Trends in extreme mean sea level quantiles from satellite altimetry

Susana M. Barbosa
INESC TEC

Campus da FEUP, 4200-465 Porto, Portugal
Tel: +351 22 209 4199 / Fax: +351 22 209 4050

susana.a.barbosa@inesctec.pt

Abstract

Satellite altimetry allows the study of sea-level long-term variability on a global and spatially uniform basis. Here quantile regression is applied to derive robust median regression trends of mean sea level as well as trends in extreme quantiles from radar altimetry time series. In contrast with ordinary least squares regression, which only provides an estimate on the rate of change of the mean of the data distribution, quantile regression allows the estimation of trends at different quantiles of the data distribution, yielding a more complete picture of long-term variability. Trends derived from basin-wide averaged regional mean sea level time series are robust and similar for all quantiles, indicating that all parts of the data distribution are changing at the same rate. In contrast, trends are not robust and diverge across quantiles in the case of local time series. Trends are under- (over-) estimated in the western (eastern) equatorial Pacific. Furthermore, trends in the lowermost quantile (0.05) are larger than the median trend in the western Pacific, while trends in the uppermost quantile (0.95) are lower than the median trend in the eastern Pacific. These differences in trends in extreme mean sea level quantiles are explained by the exceptional effect of the strong 1997-1998 ENSO event.

Keywords: mean sea level; satellite altimetry;
1 Introduction

The operation of consecutive satellite radar altimeters since the early 1990's provides a continuous data set of sea surface heights with a nearly global and spatially uniform coverage, allowing to study in an unprecedented way mean sea-level variability (Ablain et al. 2015).

The analysis of linear trends in sea level from satellite altimetry records has mainly focused on the estimation by linear regression of trends in the mean (e.g. Nerem et al. 2010; Cazenave et al. 2014; Jorda, 2014). However, the long-term variability of sea level is not necessarily spatially and temporally uniform (e.g. Llovel et al. 2011a, 2011b; Haigh et al. 2014; Dangendorf et al, 2014), and additional information can be gained by considering linear trends in quantiles of the data distribution in addition to trends in the mean. The study of extremes from satellite altimetry records is severely limited by the short period spanned by the measurements, particularly in comparison with tide gauges on islands and coastal locations, which are traditionally used for the analysis of extreme sea levels. A recent study (Woodworth and Menéndez, 2015) addressed the changes in extreme sea levels from satellite altimetry based on monthly averaged observations for a 20-years period (1993-2013). In the present study trends in extreme quantiles of mean sea level, rather than trends in extremes, are analyzed from 10-days time series of satellite altimetry observations for the 1993-2014 period using quantile regression. Here the extreme quantiles are defined as the 5% and 95% quantiles of the data distribution.

The quantile regression approach, initially developed by Koenker and Basset (1978), and applied since then in a wide range of geophysical applications (e.g. Barbosa, 2008; Jagger and Elsner, 2009; Monteiro et al. 2012; Wasco, 2014; Mendoza et al. 2015), allows to derive trends for different quantiles of the data distribution, providing a more complete description of long-term variability. While ordinary least squares regression minimizes the sum of squared residuals, median regression is the particular case of quantile regression for the 2\textsuperscript{nd} quantile (quantile 0.5) and minimizes the sum of absolute residuals, while general quantile regression minimizes the sum of asymmetrically-weighted absolute residuals. In addition to the extension to extreme quantiles, the quantile regression framework yields a more robust estimation of trends even in the central part of the data distribution, in the same sense as the median is a more robust estimate of central location than the mean.

In this work quantile regression is applied to derive robust median regression trends as well as trends in extreme mean sea level quantiles from sea level records of satellite altimetry measurements. The data considered and the applied methodology
are described in section 2. Results in terms of median regression trends and trends in extreme quantiles are first presented for regionally-averaged sea level in section 3.1, and then for local sea level in section 3.2. Concluding remarks are given in section 4.

2. Data and methods

Satellite altimetry data for the period 1993-2014 are obtained from the Radar Altimeter Database System (RADS) for Topex (cycle 11 to cycle 364), Jason-1 (cycle 22 to cycle 258), and Jason-2 (cycle 20 to cycle 237) missions. Furthermore, regional mean sea level time series resulting from basin-wide averaging of RADS data are obtained from the NOAA Laboratory for Satellite Altimetry for the North Atlantic, Atlantic, Pacific and Indian Ocean basins (Leuliette and Scharroo, 2010). Only altimetry measurements between 66 °S and 66 °N have been included. All standard instrumental and geophysical corrections (Chelton et al. 2000) are applied to the satellite altimetry measurements, including the inverse barometer correction. Sea level anomalies are computed using the CLS01 mean sea surface (Hernandez and Schaeffer, 2000). Glacial isostatic adjustment (GIA) effects on the geoid are not taken into account.

The satellite altimetry data are gridded to a 1 degree by 1 degree grid using the rads2grd routine (Scharroo, 2012), which simply spatially averages (without any smoothing or interpolation) the along-track sea level anomalies collected into 1 degree by 1 degree cells. Time series at each grid cell are derived taking the average time over the 10-days period of each cycle as the reference time. No temporal averaging is performed on the altimetry observations, and only time series with at most 50 missing values (out of a total of 809 cycles) are considered in the analysis.

As a data pre-processing step, seasonal signals are removed by sinusoidal regression. Satellite altimetry observations include a ~60-days non-geophysical signal resulting from the aliasing of tides, orbit and radiometer at the ~10-day repeat cycle (Bessières et al. 2013). Here this periodic signal is not explicitly filtered, since the focus of the present study is on non-periodic variability. However, a 30-days running median filter is applied for robust smoothing and elimination of eventual isolated outliers (Härdle and Steiger, 1995). The use of a running median instead of a running mean ensures a more robust filtering, less affected by off-lying values, and the use of a short window of only 3 observations ensures minimal smoothing and consequent preservation of the temporal features of the time series and of its correlation structure.

Quantile regression has become a mature and useful complementary tool to
conventional ordinary least squares linear regression analysis (Koenker, 2005). On the one hand, median regression as opposed to modelling of the mean in the traditional framework is considerably more robust to the presence of outliers in the data. On the other hand, quantile regression provides additional information by allowing to derive trends in the tails of a data distribution, in addition to the center. Because it makes no distributional assumption about the error term in the linear model, quantile regression offers considerable model robustness and flexibility. A common misconception is that a simple segmentation of the unconditional distribution of the response variable followed by least squares fits for the subsets would be akin to quantile regression. However, such a segmentation approach might lead to incorrect results when, for example, the data include outliers (Koenker and Hallock, 2001). Quantile regression in contrast uses all of the data, even for the extreme quantiles. In this work a constrained version of quantile regression using spline-based constraints is applied (Bondell et al. 2010) in order to guarantee that linear quantile curves for different quantiles do not cross, leading to invalid distributions (e.g. He, 1997). Technical details on the quantile regression method can be found in Koenker (2000), Koenker and Hallock (2001), Koenker (2005), and references therein, here a brief overview is presented.

The quantile function of a random variable Y is the inverse of its distribution function \( F_Y \) providing, like the distribution function, a complete description of the statistical properties of the variable. Likewise for a quantile \( \tau \) the conditional quantile function of Y given an explanatory variable X, \( Q_Y(\tau | X) \), fully captures the relationship between Y and X, and is the inverse of the corresponding conditional distribution function, \( Q_Y(\tau | X) = F_Y^{-1}(\tau | X) \). Classical ordinary least squares regression is based on the conditional mean function, the mean of the response variable Y conditional on the explanatory variable X, denoted by \( E[Y | X] \), and the minimization of the residuals \( \sum (y_i - E[Y | X=x_i])^2 \). Here the conventional estimation approach based on the QR decomposition method is applied for ordinary least squares regression. Similarly quantile regression is based on the minimization of the sum of asymmetrically weighted absolute residuals \( \sum \rho(\tau)(y_i - Q_Y(\tau | X)) \), where \( \rho(\tau) \) is the tilted absolute value function (Koenker and Hallock, 2001). Here the optimization is performed using a modified version of the simplex algorithm of Barrodale and Roberts (1974), which is described in Koenker and D’Orey (1994).

Quantile regression, as ordinary linear regression, typically assumes independent and identically distributed errors. The effect of serial dependence on quantile regression is the same as for ordinary regression, with inference based on the conventional variance-covariance matrix becoming invalid. Given that geoscience time series in
general and sea level time series in particular are typically autocorrelated, including both short-range and long-range temporal dependence (e.g. Barbosa et al. 2008), the assessment of uncertainty needs to take serial correlation into consideration.

Despite the many recent developments in quantile regression for time series data (e.g. Xiao, 2012) the parametrization of the covariance to fully capture the temporal dependence of real world time series remains a challenging task. Thus a non-parametric approach based on statistical bootstrapping is taken here for the assessment of the uncertainty of quantile trend estimates, as illustrated in Barbosa et al. (2011). Maximum entropy bootstrapping (Vinod and López-de-Lacalle, 2009) is a particularly useful resampling technique for non-stationary time series since it allows to retain the shape and temporal variability of the original time series in the bootstrap replicates. Here all uncertainty estimates are derived from the median absolute deviation (MAD) of slope values computed from a set of 500 time series replicates generated by maximum entropy bootstrap. Error bars are represented as the slope values from the original time series plus/minus MAD values from time series replicates.

Sea level trends from satellite altimetry can be significantly influenced by the specific processing of radar measurements (e.g. Fernandes et al. 2006; Ablain et al. 2015), particularly corrections such as the wet troposphere correction and orbit parameters. The uncertainty in sea level trends at the regional scale can be of the order of 2-3 mm/year (Ablain et al. 2015), the largest source of uncertainty being the orbit error (Couhert et al. 2015). In the present study only the statistical uncertainty on the trend estimate is considered.

Statistical tests are applied for inference on the trends estimated by quantile regression. The Wald test is applied to compare slopes derived by the ordinary least squares linear regression and by median regression (Fox and Weisberg, 2011). An F-test as described in Basset and Koenker (1982), is applied to infer whether trends in extreme lower (0.05) and upper (0.95) quantiles are significantly different from the median trend (quantile 0.5). Results of statistical tests are presented in terms of the corresponding p-value, which is the probability of getting a similar or more extreme result given that the hypothesis under consideration is true. The smaller the p-value the stronger the evidence in favor of the alternative hypothesis.

The data analysis is performed with the R software (R Core Team, 2015) and maps are produced with the GMT software (Wessel et al. 2013).

3. Results
First basin-wide trends in sea level quantiles are considered (section 3.1), then local trends at the global scale and corresponding maps are presented (section 3.2).

### 3.1. Regional sea level

The quantile regression analysis is illustrated in detail for the regional mean sea level time series for the Atlantic, Pacific and Indian Ocean basins. Fig. 1 shows the regional mean sea level time series and the corresponding median (quantile 0.5) and extreme (quantiles 0.05 and 0.95) quantile trends. Except for the North Atlantic, very similar slopes are obtained for ordinary and median regression, as well as for the extreme quantiles. In contrast, the North Atlantic exhibits significant differences between the central and the most extreme quantiles. The Wald test confirms that the median slope is significantly higher than the ordinary least squares slope (p-value 5E-06), and the F-test indicates that for the North Atlantic both lower and upper extreme trends are significantly lower than the median slope for a 95% confidence level (p-values 8E-07 and 4E-09, respectively).

For a more complete description of trends in regional mean sea level, quantile trends are computed for all quantiles from quantile 0.05 to quantile 0.95 in steps of 0.05 (quantiles 0.05, 0.1, 0.15, 0.2,...) and displayed in Fig. 2. The slopes for each quantile are represented by circles (°) and the corresponding uncertainty by the superimposed vertical line. For comparison the error bars for the slope value obtained by ordinary least squares regression are also displayed as horizontal dashed lines (same value across all quantiles). The results confirm that basin-wide trends are very stable and robust, since a very similar slope value is obtained across all quantiles. The spatial averaging reduces local effects producing a more homogeneous long-term variability signal. Again the exception is the North Atlantic, with slopes for the lower and upper quantiles distinct from the slopes for the central quantiles, reflecting the large fluctuations exhibited by the time series (Fig. 1(b)).

### 3.2. Local sea level

Median regression trends from local (1º grid) satellite altimetry time series are displayed in Fig. 3(a). The spatial pattern is very similar to the one obtained using ordinary least squares regression (not shown), and consistent with published maps of satellite altimetry trends (e.g. Cazenave and Llovel, 2010; Zhang and Church, 2012; Fukumori and Wang, 2013). However, statistically significant differences are obtained when comparing the slopes from median and ordinary least square regression. Figure 3(b) shows the difference of slopes (median - ordinary). Differences which are not statistically significant for a confidence level of 95% as indicated by the Wald test are set as zero. The map shows that the differences are small, typically below 2
mm/year, but spatially consistent, the largest differences occurring in the equatorial Pacific and in high-variability areas associated with current systems. The median slopes are higher in the eastern Pacific and lower than the ordinary least squares slopes in the western equatorial Pacific.

Trends for the extreme quantiles are compared with the median trends by computing the difference between the corresponding slopes as displayed in Fig. 4. Again, differences which are not statistically significant for a confidence level of 95% as indicated by the F-test are set as zero.

In Fig. 4(a) the differences tend to be positive, corresponding to quantile 0.05 slopes higher than the median slope. The largest differences are associated with western boundary current systems (Kuroshio, Gulf Stream). In the western equatorial Pacific the lower quantile trends are systematically higher than the trends in the central part of the distribution. In contrast Fig. 4(b) shows that differences tend to be negative, corresponding to quantile 0.95 slopes lower than the median slope. As for the lower quantile trends the largest differences are associated with high-variability regions and major currents. In the eastern equatorial Pacific the upper quantile slopes are significantly lower than the median slopes.

4. Discussion and conclusions

Both median and ordinary regression trends describe the long-term variability in the central part of the data distribution, and should be the same in the case of a perfectly symmetric distribution. However, median regression is more robust to non-symmetric distributions and to the presence of eventual outliers. Thus the comparison of median and ordinary regression trends provides a measure of symmetry of the data distribution and of the robustness of the derived slopes.

In the case of basin-wide averaged altimetry measurements, the median and ordinary slopes are very similar (Table 1) confirming that the trends derived for the Atlantic, Pacific and Indian Ocean basins are very robust. The difference is larger in the case of the North Atlantic, the median regression slope being significantly higher than the ordinary least squares regression slope. Furthermore for the North Atlantic the slopes for the smallest and largest quantiles are significantly lower than the median slope. The regional mean sea level time series for the North Atlantic (Fig. 1(b)) exhibits strong oscillations, in particular a marked drop in 2010-2012 that can explain the downward bias of quantile slopes for the more extreme quantiles. These features seem to be associated with large scale climate variability and the state of the North Atlantic Oscillation (NAO), which is known to affect sea-level variability in the North Atlantic (e.g. Jevrejeva et al. 2005; Woodworth et al. 2007; Tsimpis & Shaw,
In particular the lower sea level values in 2010-2012 (Fig. 1(b)) seem to be linked with the strong negative phase of the NAO index for that period.

In the case of local altimetry time series the differences between median and ordinary regression trends are larger than for regional mean sea level. In the eastern equatorial Pacific the ordinary regression slopes are underestimated when compared with the more robust (and therefore taken as reference) median slopes, while in the western Pacific the ordinary regression slopes are overestimated. These results are explained by the strong impact of the 1997-1998 ENSO event in the equatorial Pacific sea level (e.g. Nerem et al. 1999; Cazenave et al. 2012). The impact of the event in the ordinary regression slopes is enhanced by its location in the initial part of the sea level time series, as illustrated for an individual time series in Fig. 6(a). In the eastern Pacific the strong peak in the initial part of the record pulls up the first part of the time series when compared to the second half of the record leading to a reduction in the ordinary regression slope estimate. Conversely, in the western equatorial Pacific the strong trough in the beginning of the records pulls down the observations in the beginning of the record, inducing an apparent increasing trend, as illustrated in Fig. 6(b). The quantile regression slopes are more robust to the influence of the 1997-1998 ENSO event in the sea level time series (in the same sense as the median is less affected by non-symmetry than the mean) and therefore the median trends are not as under (over) estimated in the eastern (western) Pacific.

In terms of trends in extreme quantiles, the largest differences to central trends are typically larger than 3 mm/year and are also concentrated in the equatorial Pacific, reflecting the influence of the 1997-1998 ENSO event. In the eastern Pacific, the sea level time series are characterized by a strong peak in the beginning of the record. The peak influences the trend estimates for the upper quantiles, pulling-up the initial part of the time series compared to the last part of the record, and leading to an upper quantile slope estimate lower than the median slope. The lower quantile trends are not affected by the peak, thus in the eastern Pacific the lower quantile trends are similar to the corresponding median trends. The opposite mechanism is in place for the western equatorial Pacific. The sea level time series are characterized by a strong trough in the initial part, which leaves unaffected upper quantile slopes but pulls-down the initial part of the record leading to a lower quantile trend higher than the corresponding median trend.

Except for the strong influence of the 1997-1998 ENSO event, the quantile regression analysis of the satellite altimetry time series shows that the trend patterns are robust. Furthermore, similar slopes are obtained across all quantiles, indicating a
similar change in all parts of the data distribution.

The 2-3 mm/yr current uncertainty on regional sea level trends from satellite altimetry (Ablain et al. 2015) is mainly associated with the satellite measurement of sea level, including the orbit determination and geophysical corrections needed to convert the radar measurement to an height estimate. While the differences in central trends uncovered in the present study are within this 2-3 mm/yr uncertainty, they are associated only with the trend estimation itself, and result simply from applying two different approaches (ordinary least squares or median regression) to exactly the same altimetry observations, derived using the same orbit and altimeter parameters. Furthermore, the differences between ordinary and median regression estimates are well outside the statistical uncertainty for the corresponding slopes. Here uncertainty is assessed by statistical bootstrapping and corresponding time series oversampling, which leads to small uncertainties.

These differences in ordinary and median regression trends do not imply that one method is better than the other, rather that the methods reflect different aspects of the time series, particularly concerning inter-annual variability and extreme events. The comparison of slopes from ordinary least squares and median regression allows a quantitative assessment of the influence of inter-annual variability and extremes on the trend estimates. The results show a strong influence of the 1997-1998 ENSO event on trends in the equatorial Pacific. The ordinary regression trends reflect long-term variability incorporating and even amplifying, due to its location in the beginning of the time series, the effect of the ENSO event in the long-term estimates, while median regression trends reflect long-term variability downplaying that extreme event influence.

Because of the exceptional effect on sea level of the 1997-1998 ENSO event, its influence on ordinary trend estimates for the equatorial Pacific persists even as the length of the satellite altimetry record increases, since it has the same effect in the sea level time series as an outlier observation would have. Therefore the computation of trends by median regression is preferable in the equatorial Pacific, as well as in areas associated with current systems, since the median trends are less affected by the very high variability typical in sea level time series from these regions, allowing to reduce the influence of inter-annual variability and extreme events on the trend estimates.

While the present study focuses only on trends from satellite altimetry records, the influence of interannual variability on trend estimates is also a relevant issue in the case of tide gauges, since regional sea levels from tide gauges display substantial
variability and divergent patterns, despite the much longer time series (Jevrejeva et al. 2014).

The time required for the accurate determination of regional sea level trends from satellite altimetry was estimated to vary geographically between 5 and 100 years (Hughes and Williams, 2010) due to oceanic variability. These estimates do not take into account any systematic altimetric measurement system errors (Fu and Haines, 2013) and are influenced by sporadic extreme phenomena such as the 1997-1998 ENSO event, reinforcing the importance of the length of sea level time series for the accurate estimation of linear trends, particularly in the equatorial Pacific.

**ACKNOWLEDGMENTS**

This work is financed by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme, and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project POCI-01-0145-FEDER-006961 and programme IF. Radar altimetry data are provided by the NOAA Laboratory for Satellite Altimetry and the Radar Altimetry Database System (RADS).

**References**


Hernandez, F., and P. Schaeffer. 2001. *The CLS01 Mean Sea Surface: A validation with the GSFC00 surface*. CLS Ramonville St Agne, France.


<table>
<thead>
<tr>
<th>Region</th>
<th>Ordinary least squares (mm/year)</th>
<th>Median regression (mm/year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlantic Ocean</td>
<td>2.69 ± 0.045</td>
<td>2.64 ± 0.046</td>
</tr>
<tr>
<td>North Atlantic</td>
<td>1.79 ± 0.033</td>
<td>2.08 ± 0.048</td>
</tr>
<tr>
<td>Pacific Ocean</td>
<td>2.57 ± 0.046</td>
<td>2.62 ± 0.057</td>
</tr>
<tr>
<td>Indian Ocean</td>
<td>3.68 ± 0.056</td>
<td>3.70 ± 0.062</td>
</tr>
</tbody>
</table>

Table 1: Trends (mm/year) from regional mean sea level time series. Uncertainties are derived by bootstrapping for both ordinary least squares regression and median regression.
List of Figures

**Figure 1:** Regional mean sea level time series plots and trend lines for quantiles 0.05 (....), 0.5 (--), and 0.95 (---): (a) Atlantic Ocean, (b) North Atlantic, (c) Pacific Ocean and (d) Indian Ocean.

**Figure 2:** Quantile regression slopes for regional mean sea level: (a) Atlantic Ocean, (b) North Atlantic, (c) Pacific Ocean and (d) Indian Ocean. The slopes are represented by open circles (°) and the corresponding uncertainty by the super-imposed vertical lines. The horizontal lines depict the error bars for the ordinary least squares slope (same for all quantiles).

**Figure 3:** Median regression trends: (a) slopes and (b) difference relative to ordinary least squares regression (median - ordinary).

**Figure 4:** Difference between trends for extreme quantiles and the median trend (quantile – median) for (a) quantile 0.05 and (b) quantile 0.95.

**Figure 5:** Local time series of sea level anomalies at (a) 97 °W 0 °N (eastern Pacific) and (b) 164 °E 2 °S (western Pacific).
Figure 1: Regional mean sea level time series plots and trend lines for quantiles 0.05 (---), 0.5 (—), and 0.95 (——): (a) Atlantic Ocean, (b) North Atlantic, (c) Pacific Ocean and (d) Indian Ocean.
Figure 2: Quantile regression slopes for regional mean sea level: (a) Atlantic Ocean, (b) North Atlantic, (c) Pacific Ocean and (d) Indian Ocean. The slopes are represented by open circles (°) and the corresponding uncertainty by the super-imposed vertical lines. The horizontal lines depict the error bars for the ordinary least squares slope (same for all quantiles).
Figure 3: Median regression trends: (a) slopes and (b) difference relative to ordinary least squares regression (median - ordinary).
Figure 4: Difference between trends for extreme quantiles and the median trend (quantile – median) for (a) quantile 0.05 and (b) quantile 0.95.
Figure 5: Local time series of sea level anomalies at (a) 97 °W 0 °N (eastern Pacific) and (b) 164 °E 2 °S (western Pacific).