

Manuscript prepared for Nat. Hazards Earth Syst. Sci. Discuss.
with version 2015/09/17 7.94 Copernicus papers of the \LaTeX class copernicus.cls.
Date: 21 November 2015

Statistical properties of coastal long waves analysed through sea-level time-gradient functions: exemplary analysis of the Siracusa, Italy, tide-gauge data

L. Bressan and S. Tinti

Department of Physics and Astronomy (DIFA), University of Bologna, Bologna, Italy

Correspondence to: S. Tinti (stefano.tinti@unibo.it)

Abstract

This study presents a new method to analyse the properties of the sea-level signal recorded by coastal tide gauges in the long wave range that is in a window between wind/storm waves and tides and is typical of several phenomena like local seiches, coastal shelf resonances and tsunamis. The method consists of computing four specific functions based on the time gradient (slope) of the recorded sea level oscillations, namely the instantaneous slope IS, and three more functions based on IS, that are the reconstructed sea level RSL, the background slope BS and the control function CF. These functions are examined through a traditional spectral FFT analysis and also through a statistical analysis showing that they can be characterised by probability distribution functions PDFs such as the Student's t distribution (IS and RSL) and the Beta distribution (CF). As an example, the method has been applied to data from the tide-gauge station of Siracusa, Italy.

1 Introduction

This paper is focused on the analysis of coastal long waves, for which we mean the period window above the wind and storm waves and below the tides, and that is typical of phenomena like harbour resonant oscillations, coastal-basin seiches, meteo-tsunamis, tsunamis, etc. The study of such long waves is an important component of sea-level studies for coastal areas, also because it is known that they can be very damaging. Large amplitude waves, like tsunamis or extreme seiches, may cause severe flooding of coastal areas, with concomitant danger and damage to people, facilities and properties. And even waves of smaller amplitudes, especially in harbours, are capable of inducing large oscillations and strong currents that might limit operability and stress/break moorings destabilizing ships (Wiegel, 1964; Miles, 1974; Rabinovich, 2009; Kwak et al., 2012).

One of the main topics of the literature on coastal long waves regards resonant properties of harbour basins, which is understandable due to the great economic and strategic relevance of port installations and facilities in all marine countries of the world. Nowadays

free-oscillation and longwave analysis are performed mostly through numerical simulations and hydraulic laboratory modelling (Beltrami et al., 2003; Bellotti, 2007; Luick and Hinwood, 2008; Bellotti et al., 2012; Hinwood and Luick, 2012; Lopez et al., 2012; Guerrini et al., 2014). The main goal in harbor and port engineering studies is to design breakwaters, wharves and quays that are able to withstand the meteo-marine induced wave and current conditions and to design berths and quays such that the moored vessel movements of various types would enable sufficient yearly operability of the vessels at berth during loading or unloading operations.

Despite the recognized importance of longwave studies and coastal resonances, direct measurements of long waves are not common, and they have been carried out especially for research purposes (Okihiro et al., 1993; Okihiro and Guza, 1996; Lara et al., 2004; Bellotti and Franco, 2011; Guerrini et al., 2014). In recent years, the upgrade of the traditional coastal tide gauge networks, that usually recorded data every 10 or 15 min, to faster sampling and recording rates (1–10 s) has allowed the use of sea-level data to study waves in the longwave range, which are however usually analysed only for extreme events as tsunamis (Rabinovich and Thomson, 2007; Bressan and Tinti, 2011; Rabinovich et al., 2013).

In this paper, a study of sea level in the longwave range has been carried out for the harbour of Targia, located a few km north of the town of Siracusa, in Italy, as an example of application of a new method of analysis of coastal tide-gauge records. Since the toponym of Siracusa is known much better than Targia, we will hereafter refer to this station as the Siracusa tide-gauge station. From sea-level traces the method computes four functions that are analysed in two ways: (i) considering them as functions of time, one computes and examines their FFT spectra and (ii) interpreting them as time series of random wave fields, one computes their statistical properties by searching for the probability density functions (PDFs) that best fit the experimental occurrences. The functions we introduce here were first defined in a different context, i.e. within a Tsunami Early Detection Algorithm (TEDA) with the purpose of detecting tsunamis and hazardous long waves (see Bressan and Tinti, 2011, 2012 and Bressan et al., 2013) and were proven to be

suitable to characterize the sea-level signal in the tsunami frequency band. Before using TEDA for a specific site, a calibration process is needed to determine the parameters of TEDA computational procedures and mathematical expressions. It was during the TEDA calibration for the Siracusa tide gauge that the new analysis method presented in this paper was first conceived. This is the reason why (i) the method is applied here to Siracusa station data, (ii) the longwave functions are calculated with the TEDA setting resulting from calibration and (iii) the calibration procedure is synthetically illustrated in the Appendix, with the main body of the paper dedicated to the new method.

In the next sections of the paper, we first introduce the four functions of the method and the basic data set we use. Then we illustrate the results of the FFT analysis of these functions, which is the first part of the method, and the results of the statistical analysis, which is the second part. Discussion of the outcomes and of their potential will conclude the paper.

2 Longwave functions

The method for longwave analysis we propose is based on the definition of four functions, the first of which is the instantaneous sea-level slope IS. From it, further longwave functions are derived: the reconstructed sea-level RSL, the background slope BS and the control function CF. These functions will be defined here below, and an example of them is shown in Fig. 1.

2.1 Instantaneous slope IS

The function IS is computed by least-squares fitting the detided time series of the sea-level data within the time interval D_{IS} of length T_{IS} . Technically speaking, let's suppose that t_i is the current sampling time, that $dt = t_i - t_{i-1}$ is the sampling interval and that D_{IS} is the backward time interval specified by the end times t_{i-L} and t_i , with $T_{IS} = t_i - t_{i-L} = Ldt$. Let's further suppose that h_k is the sea level reading in the interval D_{IS} , i.e. $i - L < k \leq i$. Then $IS(t_i)$ is the slope coefficient of the straight line fitting the set of h_k in a least-squares

sense, which can be denoted as:

$$\text{IS}(t_i) = \text{slope}(h_k)$$

We observe that this procedure cuts periods smaller than T_{IS} and that detiding cuts periods of tidal interest, which means that the function IS contains oscillations in the intermediate interval band between T_{IS} and tides and can be denoted as longwave instantaneous sea-level slope. The detiding method we adopted will be explained in detail later on. In this application, $T_{\text{IS}} = 4$ min.

2.2 Reconstructed sea level RSL

The longwave reconstructed sea level RSL represents a filtered sea-level signal, and it is computed by integrating the function IS over the time interval D_{RSL} , of length $T_{\text{RSL}} = Ndt$, which for discretized values transforms into the following N -term summation:

$$\text{RSL}(t_i) = dt \sum_{k=i-N+1}^i \text{IS}(t_k)$$

Since the sea-level slope IS is detided, the function RSL does not require additional tidal corrections. The “filtering” effect comes from the joint contribution of the selected durations, T_{IS} and T_{RSL} , of the respective computation intervals D_{IS} and D_{RSL} (see Bressan and Tinti, 2011, for details). In this application, $T_{\text{RSL}} = 10$ min.

2.3 Background slope BS

The background slope BS is defined as the maximum sea-level slope (in absolute value) computed over the time interval D_{BS} of length $T_{\text{BS}} = N_{\text{BS}}dt$, where D_{BS} is the backward interval specified by the end times $t_{i-N_{\text{BS}}-N_{\text{G}}}$ and $t_{i-N_{\text{G}}}$ and is separated from the current sampling time t_i by the gap interval of length $T_{\text{G}} = N_{\text{G}}dt$. Formally, this definition can be given the form:

$$\text{BS}(t_i) = \max(|\text{IS}(t_k)|), \quad \text{with } k = i - N_{\text{BS}} - N_{\text{G}+1} + 1, \dots, i - N_{\text{G}}.$$

In this application, $T_{BS} = 60$ min and $T_G = 15$ min.

2.4 Control function CF

The control function CF is defined as the positive ratio of the sea level and background slope, i.e. as:

$$CF(t_i) = |IS(t_i)|/BS(t_i)$$

It was introduced mainly as a tool to monitor the departure of the instantaneous slope from the background values, and hence to identify anomalies for tsunami and hazardous longwave detection (see Bressan and Tinti, 2011). It is important to remark that $IS(t_i)$ and $BS(t_i)$ refer to independent computation intervals if the gap T_G is larger than T_{IS} , as it occurs in the present application.

2.5 Detiding method

The detiding method we adopted works by removing the tide trend directly from the sea-level slope IS , rather than by removing a fitting synthesis of tidal harmonics from the sea-level record. The detiding procedure consists of the following steps. First, one computes a temporary raw sea-level slope $IS'(t_i)$ on the raw sea-level data (including tide), by using the same algorithm illustrated before for the computation of $IS(t_i)$. Second, one calculates the trend of $IS'(t_i)$ in the time interval D_T of length T_T that precedes the current time t_i by the gap T_{GT} , and that is thus defined as $D_T \equiv [t_i - T_T - T_{GT}; t_i - T_{GT}]$. This trend, say $IS_T(t)$, which is assumed to be due to the tide, is computed by least-squares fitting the time function $IS'(t_i)$ within D_T by means of a polynomial function. Third, the tidal slope at the time t_i , i.e. $IS_T(t_i)$, is obtained from $IS_T(t)$ by a proper extrapolation that makes use of the same polynomial function. Finally, one computes the instantaneous slope by subtracting the tidal slope from the raw slope, i.e. $IS(t_i) = IS'(t_i) - IS_T(t_i)$.

In general, for small tidal heights (in the range of 0.2–0.3 m) as in most of the Mediterranean stations, a fitting polynomial of degree 0 (the mean) is sufficient for a correct

detiding, computed over an interval $T_T = 60$ min and anticipated by the gap $T_{GT} = 15$ min, as used here and in Bressan et al. (2013). We observe that for coastal oceanic stations, the optimal parameters for a correct detiding might vary greatly. For example, for the analysis of tide-gauge stations along the Pacific Ocean coasts Bressan and Tinti (2012) found it appropriate to use a polynomial of degree 0 and a computation interval length of $T_T = 60$ min for about 20 % of coastal stations, while for the remaining 80 % a degree 2 polynomial was found more adequate with T_T in the range of 45–330 min and an average value of 210 min.

3 The Siracusa sea-level data

In the frame of the TSUNET project, a local sea-level monitoring network has been installed by the University of Bologna for the coasts of eastern Sicily, Italy, including three coastal tide-gauges in Tremestrieri (south of Messina), Catania and Targia (north of Siracusa), that will be named as the Siracusa tide gauge in this paper. This tide gauge is installed in the inner wall of the main breakwater forming the little harbour of Targia in the bay of Augusta, that is divided in two sectors (north and south) by the natural peninsula of Magnisi (see Fig. 2). The bay has a great strategic value for military, commercial and industrial reasons also related to the industrial area of Priolo (with petrochemical-, refinery-, electric power station-installations) located in the northern sector.

The Siracusa tide-gauge station started recording on 4 May 2011 with a 5 s sampling period, adequate to measure and detect tsunamis. The dataset analysed in this study is composed by sea-level data embracing a period of about 3 years from 5 May 2011 to 31 May 2014.

4 Spectral analysis

In this section we compute traditional FFT spectra of the sea-level signal including tides for a subset of the available data spanning the time interval from May 2011 to June 2013. We

calculate power spectral densities (PSD), in a way similar to Welch (1967), over consecutive 12 h time windows, which means that we obtain a spectrum every 12 h for a total period of about 25 months. For the analysis, FFT periodograms have been averaged over the entire data interval, over years and over calendar months. The results are shown in Figs. 3 and 4. Figure 3 displays averaged spectra of two year-long consecutive time intervals starting from May 2011 and from May 2012, as well as spectra averaged over the whole period (denoted as “Total”). It may be seen that all spectra are very similar to one another, with almost negligible differences. Figure 4 portrays average spectra per calendar months from January (month 1) to December (month 12) as well as the total selected period spanning from June 2012 to May 2013 for comparison. In this case, spectra are similar in shape, that is they exhibit the same peaks, but have different level of intensity. The main conclusion is that the spectral content of the sea-level signal is stable in time and there are no sensible changes from one year to the other (see Fig. 3), while on the contrary there are relevant changes in the PSD intensity associated with the seasonal cycle, with high energy spectra in spring months (March–May) and low energetic spectra in summer months (June–September) (see Fig. 4). The stability of the spectral shape allows one to recognise a typical power spectrum for the Siracusa tide-gauge sea level that is characterised by many distinct spectral peaks, corresponding to periods of 21.2–21.8, 19.5, 11.0–12.5, 9.6, 7.6, 6.0, 5.25, 4.5, 4.1, 2.5–3.0, and 1.0–1.2 min. As for the PSD trend, two typical trends can be identified with transition approximately at 15 min, that corresponds to the end of the longest resonant peak: for lower frequencies, spectra decay as ω^{-2} , while at higher frequencies as ω^{-1} .

The yearly stability and seasonal variability of sea-level spectra in the intermediate wave regime has been found for many coastal stations (see e.g. Rabinovich, 2009; Bressan and Tinti, 2011, 2013), with typical spectra changing from station to station. It has to be noticed that, while offshore it is possible to refer to a universal shape for the shortwave amplitude spectrum, as for example the JONSWAP (Hasselmann et al., 1973) that can be adapted to local conditions (such as fetch), on the contrary spectral characteristics at the coast differ substantially from place to place. Indeed, spectral peaks may change in number, position and width, depending on the morphology and bathymetry of the coastal area. In general, the

association between spectral peaks and coastal basins or sub-basins is not trivial and has not been carried out in this study. We simply suggest that for Siracusa tide-gauge station, the semi-closed Targia harbour basin, the southern part of the bay of Augusta bounded by the Magnisi peninsula, and the whole Augusta bay (see Fig. 2) might act as resonant basins and might be responsible for some (if not all) of the observed peaks.

5 Spectral analysis of the longwave functions

In the first step of our method the longwave functions introduced above are seen as a function of time, since they can be computed at every sampling time t_i and can be studied by means of the same spectral analysis as the original sea level signal. Average power spectra for the same periods selected in the previous section, i.e. May 2011–April 2012 and May 2012–April 2013 are displayed together with the whole period average (Total) for the functions IS and RSL in the respective Figs. 5 and 6. Two observations hold for both Figures. First, the stability of the PSD over years is well visible. Second, spectral peaks in the window 4–120 min are the same as the ones of the ordinary sea level average PSD (compare Figs. 5 and 6 with Fig. 3), even if spectral peaks of 5.25 and 4.5 min are slightly reduced in intensity, while below 4 min they happen to be modified, this latter feature being controlled by the specific choice for $T_{IS} = 4$ min. As for trend, the PSD of IS is flat (ω^0) up to about 10^{-3} Hz (15 min) and growing as ω for larger frequencies up to $4-5 \times 10^{-3}$ Hz (5-4 min). The similarity of peaks suggests that power spectra of either IS or RSL have the same information content as the PSD of the original signal and can be used to characterise the resonant properties of the tide-gauge station place.

As regards the seasonal cycle, this also can be identified through the power spectra of the longwave functions. Variations within the year going from June 2012 to May 2013 for these functions are shown in Figs. 7 in a way that is an alternative to the graph type proposed in Fig. 4. For all functions we compute the maximum envelope spectrum, say PSD_{MAX} that is obtained by taking the largest PSD among all the average monthly spectra. Likewise, we calculate also the minimum envelope spectrum PSD_{MIN} . The year average PSD together

with the upper and lower envelopes are plotted in the four Fig. 7a–d in the order for the four functions IS, RSL, BS and CF. The distance or gap between envelopes provides a measure of the yearly variability, that we know is due to the seasonal cycle. What one can derive from these graphs is that the gap tends to grow with the frequency for functions IS and RSL, but that it is constant for BS and CF. Further, the gap is much smaller for CF for which the ratio between envelopes PSD_{MAX} and PSD_{MIN} is about 3–4, while for all other functions it falls in the wider range 10–10². It is further to notice that the gap for IS and RSL grows in correspondence with peak maxima, which means that resonances enhance seasonal variations.

Figure 7c-d allows one also to see that the power spectra of functions BS and CF cannot be used for peak characterisation since they lose either substantially or totally the information about resonances. Moreover, it is worth observing that the PSD of BS is almost linear with trend of ω^{-2} .

6 Statistical analysis of the longwave functions

In the second step of the analysis method, one regards the longwave functions IS, RSL, BS and CF as the result of a random process and the empirical value at each sampling time t_i are considered random independent variables $X(t_i)$ that are assumed to obey a probability distribution. In practice, we consider the set formed by all occurrences $X(t_i)$ with t_i belonging to the time interval of analysis as an experimental sample with an empirical frequency distribution (EFD) corresponding to a parent probability distribution function (PDF). The purpose of the analysis is the determination of the parent PDF. In principle, the independence of the observations $X(t_i)$ is not true, since the sea-level signal is a multiscale time function with values at any time depending from previous occurrences, as also shown above through spectral analysis. In order to address the problem of variables independency we have also considered subsets where the observables $X(t_i)$ were randomly selected. Since we have obtained the same statistical results as for the entire set, in the following we illustrate our method by applying our statistical analysis to sets formed by all empirical

variables $X(t_i)$ occurring in the time interval under study with occurrences separated by the sampling interval $dt = 5$ s. The statistical analysis has been carried out over different time intervals, i.e. over the whole dataset, that in this case spans the three-year period from May 2011 to April 2014 and over the corresponding calendar months. Therefore, the shortest analysis time interval considered in this paper has the duration of one month. Indeed we have also tried with shorter duration intervals and found that results are stable for intervals of several days, while they are not for intervals less than one hour. As an example, the normalised EFDs of the variable IS(t_i) for an interval 20 day long (including 3.456×10^5 samples) and for a 20 min interval with 240 samples are shown together for comparison in Fig. 8 and it can be seen that they differ remarkably. Typically, monthly EFDs count more than 5×10^5 samples, and the 3-year EFD over 6.2×10^6 occurrences.

The most important result is that the normalised EFDs of the longwave variables are stable, that is they exhibit a characteristic shape over month- and multiyear periods. These typical shapes are illustrated in Fig. 9 and are bell-like, unimodal, symmetric, centred around the origin, for IS and RSL, whereas they are half-bell, unimodal, positive and decreasing for CF, and positive, asymmetric, long right-tailed for BS. These results do not change even if one considers different time constants such as e.g. T_{IS} and T_{RSL} .

The following step of the statistical analysis is to find the parent PDF that might be associated to the obtained EDFs. We have tried with a number of classical PDFs and have measured the misfit or error E by means of the following formula:

$$E = 0.5 \sum dx |\text{EFD} - \text{PDF}|$$

where dx is the bin width and the absolute value of the difference between EFD and the tried PDF is integrated over the range of the possible occurrences. Consider that, since both EFD and PDF are positive and normalised (i.e. their integral is equal to 1), the above definition ensures that E is an index ranging between 0 (perfect fit) and 1 (worst case with curves totally not overlapping) and that the best fitting solution is the one minimizing the misfit E .

The second relevant result is that we have been able to find that IS and RSL can be satisfactorily fitted by means of a Student t distribution and that CF can be well described by a Beta distribution function. On the contrary we were not able to find distributions with satisfactory fit to the EFDs of BS, and this issue will be left to future studies. This is exemplified in Fig. 9 where EFDs are displayed together with the corresponding PDFs. We point out that for the present study, fitting has been carried out by estimating the parameters of the selected PDF through the maximum likelihood method (Jones et al., 2015). In the following we restrict the analysis only to the three variables for which we have identified the parent PDF, i.e. IS, RSL and CF with the main focus on the multiyear (long term) and seasonal characterisation.

6.1 Instantaneous slope IS

The variable IS has been found to obey a Student t distribution. The normal t distribution is a one-parameter continuous PDF where the real parameter is usually denoted by n and called the degree of freedom, with $n > 0$. This distribution has mean value equal to zero for $n > 1$ and variance equal to $n/(n-2)$ for $n > 2$. With n tending to infinity, it converges to the Normal distribution. Examples of such PDF are given in Fig. 10 for different values of the degree of freedom. The non-standardised t distribution is the generalization of the normal t distribution with two more parameters, the mean μ and the scale S . In our case, since the observed EDFs for IS are centred on the origin, we assume that the parent PDF has mean equal to zero. Indeed, we have also tried a three-parameter fitting, but it turned out that the estimated mean is very close to zero, which justifies our simplifying assumption. We have found that the best possible fit is ensured by a two-parameter t distribution of the type:

$$t(X|n, S) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)\sqrt{\pi n}S\left(1 + \frac{X^2}{nS^2}\right)^{\frac{n+1}{2}}}$$

where X is the random variable, n and S are parameters and Γ is the Gamma function. Fitting has been carried out through the maximum likelihood method under the constraint

that the degree of freedom n cannot assume values larger than 20. As shown before, for $n = 20$ a t distribution is already very close to a Gaussian distribution (to which it tends for increasing n). Practically in this range, two EFDs, very similar to each other, might be fitted by t distributions with very different values of degree of freedom for $n > 20$, which leads to instability in the fitting process.

For the variable IS we have estimated the parameters n and S for the entire 3 year period, for the long-term characterisation of the distribution. These are $n = 2.70$ and $S = 0.19 \text{ cm min}^{-1}$, which leads to a variance $\sigma^2 = nS^2/(n-2) = 0.14 \text{ cm}^2 \text{ min}^{-2}$ and a corresponding standard deviation of 0.37 cm min^{-1} . The low value of the degree of freedom reveals that the PDF is rather far from being a Gaussian, with a smaller modal value and larger tails. We have performed the same estimates also for the corresponding calendar months to study the seasonal variations. To ensure reliability, we have restricted calculations only to months with more than 85 % of high-quality data, which implies that only 24 out of 36 months have been taken into account. Estimates of the monthly datasets have been made in three different ways, more specifically: (1) free case, as a two-parameter estimation process, (2) n constrained case, where only the scale parameter S has been estimated while n was attributed the long-term value, (3) S constrained case, where viceversa only the degree of freedom n was estimated, with S assuming the long-term value. An example is given in Fig. 11, where the IS EFD of September 2013 is fitted with three PDFs as explained before, and in addition with the long-term PDF.

The analysis of the t distribution changes over months is synthesised in Fig. 12 that is formed by five panels. Two of them show the time evolution of mean and variance of the monthly EDFs of the variable IS, from which one sees that mean is practically zero and variance exhibits strong variations with alternation of highs and lows. The result of fitting is shown in panels where estimates of n and S are plotted vs. time. The graph for n gives the result of the case where n is kept fixed to the long-term estimate (black), the result of the one-parameter fitting case where n is estimated under the constraint of a fixed S (red) as well as the outcome of the case where two-parameter fitting is performed (violet). Likewise, the graph for S displays the case with S kept constant at the long-term

value (black), the case with only S estimated (blue) and eventually the case with both parameters estimated (violet). The key-tool to help us to formulate a judgement about all these cases of fitting is the graph of the time-dependent misfits. As expected, the misfit between the EFDs and the theoretical PDFs is larger (worst cases) when the monthly PDFs are all computed with the same long-term values ($n = 2.70$, $S = 0.19 \text{ cm min}^{-1}$). Misfits fall in the range 5–20% (black). Equally expectedly, the smallest misfits are attained when EDFs are fitted with the two-parameter t distribution with variations in the range of 1–5% (violet). And when only one-parameter fitting is attempted the outcoming misfits fall in the intermediate range. What is worth observing is that when n is varied and S is kept constant (red) misfits vary between 4 and 16%, while when on the contrary S is allowed to vary and n is kept constant (blue) misfit values remain confined to 1–6%, which is approximately the same range as the two-parameter misfit. This suggests clearly that passing from two- to one-parameter fitting gives no substantial deterioration of the fitting if n is kept fixed at its long-term value. Consequently, one can regard the monthly IS distribution as a t distribution where the degree of freedom is governed by the long-term regime and the scale is dominated by the seasonal variations. This allows one to characterise the site by a normal Student's t distribution of given n and to report the monthly behaviour by plotting S . From the corresponding graph of S in Fig. 12, following the blue line, one can see that the largest values of S ($S > 0.25 \text{ cm min}^{-1}$) are obtained for February 2012, April 2012, December 2012, May 2013 and February 2014. On the contrary, the smallest values ($S < 0.15 \text{ cm min}^{-1}$) are attained from June–October 2012 and from August–September 2013. This is suggestive that in summer season one can expect that S is smaller (below the mean) than in winter season (above the mean).

6.2 Reconstructed sea level RSL

The variable RSL can be studied in the same way as the variable IS, and remarkably, the finding is the same. First the Student's t distribution is found to be the most suitable theoretical PDF for RSL. The long-term t distribution maximum-likelihood fitting leads to the values of $n = 4.25$ and $S = 0.64 \text{ cm}$, which shows that even the reconstructed sea level

variable RSL departs remarkably from a Gaussian distribution exhibiting higher tails. By using the case of September 2013, Fig. 11 shows the fitting PDFs computed by letting both parameters to vary, or only one (either S or n) as well as the PDF with the long-term parameter values. Eventually, Fig. 13 is analogous to Fig. 12. It provides monthly mean values and variances of the RSL EFD, showing that mean is practically zero, while variance changes significantly over months. The misfit graph provides the same ranking found for the variable IS. Misfits corresponding to the long-term PDF (black) are the worst, ranging from 4–18%, while misfits attained with a two-parameter fitting (violet) are the best, falling between 1–4%. When only the parameter n is estimated (red), misfits go from 2–14%, and eventually, when only S is evaluated (blue), misfits are restricted to the interval 1.5–6%. As a consequence, also for the variable RSL, one can state that monthly EDFs follow a t distribution where the degree of freedom is governed by the long-term regime, while the scale is governed by seasonal changes. The blue line in the S graph of Fig. 13 corresponds to PDF with $n = 4.25$ and shows us the monthly changes of the scale parameter for RSL. It turns out that the months with the largest values of S ($S > 0.76$ cm) are exactly the same as for IS (i.e. February 2012, April 2012, December 2012, May 2013 and February 2014), and analogously for the smallest values ($S < 0.55$), that is June–October 2012 and August–September 2013.

6.3 Control function CF

The EFDs of the variable CF are unimodal, positive, monotonically decreasing, right-tailed, with one inflection point. We found that they can be satisfactorily approximated by a three-parameter Beta distribution, as given here below:

$$B(X|a, b, S) = \frac{\Gamma(a+b)(X/S)^{a-1}(1-X/S)^{b-1}}{S\Gamma(a)\Gamma(b)}$$

where S is the scale parameter and a and b are two shape parameters, with $1 < a < 2$ and $b > 2$. Notice that the random variable X belongs to the limited domain $(0, S)$. In general S is found to be much larger than the largest empirical value of CF. Mean and variance are

given respectively by the following expressions:

$$\text{mean} = \frac{aS}{a+b}$$

$$\text{variance} = \frac{abS^2}{(a+b)^2(a+b+1)}$$

Estimates of the Beta distribution parameters for the entire 3-year period are $S = 5.1$, $a = 1.21$ and $b = 19.5$ yielding a long-term mean = 0.30 and variance = 0.66. The results of the statistical analysis concerning monthly EFDs are given in Fig. 14, that has the same structure as Figs. 12 and 13. Monthly mean and variance are given in the first two plots on the left. Misfits are shown in the bottom plot on the left. The misfit graph contains five cases: misfits computed with the long-term parameters given above (black), misfits obtained with full three-parameter fitting (violet), and misfits calculated with one free and two fixed parameters: a free (red), b free (green), S free (blue). In principle, since we use a three-parameter distribution, interpretation of the results is expected to be rather complicated. In practice, however, it is not and the key interpretation tool is once more the misfit graph. Misfits for the long-term PDF range between 2.5 and 6.5 %, while they reduce to 1.5–3 % when all parameters are free. In cases with constrained parameters they fall in between. Most importantly, one can observe that, differently from the cases of the variables IS and RSL, misfits are always rather small (less than 7 %), which means that the Beta distribution with the long-term parameters fits satisfactorily even all monthly EFDs. This suggests that the variable CF is not affected by seasonal variations and long-term estimates are sufficient to characterise the CF distribution.

7 Conclusions

A study of recorded sea level in the longwave range has been carried out for the harbour of Siracusa, Italy, as an example of application of a new method of analysis for coastal oscillations that is based on functions originally introduced in the early detection algorithm

called TEDA, devised for real-time identification of tsunamis and high-amplitude long waves on coastal recordings.

The classical tool for studying sea level is spectral analysis by means of which one can (1) characterise high-amplitude peaks and resonant properties of the basin or basins where the tide gauge is installed and (2) observe seasonal (inter-year) amplitude variations with respect to a multiyear typical spectrum. On carrying out the first step of our method, consisting in computing the FFT spectra of functions IS, RSL, BS and CF, we have proven that both issues, i.e. spectral peak sequence and seasonal cycle, can be recognised also in spectra of IS and RSL, that can therefore be considered as carrying the same information content as the spectra of the original sea level in the longwave window. Spectra of BS and CF have been found to show totally different features. In particular, we have seen that BS spectra show no peaks, but only a linear trend, and that CF spectra are stable and show no appreciable seasonal variations.

The second step of the method is the statistical analysis of IS, RSL, BS and CF where these quantities are seen as random variable occurrences. The main result is that the slope IS and the sea level RSL have been found to be distributed as a two-parameter Student's t distribution where the parameters are the degree of freedom n and the scale parameter S . Even more interestingly, it has been also found that n tends to be stable over years and can be considered as a long-term characterising parameter, while S shows seasonal variations with larger winter values and smaller summer values. As for CF, it has been found to follow a three-parameter Beta distribution with shape parameters a and b and scale parameter S , and furthermore it has been observed that the distribution remains stable over years with no influence of a seasonal cycle, which is consistent with the spectral analysis findings.

In conclusion, if spectral analysis of the longwave functions (step 1 of the method) cannot be considered an added value to the classical spectral analysis of sea level data, the statistical analysis (step 2) provides an innovative tool since characterises the tide-gauge site by means of well known theoretical distributions. This has the advantages that their parameters are easy to determine (e.g. through the maximum likelihood method adopted here) and that they are easy to handle in order to make any kind of probability assessments:

for example, it would be easily possible to estimate the exceedance probability of a given threshold for the reconstructed sea level or the sea level time gradient. Also noticeable is the separation between the value content of the parameters for IS and RSL statistics with n carrying information on the long term and S being representative of the seasonal cycle. Moreover, the stability of the CF statistics confirms that it is rather insensitive to changes of the sea conditions and thus it is the ideal function to use in algorithms for detecting alarming long waves, like TEDA.

We believe that our new analysis method provides a general and synthetic characterization of longwaves for coastal locations like Siracusa that could be routinely used to complement engineering studies in coastal areas, especially in harbours, where free oscillations and resonances might limit port operativity, even when their amplitudes do not imply flooding.

Appendix A

TEDA is an early detection algorithm designed to work in real-time that is composed by two parallel algorithms to detect long waves: the tsunami detection method (TDM) to detect long waves arriving with an impulsive front, and the secure detection method (SDM) to detect long waves that pass a specified threshold amplitude. The functions used by TEDA are the same defined and analysed in this paper. The sea-level slope IS (called instantaneous slope in TEDA) is computed in a short time interval including the most recent datum. The background slope BS represents the present sea-state in absence of hazardous long waves, and it is computed over a longer time interval. The main difference between TEDA and other detection algorithms, such as the ones based on the ratio STA/LTA is that BS is not computed on sea-level data on a longer time interval, but instead as a statistics of previous values of IS: e.g. $BS = \max(|IS_{DBS}|)$. The TDM compares IS and BS through the control function $CF = |IS|/BS$, which is shown to be rather insensitive to the sea-state variations. A tsunami detection is triggered when $|IS| \geq \lambda_{IS}$ and $CF \geq \lambda_{CF}$, where λ_{IS} and λ_{CF} have to be determined by means of a site-dependent calibration process.

The SDM is based on the filtered and detided reconstructed sea-level function RSL and is triggered every time RSL passes a set amplitude threshold, i.e. when $|RSL| \geq \lambda_{SD}$.

The detailed description of TEDA can be found in Bressan and Tinti (2011) and in Bressan et al. (2013).

A1 TEDA calibration

Sea-level signals in the longwave range at the coast depend strongly on the recording site since they are influenced by local conditions. Because of this site-specific characterization, TEDA, as well as other tsunami detection algorithms, should be calibrated, i.e. optimized for the local sea-level characteristics and for the tsunamis and longwave events typical of the site. The calibration procedure can be summarized with the following steps (for details see Bressan et al., 2013):

- Build a sea level background database;
- Build a tsunami database;
- Determine the parameters to test;
- Select thresholds by requiring no false detections on the background database;
- Test the tsunami database with all different combinations and thresholds;
- Select the parameters combination that detects most events in the shortest time.

In this application, the focus is posed on the durations, T_{IS} and T_{RSL} , of the time intervals to compute IS and RSL, and on the TDM thresholds λ_{IS} and λ_{CF} , that are the most important parameters. Instead, the other parameters are kept constant: namely T_{BS} is set to 60 min, T_G is set to 15 min, λ_{RSL} is set to 15 cm, and the alert time for the secure detection is set to 60 min. The lengths of the computation interval T_{IS} and T_{RSL} tested for the calibration of Siracusa tide-gauge are 4, 6, 8, 10 and 12 min.

A2 The databases

Four different background conditions have been selected for the calibration, including calm sea, rough sea, a seiche event and anomalous signal disturbances, that are probably due to the passage of boats close to the tide-gauge installation.

As for tsunamis, considering that there are no historical tsunami records for Siracusa, building a tsunami database means computing tsunami signals from potential sources. In our case we used sources identified by Tonini et al. (2011) to assess tsunami hazard for the Catania coasts, since they are considered relevant also for Siracusa. Tsunami scenarios consider three remote seismic faults responsible for the 365 earthquake and tsunami, located in the Hellenic Arc west of Crete, two local sources (a seismic fault and a landslide) that could be responsible for the 1693 tsunami and two potential sources (seismic fault and a combination of a fault and a submarine landslide) that could be the cause of the 1908 Messina tsunami (see Tonini et al., 2011; Bressan et al., 2013). In total we consider 7 tsunami cases as schematised in Table A1. Tsunami simulations have been carried out by means of the model UBO-TSUFDF developed and maintained by the Tsunami Research Team of the University of Bologna (see Tinti and Tonini, 2013).

The signals on which TEDA is tested are formed by combining, that is adding, the 4 selected recorded background conditions with the 7 computed tsunami time histories, giving a total of 28 cases.

A3 Calibration results

From the calibration process the best performing parameter configuration turns out to be $T_{IS} = 4$ min and $T_{RSL} = 10$ min, with $\lambda_{IS} = 5.35$ cm and $\lambda_{CF} = 1.65$, since it detects 26/28 events in an average time of 5.8 min. The results are given in Fig. A1, where TEDA detections are shown over the 7 synthetic tsunami signals forming the tsunami database for all 4 background conditions. TEDA calibration results seem to provide an efficient tsunami and long wave detection, except in case of very low amplitude events, like the two missed cases.

Acknowledgements. This work has been carried out in the framework of the EU project ASTARTE, grant agreement no. 603839, within FP7 ENV2013 6.4-3 and of the Italian project RITMARE.

References

- Bellotti, G.: Transient response of harbours to long waves under resonance conditions, *Coast. Eng.*, 54, 680–693, 2007.
- Bellotti, G. and Franco, L.: Measurements of long waves at the harbour of Marina di Carrara, Italy, *Ocean Dynam.*, 61, 2051–2059, 2011.
- Bellotti, G., Briganti, R., and Beltrami, G. M.: The combined role of bay and shelf modes in tsunami amplification along the coast, *J. Geophys. Res.*, 117, C08027, doi:10.1029/2012JC008061, 2012.
- Beltrami, G. M., Bargagli, A., and Briganti, R.: Gravitational mode calculation of basins discretized by orthogonal curvilinear grids, *Ocean Eng.*, 30, 833–853, 2003.
- Bressan, L. and Tinti, S.: Structure and performance of a real-time algorithm to detect tsunami or tsunami-like alert conditions based on sea-level records analysis, *Nat. Hazards Earth Syst. Sci.*, 11, 1499–1521, doi:10.5194/nhess-11-1499-2011, 2011.
- Bressan, L. and Tinti, S.: Detecting the 11 March 2011 Tohoku tsunami arrival on sea-level records in the Pacific Ocean: application and performance of the Tsunami Early Detection Algorithm (TEDA), *Nat. Hazards Earth Syst. Sci.*, 12, 1583–1606, doi:10.5194/nhess-12-1583-2012, 2012.
- Bressan, L., Zaniboni, F., and Tinti, S.: Calibration of a real-time tsunami detection algorithm for sites with no instrumental tsunami records: application to coastal tide-gauge stations in eastern Sicily, Italy, *Nat. Hazards Earth Syst. Sci.*, 13, 3129–3144, doi:10.5194/nhess-13-3129-2013, 2013.
- Guerrini, M., Bellotti, G., Fan, Y., and Franco, L.: Numerical modelling of long waves amplification at Marina di Carrara Harbour, *Appl. Ocean Res.*, 48, 322–330, 2014.
- Hasselmann, K., Barnett, T. P., Bouws, E., Carlson, H., Cartwright, D. E., Enke, K., Ewing, J. A., Gienapp, H., Hasselmann, S., Krusemann, P., Meerburg, A., Muller, P., Olbers, D. J., Richter, K., Sell, W., and Walsen, H.: Measurements of wind wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP), *Erganzungsheft zur Deutschen Hydrographischen Zeitschrift, Reihe A(8)*, 1–95, 1973.
- Hinwood, J. and Luick, J.: Closed basin modes of a dual basin harbour, *Pure Appl. Geophys.*, 170, 1881–1897, 2013.

- Jones, E., Oliphant, E., Peterson, P. et al.: SciPy: Open Source Scientific Tools for Python, available at: <http://www.scipy.org/> (last access: 6 August 2015), 2001.
- Kwak, M., Moon, Y., and Pyun, C.: Computer simulation of moored ship motion induced by harbour resonance in Pohang New Harbour, 33rd Conference on Coastal Engineering Proceedings, 1–6 July 2012, Santander, Spain, 1, 714–725, 2012.
- Lara, J. L., Martin, F. L., Losada, I. J., and Diaz, G.: Experimental analysis of long waves at harbour entrances, 29th Conference on Coastal Engineering, 19–24 September 2004, Lisbon, Portugal, 2, 1308–1320, 2004.
- Lopez, M., Iglesias, G., and Kobayashi, N.: Long period oscillations and tidal level in the Port of Ferrol, *Appl. Ocean Res.*, 38, 126–134, 2012.
- Miles, J. W.: Harbor seiching, *Annu. Rev. Fluid Mech.*, 6, 17–35, 1974.
- Luick, J. L. and Hinwood, J. B.: Water levels in a dual-basin harbour in response to infragravity and edge-waves, *Prog. Oceanogr.*, 77, 367–375, 2008.
- Okihiro, M. and Guza, R. T.: Observation of seiche forcing and amplification in three small harbours, *J. Waterw. Port C. Div.*, 122, 232–238, 1996.
- Okihiro, M., Guza, R. T., and Seymour, R. J.: Excitation of seiche observed in a small harbour, *J. Geophys. Res.*, 98, 18201–18211, 1993.
- Rabinovich, A. B.: Seiches and harbour oscillations, in: *Handbook of Coastal and Ocean Engineering*, edited by: Kim, Y. C., World Scientific, Singapore, 193–236, 2009.
- Rabinovich, A. B. and Thomson, R. E.: The 26 December 2004 Sumatra tsunami: analysis of tide gauge data from the world ocean Part 1. Indian Ocean and South Africa, *Pure Appl. Geophys.*, 164, 261–308, 2007.
- Rabinovich, A. B., Thomson, R. E., and Fine, I. V.: The 2010 Chilean tsunami off the west coast of Canada and the northwest coast of the United States, *Pure Appl. Geophys.*, 170, 1529–1565, 2013.
- Tinti, S. and Tonini, R.: The UBO-TSUFDF tsunami inundation model: validation and application to a tsunami case study focused on the city of Catania, Italy, *Nat. Hazards Earth Syst. Sci.*, 13, 1795–1816, doi:10.5194/nhess-13-1795-2013, 2013.
- Tonini, R., Armigliato, A., Pagnoni, G., Zaniboni, F., and Tinti, S.: Tsunami hazard for the city of Catania, eastern Sicily, Italy, assessed by means of Worst-case Credible Tsunami Scenario Analysis (WCTSA), *Nat. Hazards Earth Syst. Sci.*, 11, 1217–1232, doi:10.5194/nhess-11-1217-2011, 2011.

- Welch, P. D.: The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, IEEE Transactions on Audio and Electroacoustics, 15, 17–20, 1967.
- Wiegel, R. L.: Oceanographical Engineering. Englewood Cliffs, New Jersey, Prentice Hall, 532 p, 1964.

Table A1. Tsunami sources used for the tsunami hazard studies for the Catania coasts (Tonini et al., 2011) Fx = Fault x; E = Earthquake, L = Landslide.

Event	Type of source
365	F1, F2, F3
1693	E, L
1908	E, E+L

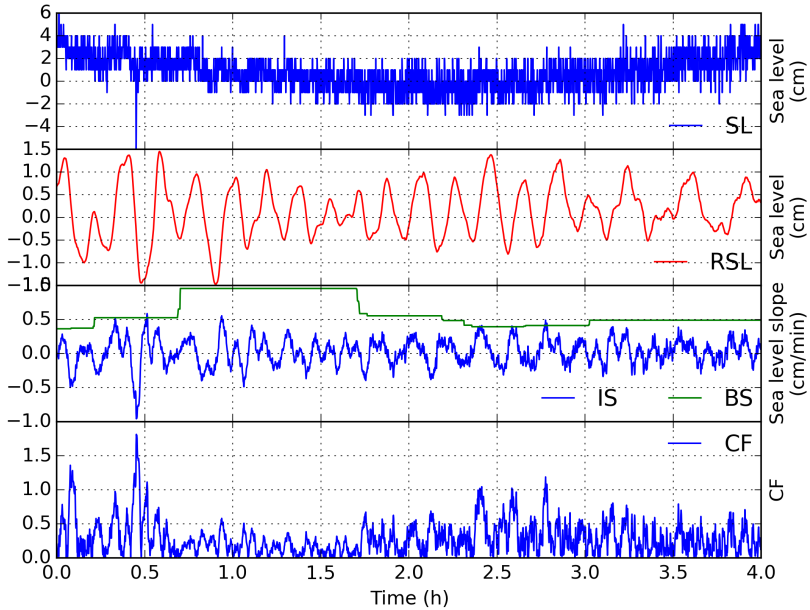


Figure 1. Sea-level of the Siracusa tide-gauge (top panel), the detided reconstructed sea-level RSL (second panel), the sea-level slope IS and background slope BS (third panel) and the control function CF for a sample of 4 h.



Figure 2. (a) Map of the Siracusa Targia harbour with position of the tide-gauge station belonging to the TSUNET network. (b) Augusta Bay, eastern Sicily. (c) Italy and Sicily. Images reprocessed from Google Earth.

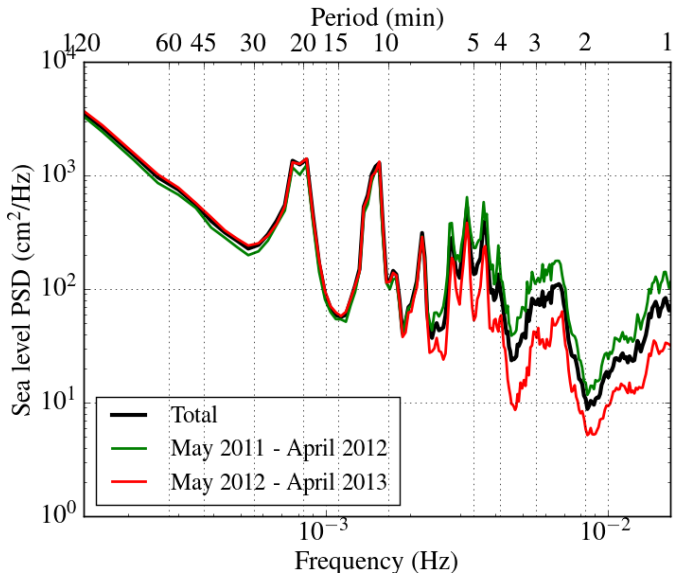


Figure 3. Average power spectra of the entire 25 month long interval of analysis from May 2011 to June 2013 (Total) and of two years, the first from May 2011 to April 2012 and the second from May 2012 to April 2013. The stability of the spectra is evident, which allows the identification of typical spectral peaks.

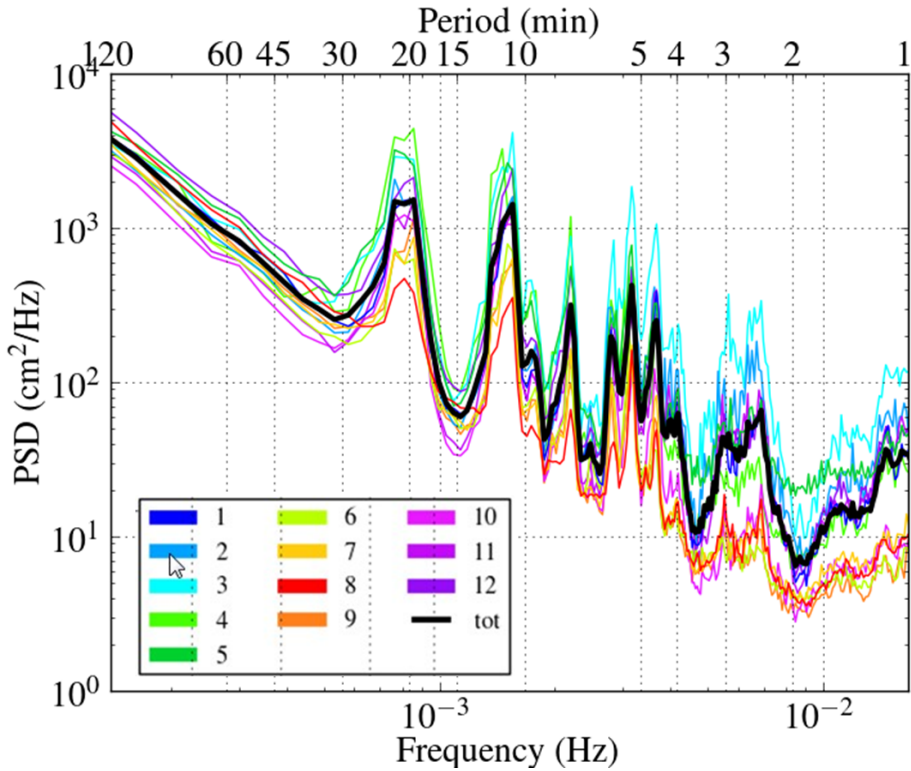


Figure 4. Average power spectral densities (PSD) for the year-long interval June 2012–May 2013 (tot), and for the corresponding calendar months from 1 = January to 12 = December.

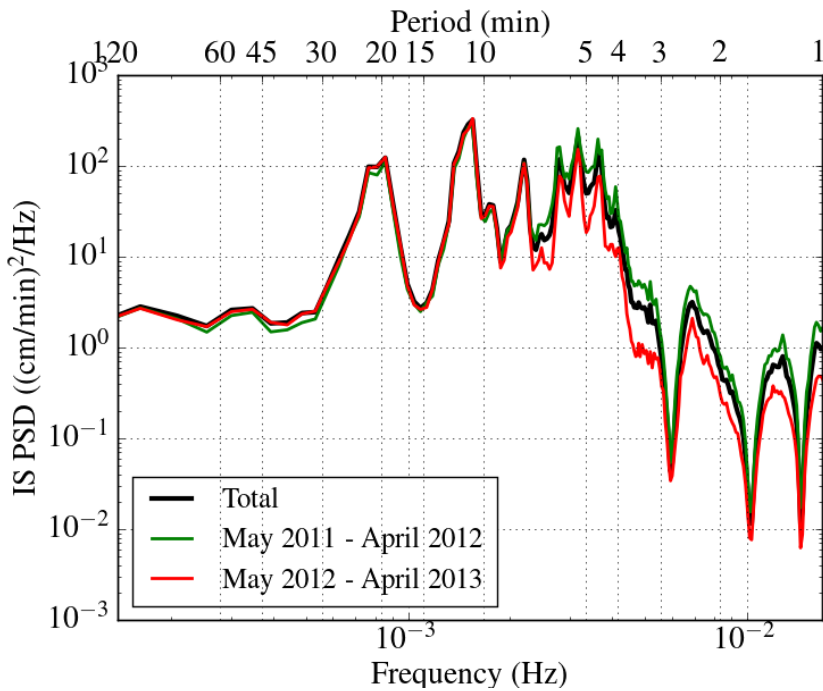


Figure 5. Average power spectra of the sea-level slope function IS computed with $T_{\text{IS}} = 4$ min. For further details see caption of Fig. 3.

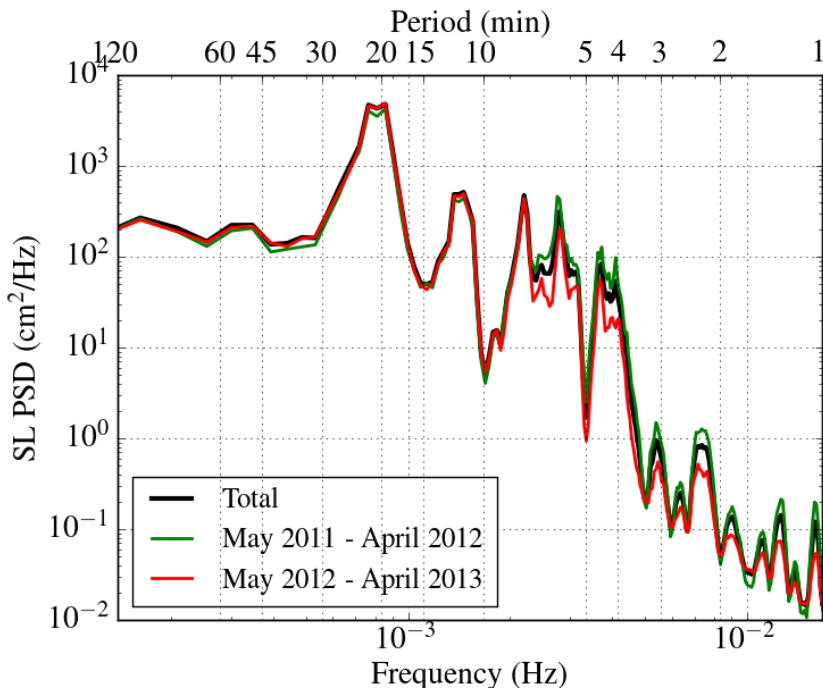


Figure 6. Power spectra of the long-wave reconstructed sea-level function RSL, computed with $T_{\text{IS}} = 4$ min and $T_{\text{RSL}} = 10$ min. For further details see caption of Fig. 3.

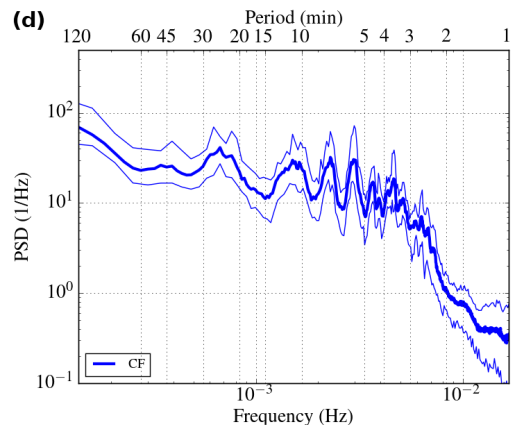
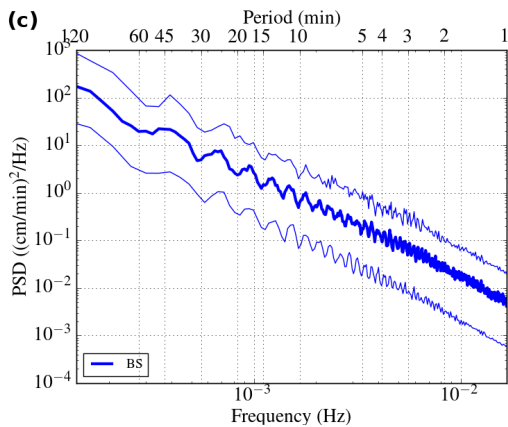
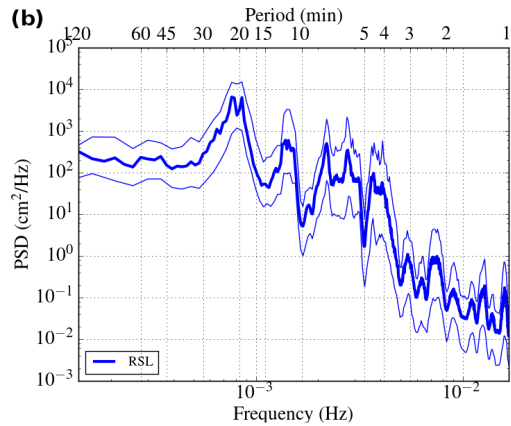
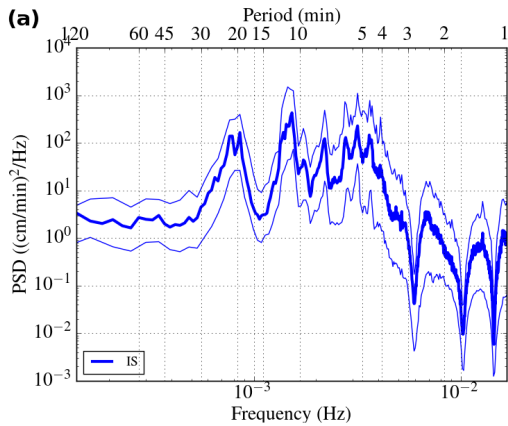


Figure 7. (a) Power spectra of IS together with the upper envelope PSD_{MAX} and the lower envelope PSD_{MIN} computed in the time interval June 2012–May 2013. The distance between envelopes measures the yearly dispersion which goes from one to two orders of magnitude and tends to grow with frequency. In computations $T_{\text{IS}} = 4$ min. **(b)** Power spectra of RSL together with the upper envelope PSD_{MAX} and lower envelope PSD_{MIN} computed in the time interval June 2012–May 2013. Even for RSL the gap between envelopes is in the range 10–100 and tends to grow with frequency. In computations $T_{\text{IS}} = 4$ min and $T_{\text{RSL}} = 10$ min. **(c)** Power spectra of BS together with the upper envelope PSD_{MAX} and lower envelope PSD_{MIN} computed in the time interval June 2012–May 2013. No resonant peaks appear in the graph. The ratio between envelopes is around 30–40 and frequency-independent. In computations $T_{\text{IS}} = 4$ min, $T_{\text{G}} = 15$ min and $T_{\text{BS}} = 60$ min. **(d)** Power spectra of CF together with the upper envelope PSD_{MAX} and lower envelope PSD_{MIN} computed in the time interval June 2012–May 2013. The ratio between envelopes is around 3–4, at least one order of magnitude less than for all other longwave functions, and frequency-independent in the range of interest 4–120 min. In computations the time constants are $T_{\text{IS}} = 4$ min, $T_{\text{RSL}} = 10$ min, $T_{\text{G}} = 15$ min and $T_{\text{BS}} = 60$ min.

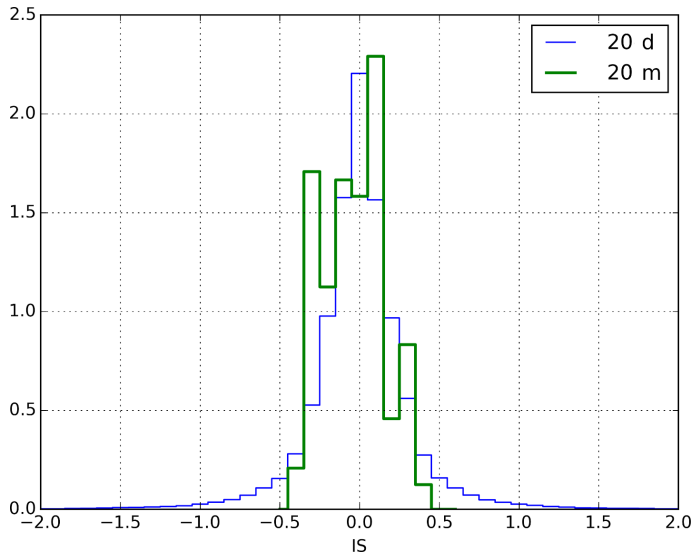


Figure 8. Normalised empirical frequency distributions for the variable IS corresponding to different time intervals: 20 min (green) and 20 days (blue). Longer time intervals lead to EDF of the same type as the blue curve. The unit of IS in the horizontal axis is cm min^{-1} . The width of each bin is 0.1 cm min^{-1} .

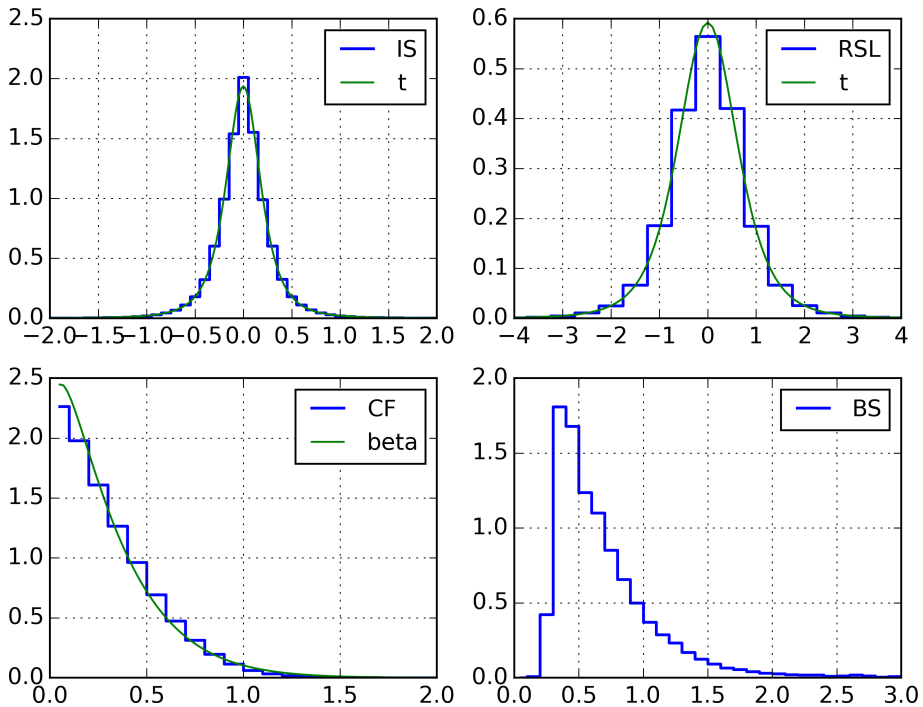


Figure 9. Examples of normalised empirical distributions for the four longwave functions of the method (IS upper left, RSL upper right, CF lower left, BS lower right) and the corresponding fitting PDFs (solid green curve). Units are cm min^{-1} for IS and BS, cm for RSL, whereas CF is dimensionless. The width of the bins in the proper units is 0.1 in the IS, CF and BS graphs, and 0.5 in the RSL graph.

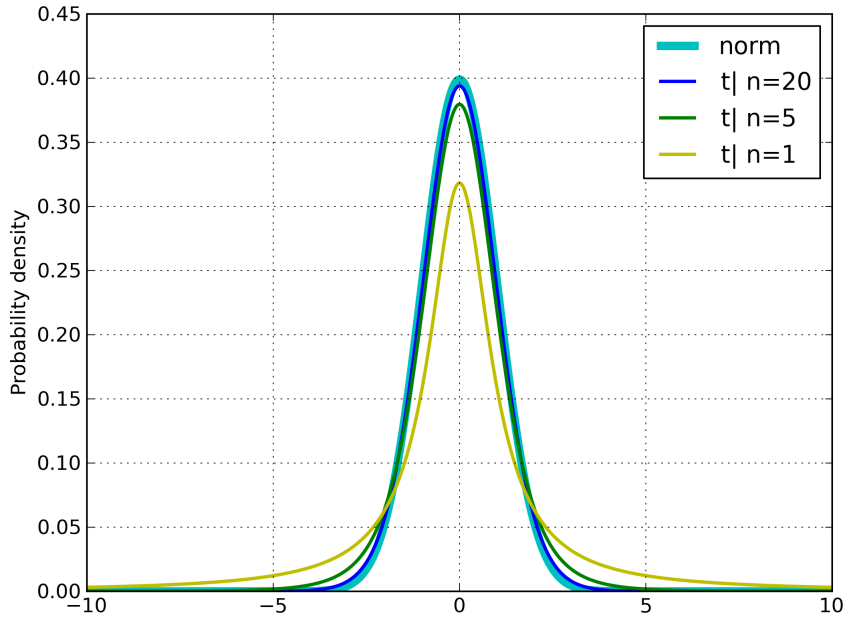


Figure 10. Normal Student t distribution for different values of degree of freedom n i.e. $n = 1, 5, 20$. It is a symmetric bell centred on the origin with tails higher than the Gaussian bell. When n grows larger, it tends to the Normal distribution (denoted by "norm"). With $n > 20$ the difference between Student's t and Normal distribution is quite small.

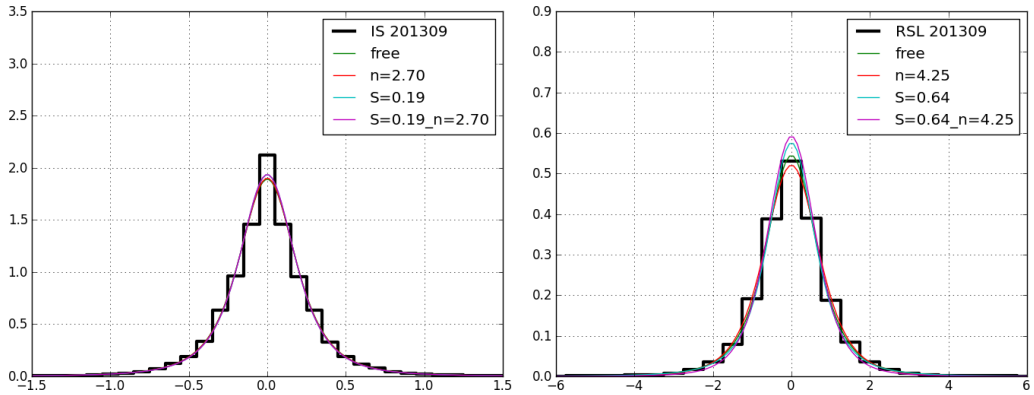


Figure 11. Normalised empirical frequency distributions of September 2013 for the variable IS (left) and RSL (right) with four different fitting Student's t distribution functions. The long-term parameter PDF is compared to PDF with free parameter estimation, and with PDFs where one of the two parameters is assumed to equal the long-term value. The horizontal axis unit for IS is cm min^{-1} and for RSL is cm. The respective bin widths are 0.1 cm min^{-1} and 0.5 cm .

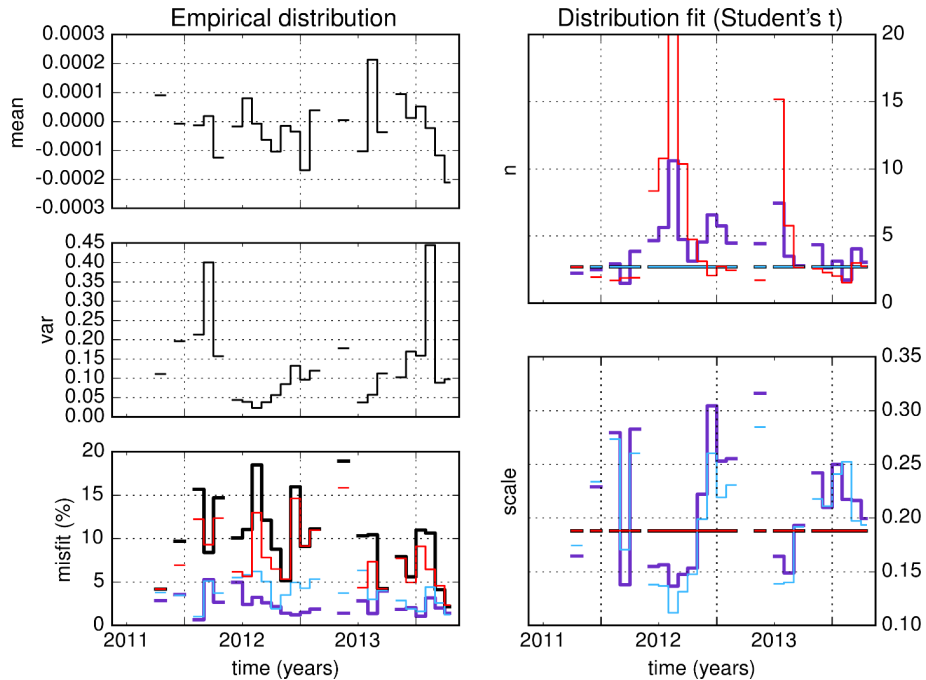


Figure 12. Statistical analysis of IS. Monthly EDFs mean (in cm min^{-1}) and variance (in $\text{cm}^2 \text{min}^{-2}$) are given in the first two plots on the left. Gaps in the graphs are due to the fact that only months with at least 85% of high-quality data are used. Misfits (last plot on the left) are computed for four cases: n and S kept constant (black); only S kept constant (red), only n kept constant (blue), full two-parameter fitting (violet). The same colour code is used for the graphs on the right showing monthly values of n (above) and of S (below, with scale unit in cm). The blue lines in the misfit and in the S plots can be taken as the final result of the analysis: the normalised variable IS/S obeys the Student's t distribution law with degree of freedom ($n = 2.70$) determined by the long-term regime.

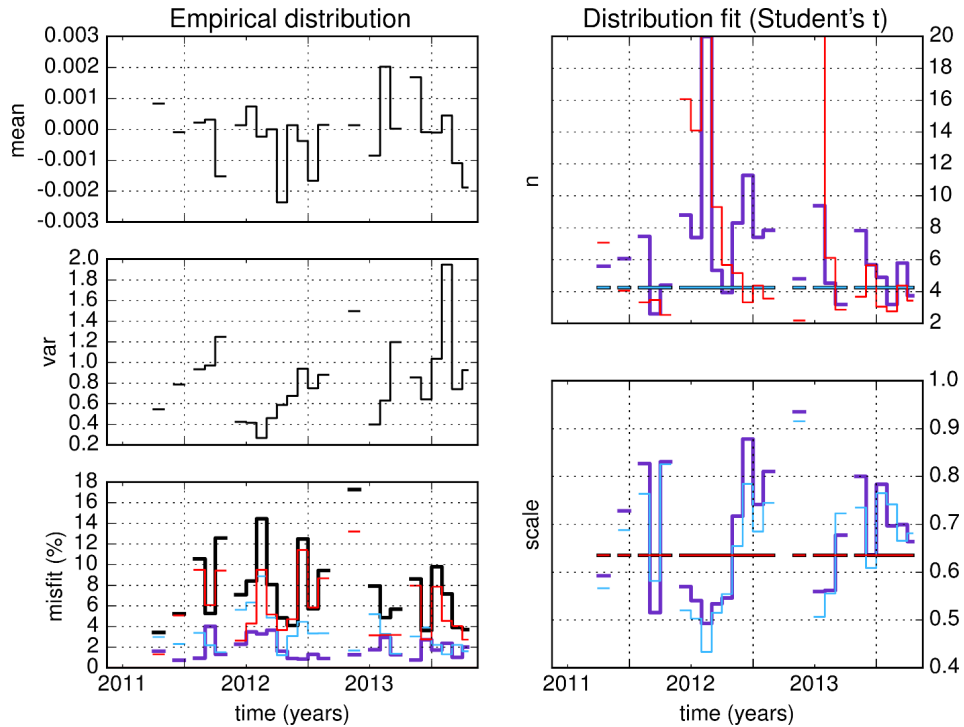


Figure 13. Statistical analysis of RSL. See captions of Fig. 12 for details. The measure units are cm for the mean and for the scale parameter S , and cm² for the variance. The long-term estimates for n and S are respectively 4.25 and 0.64 cm.

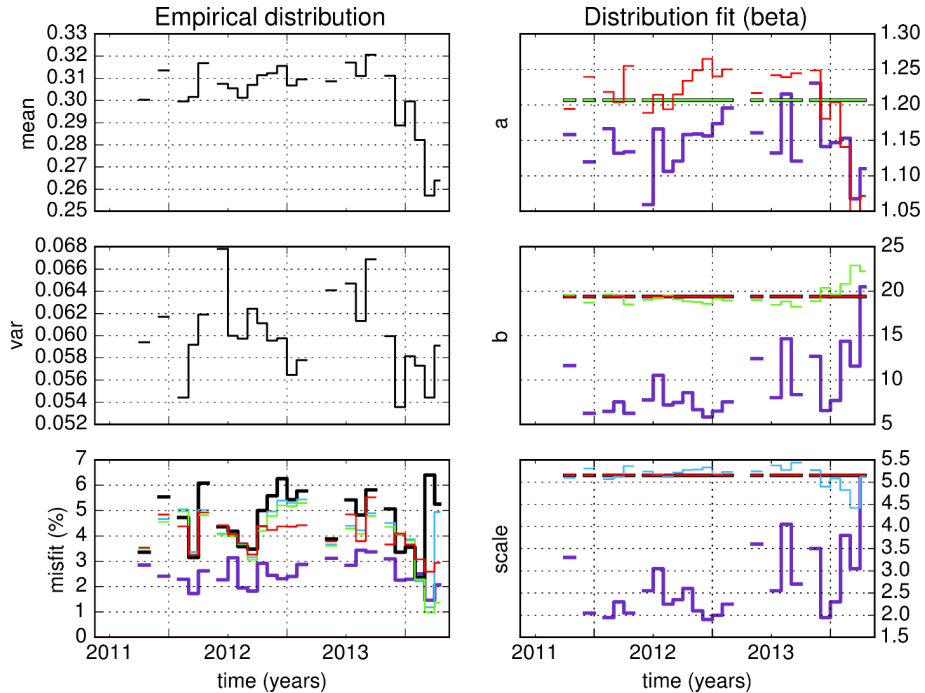


Figure 14. Statistical analysis of the variable CF. Since CF is dimensionless, also dimensionless are mean, variance and scale. Misfits are computed by keeping all the three parameters a , b and S of the Beta distribution constant (black), by estimating them one at a time, i.e. only a (red), or only b (green) or only S (blue), and by estimating all of them together (violet). The same colour code is used in the plots of the parameters on the right. The long-term values of the Beta PDF parameters are: $a = 1.21$, $b = 19.5$, $S = 5.15$.

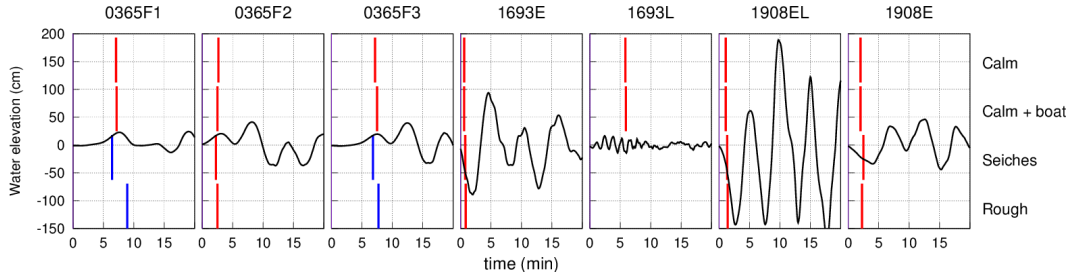


Figure A1. Results of TEDA detections with the optimal configuration setting for the Siracusa tide gauge. Each panel shows the tsunami wave computed for the Siracusa tide gauge from a given source. Vertical lines mark the time of TEDA detection triggered either by the TDM (red) or by the SDM (blue). From top to bottom the lines refer to different sea conditions of increasing amplitude: calm, calm + boat signal, typical seiche, and rough sea. Notice that for the weak tsunami case due to a landslide (1693L) TEDA could detect the event only in low-amplitude background. Notice further that of the 26 TEDA detections, as many as 22 are due to the TDM and only 4 to the SDM.