed to trends in the quantile time series of astronomical tidal predictions. In the southeast, on the other hand, trends do not change much, but become significant only because of the reduced variability. Since we are primarily interested in identifying changes in the meteorological component of ESLs, we do not expand the discussion on changes in astronomical tides here; more detailed explanations and results from global assessments and for the US coast can be found for example in Jay [2009], Müller [2011], Müller et al. [2011], Ray [2006, 2009], or Woodworth [2010].
Finally, comparing the T995 (Figures 3e and 3f) and SK9875 (Figures 3g and 3h) data sets uncovers some differences. This is not surprising since the data sets are different in the way they are developed (although both represent “extreme event indices”). In general, the quantile time series derived from hourly values are less susceptible to errors. The skew surge approach reduces the risk (compared to using nontidal residuals) that timing errors in the data and/or tidal prediction distort the results; however, if larger timing errors occur the results are still affected. Furthermore, calculating skew surge values can be challenging in regions with complex tidal characteristics—identifying observed tidal high waters is for example not always straightforward when the time series are noisy. Finally, a high skew surge that occurred around neap tide may not show up as an extreme event in the total water level time series, and hence in the T995 data set.

Menéndez and Woodworth [2010] found similar differences from assessing trends in quantile time series derived from hourly total still water levels and nontidal residuals (instead of skew surges; see Figure 5 in their paper).

It is noteworthy that in the SK9875 data set, winter trends in the southeast are also positive, but not significant at the 90% confidence level. Thus, taking the results from both approaches into account this is the only region and season where we find a larger number of neighboring TGs exhibiting positive (and partly significant) long-term ESL trends in the order of 0.5 mm yr$^{-1}$ unrelated to MSL or tidal changes. Other individual sites (and seasons) where both approaches identify significant trends are Pensacola (summer) and La Jolla (summer). These two and all other sites where we find significant trends in at least one of the two data sets and seasons are considered for the next analysis step.

Figure 4. Linear quantile regression of observed hourly values for selected TGs (sites shown have significant trends in Figures 2e–2h) and winter (blue) and summer (red) seasons; shaded bands denote standard errors of the 50th quantile slopes against which the higher quantile slopes are compared (with standard errors shown as vertical bars). Note the different scaling of the y axes across rows.
Linear quantile regression is applied to hourly data to explore the consistency of the T995 results across a wider range of high quantiles (Figure 4). Results are shown for the 50th quantile (representing changes in MSL) and for the 90th, 95th, 97.5th, 99th, and 99.5th quantiles (with standard errors derived via bootstrapping). All hourly records were linearly detrended before quantile regression was applied to summer and winter seasons separately; therefore, the slopes derived for the 50th quantiles are not exactly zero. The results are broadly consistent with those derived from the previous analysis, but there are some sites and seasons where significant changes above or below MSL become evident that have not been detected with the linear trend analysis of seasonal quantiles (e.g., for winter in San Francisco, La Jolla, and New York). The results also highlight that (with very few exceptions) at sites where trends are significant in the T995 and SK9875 data sets (i.e., 99.5th and 98.75th quantiles, respectively), the slopes for other high quantiles are also significantly different from those in MSL.

Finally, we test by how much the detected trends vary when other time spans than the 1929–2013 period are considered. We calculate trends in T995 for all window lengths (>20 years) and start dates, and the results of this are shown for La Jolla (summer) and Mayport (winter) in Figure 5. For both examples trends become consistently significant when the underlying data sets exceed a certain length. For La Jolla the threshold is ~60 years. Below that, i.e., window length of 40–60 years, trends are significant in approximately half of the cases. For Mayport the threshold is higher and at least ~80 years of data are required for trends to be consistently significant. For shorter time series (70–80 years) trends tend to be significant only when the latter part of the entire record is considered. The results for both sites also highlight that for data lengths <40 years trends vary considerably, between −1.5 to +1.5 mm yr⁻¹ in La Jolla and −4 to +4 mm yr⁻¹ in Mayport, suggesting that the ESL time series exhibit substantial multidecadal fluctuations which are not captured when a simple linear model is applied to the entire data sets.

4.2. Variability
The results presented in Figure 5 confirm the existence of significant variability in ESL time series, but only for two selected sites and seasons. In order to better understand the spatial coherency of these temporal variations across the study area, we calculate linear correlation coefficients between all sites using the T995 and SK9875 data sets (Figure 6). Not surprisingly, the highest correlation exists near the diagonal between neighboring stations.

However, the correlation matrices also exhibit common (spatial) features that allow the definition of regions with broadly coherent variability. Carefully considering the results from all subpanels of Figure 6, we identify the following six regions: North Pacific coast (NP; TGs 1–3), South Pacific coast (SP; TGs 4–7), Gulf of Mexico
Scattered significant correlation coefficients are also found over much larger spatial scales (especially in Figure 6d), including between east and west coast TGs; the coefficients are however small and these correlations are not further explored or discussed here.

The correlation analysis is conducted based on the seasonal percentile time series and therefore captures mainly the interannual variability. The coherence of multidecadal fluctuations within the identified six subregions is assessed with the nEVA approach described in the section 3, i.e., the GEV is fitted to 37 year running windows of the Tmax data set and 100 year RWLs are calculated. To prioritize the data that need to be further explored in order to capture only the relevant temporal variations, we first identify the seasons that dominate the storm surge climate in the six regions. To do this, we compare the 100 year RWL values derived with the running window approach for summer and winter seasons at all 20 TGs. If winter RWL values are consistently higher than summer RWLs this is the relevant season, and vice versa; if the RWL time series intersect both seasons are deemed relevant.

Winter storm surges dominate in the NP, SP, and NA regions, while storm surges associated with tropical summer storms are more relevant in the GOM. In the SA and MA regions, the summer season is relevant at some sites and both seasons at others (Figure 7). In Key West and Charleston summer RWL values are always higher, although in some instances only by a few centimeters. Therefore, from here on we focus on the winter season for the NP, SP, and NA regions, the summer season for the GOM coast, and both seasons for the SA and MA regions.
Applying the nEVA to all Tmax time series for the relevant seasons (Eastport is omitted at this stage because of too many data gaps) reveals that the multidecadal variations within the six regions are broadly coherent (Figure 8), confirming the results from the correlation analysis; note that in Figure 8 the mean has been removed from the RWL time series as we are interested in identifying common temporal variations instead of comparing absolute values across sites.

However, the amplitudes of the observed changes vary largely across regions, from ~10 cm in winter for the NP, SP, and SA regions, to up to almost 100 cm for the other regions and relevant seasons; changes are generally larger in summer. In some instances two or more records within a region show very similar fluctuations while one record exhibits a very different signal (gray dashed lines in Figure 8). This is likely a result of local effects such as freshwater inflow from rivers (e.g., at TG Astoria) or changes in the bathymetry superimposed onto climate related ESL.
changes, or due to trends in the astronomical tides which are still included in the Tmax data set. The SA coast (summer) is the only region where the RWL time series from all TGs are distinctly different and we cannot identify a coherent signal that could be attributed to climate variability. However, we also note that the range of variability is small compared to some of the other regions. In this case alternative approaches merging the data from all sites in the particular area into an index time series [e.g., Grinsted et al., 2012] may be more appropriate to identify the prevailing decadal RWL variations associated with changes in storminess.

Similarly to the long-term trend analyses, we want to make sure that the results presented in Figure 8 are robust against the applied analysis and data selection technique(s). Therefore, we pick one TG from each of the six regions (and the relevant season) and assume that it is representative of the changes along this particular coastline stretch. This is justified because we find at least two (in most cases three) sites with coherent changes in each of the regions. As discussed above the SA coast is the only exception; here we use the data from the long Mayport record for further exploration.

The results from applying different analysis techniques (i.e., nEVA and low-pass filtering) to the different data sets (Tmax, SKmax, and T999) are consistent (Figure 9); note that the time series were normalized (i.e., mean subtracted and divided by standard deviation) to allow direct comparison. For most sites all four, but at least three of the time series show coherent temporal changes and the RWLs derived from the Tmax data

Figure 9. Variations in 100 year RWLs at selected TG sites being representative of a particular region and for the relevant season derived with the nEVA approach applied to the Tmax (red) and SKmax data sets (green), and 37 year running averages of the T999 (blue) and SKmax (black) data. Shaded bands represent standard errors from the nEVA approach. Data sets have been normalized for comparison purposes (see text).
set are always among those. The largest discrepancies are found for San Francisco, where uncertainties are large and the temporal patterns at multidecadal time scales are nonetheless broadly coherent. Hence, we are confident that the results from the nEVA applied to the Tmax data sets are robust against the employed analysis technique.

We note that there are other (more complex) models available to explore trends, accelerations, or decadal fluctuations in time series. Some of them have already been used in sea level science (mostly to analyze MSL) such as singular system analysis (SSA) [e.g., Wahl et al., 2010, 2011], the concept of sea level rise differences (SLRD) [e.g., Sallenger et al., 2012; Calafat and Chambers, 2013], or Empirical Mode Decomposition (EMD) [e.g., Ezer, 2013; Ezer et al., 2013] (which according to Chambers [2014] has to be applied with extreme caution when analyzing sea level data). The results obtained with the four independent analysis techniques employed here are consistent. Therefore, adding additional methods (and more complexity) to the analysis would very likely not reveal any new relevant information.

The RWL time series depicted in Figures 8 and 9 were derived with the GEV, which consists of three parameters. Two of those (i.e., location and scale parameters) are allowed to fluctuate within the nEVA, while the third one (the shape parameter) is assumed stationary over the entire record lengths. Hence, the RWL changes can be due to changes in the location parameter, being representative of the mean, or the scale parameter, being representative of the standard deviation, or both. Comparing the two parameters derived for the consecutive 37 year periods highlights that at the majority of sites both exhibit almost identical temporal variations (Figure 10). Discrepancies are found for Mayport and San Francisco, representing the SP

---

**Figure 10.** Changes in scale (blue) and location (red) parameters from the nEVA (37 year running windows) at the same TG sites shown in Figure 8; shaded bands denote the uncertainty range of the stationary approach (i.e., GEV is fitted to the entire record).
and SA regions, respectively. Selecting other sites from the two regions confirmed that temporal changes in location and scale parameters were not consistent along these coastline stretches. This suggests the existence of spatially more complicated links between ESL changes and large scale climate variability. This has to be accounted for when attempting to explain the mechanisms underlying the overall variations in RWL time series. The results also show that at each individual site both the scale and location parameter time series deviated significantly (90% confidence) at some point during the last century from the results obtained under the stationary assumption (i.e., GEV fitted to the entire records).

Thus far, we have only focused on the 100 year RWL, which is often most relevant to coastal managers and engineers in the infrastructure design context. However, depending on the project type and location, lower or much higher RWL values are relevant (e.g., to protect critical infrastructure). To explore the magnitude of variability across a wider range of (relevant) RWLs, the entire distribution functions derived with the nEVA for consecutive 37 year periods are plotted as gray lines in Figure 11 and are compared to the results derived under the stationary assumption (blue lines; i.e., GEV fitted to the entire record length). Again, the nEVA results deviate significantly (90% confidence) from the stationary assumption at all eight study sites; note that uncertainties in the shape parameter were not accounted for to construct confidence intervals in order to allow comparison to the nEVA results (where the shape parameters is assumed stationary). For the 50, 100, and 200 year RWLS the magnitudes of the low-frequency variations are also indicated. They range from ~10 cm for the 50 year RWL to 110 cm for the 200 year RWL, with the 100 year RWL in between

Figure 11. GEV distributions derived with the 37 year running window approach (gray) and under the stationary assumption (i.e., GEV fitted to entire record) with 90% confidence intervals (blue); black vertical lines and numbers denote the magnitude of variability in 50, 100, and 200 year return water levels derived with the nEVA. RWLS are referred to MSL datum (for the 1983–2001 epoch).
exhibiting changes in the order of 10–80 cm. We note that the 200 year RWL is slightly beyond what is commonly thought to be the maximum allowable extrapolation period when using extreme value analysis (i.e., 3–4 times the time series length; here we use 37 year windows) [e.g., Pugh, 2004]. However, existing guidelines often oblige planners and engineers to provide extreme water level estimates for much longer return periods—e.g., $10^4$ (coastal protection in the Netherlands) or $10^6$ (protection of nuclear power plants along the coast or tidal rivers) years [e.g., Pugh and Woodworth, 2014]—even if the available data sets do not meet the criteria; in these RWLs decadal variations are even larger. Numerical model runs can be used to complement TG observations to improve the data basis for extreme value analyses temporally (allowing for more robust estimates) and/or to obtain values for ungauged coastline stretches [Arns et al., 2013, 2015; Haigh et al., 2014b, 2014c; Zhang and Sheng, 2013].

5. Conclusions

Multidecadal variations in MSL time series have recently been the focus of several studies which have highlighted that such fluctuations may (1) hamper the early detection of accelerated sea level rise; (2) result in the exceedance of critical thresholds much earlier than suggested by smooth sea level projections as published for example by the IPCC, and (3) lead to higher sea levels toward the end of the century than currently thought plausible. These low-frequency changes will also affect extreme sea level events, which pose a much greater risk to highly developed and densely populated low-lying coastal areas than the direct effects of MSL rise. In this context it has been (and still is) widely assumed that ESL changes broadly follow MSL changes without showing distinct significant long-term trends or (multidecadal) fluctuations. At its simplest, this implies that taking, let us say, a MSL scenario of 50 cm (for 2100) into account to plan a coastal structure (e.g., a dike or a flood gate) with an anticipated lifetime of ~90 years just requires adding 50 cm to the relevant design RWL (e.g., 100 or 1000 years). This approach does not account for decadal fluctuations in mean and extreme sea levels which may affect RWLs in a way that the (guaranteed) safety standards are not met throughout the entire lifetime of the structure. Other approaches account for uncertainties in future projections in sea level rise and/or changes in storminess by determining the required adaptation in order to preserve the present-day likelihood of flooding [e.g., Hunter, 2012; Hunter et al., 2013; Pugh and Woodworth, 2014]. However, in order to adequately take low-frequency MSL/ESL variations superimposed onto long-term trends/accelerations into account for coastal adaptation, projections of these multidecadal variations need to become available. With the present study we make an important step toward this long-term objective.

We analyzed hourly water level records from 20 TGs around the contiguous US coastline systematically to identify changes in ESL different to those in MSL. We employed different data sampling and analysis techniques that have been proposed in the literature to assure robustness of the results against the chosen methodology. All analyses were performed separately for summer and winter seasons to account for the differences between tropical (dominating in summer) and extratropical (dominating in winter) storms in their formation and distinct links to large-scale climate variability.

We find significant long-term trends for the 1929–2013 period at individual sites along most coastline stretches with one of the methods and for one of the two seasons (Figure 3). However, the results obtained with various extreme indices differed, suggesting that one has to be careful in interpreting trends found at a particular site with a particular analysis technique as a climate related trend. The winter season in the SA region is the only case where we find a larger number of neighboring TGs exhibiting positive and (partly) significant linear trends in the order of 0.3–0.5 mm yr$^{-1}$. If these trends continue it would lead to changes <$5$ cm in ESLs above MSL rise until the end of the century. This is only a fraction of the 28–98 cm of GMSL rise currently assumed plausible [Church et al., 2013] for the end of the century (regional relative MSL changes may be much larger/smaller). Overall, these findings are broadly consistent with those derived by Menéndez and Woodworth [2010] from analyzing long US TG records and support the assumption that long-term trends in ESLs over the last 100–150 years were primarily due to MSL rise along most of the US coastline. This is also in general agreement with studies that focused on long-term changes in storminess in the Atlantic and Pacific using direct observations, proxy data, or model output. Those revealed no or only little evidence for significant storminess trends at centennial time scales, but decadal variations which have caused trends over particular time periods [e.g., Aucan et al., 2012; Bromirski et al., 2003; Feser et al., 2014; Seneviratne et al., 2012].
Further analyses showed that if significant ESL trends exist at a TG site, they generally persist over a wider range of high quantile time series (Figure 4), although the magnitudes may vary. Analyzing trends for different window lengths and start years showed that approximately 60–80 years (depending on the TG location and time series variability) of data are required to identify significant long-term ESL trends, which are robust against the selected time period. Results from the same analysis also confirmed the existence of significant multidecadal variations in ESL time series (Figure 5). These variations, especially in design relevant RWLs, are much larger and hence more important than long-term trends, also suggesting that linear trend models alone may be inadequate to describe all relevant ESL changes or include them into nonstationary extreme value analysis. The multidecadal ESL variability is also much more coherent, both spatially and when comparing results from employing different methods (Figures 6–9).

We find six regions along the US coast, each one comprising at least three long TG records, exhibiting consistent variability and we identified for each region the relevant seasons which dominate the storm surge climate. In total, four analysis techniques were applied to assure robustness of the results. The amplitudes of the multidecadal variations range from ~10 to 110 cm for the 50–200 year RWLs at selected sites (with each one being representative of an entire region) and the relevant seasons (Figure 11).

These findings clearly challenge the applicability of the “MSL-offset” method. Taking Atlantic City (winter) and the design of a coastal protection structure as an example, between the 1940s and today the 100 year RWL has increased by almost 40 cm (unrelated to MSL changes); hence, just adding a certain amount of MSL rise linearly to the 100 year RWL in the 1940s would have resulted in a significant decrease in the level of protection provided by the measure. On the other hand, in Baltimore (summer; and other sites nearby), the 100 year RWL has decreased by almost 80 cm; RWLs from TGs in other areas, such as the GOM or SP regions do not show any trends but fluctuations in the order of 10–100 cm.

Overall, the results presented here highlight the utmost need to further explore multidecadal ESL variations and their physical driving factors in order to ultimately be able to include the information into (quasi-)non-stationary extreme value models when planning coastal adaptation measures with the purpose of providing predefined levels of protection throughout their anticipated lifetimes. In a future study, we plan to investigate the relationships between the ESL variations along the US coastline, as identified here, and (large scale) climate variability. If such links can be established, the results may be used along with climate model output and statistical downscaling to infer information about future ESL changes and the potential effects on RWLs. In the present study we identified the relevant temporal changes (which are robust against the employed data sampling and analysis techniques) and also considerably reduced the amount of data that need to be further investigated. Being able to explain the observed changes in the eight time series shown in Figure 8 (or associated location and scale parameters shown in Figure 10) means being able to explain a major fraction of the relevant multidecadal ESL variability along the entire US coastline.

References


