

Structure and performance of a real-time algorithm to detect tsunami or tsunami-like alert conditions based on sea-level records analysis

L. Bressan and S. Tinti

Physics Department, University of Bologna, Bologna, Italy

Received: 9 December 2010 – Revised: 4 March 2011 – Accepted: 8 March 2011 – Published: 19 May 2011

Abstract. The goal of this paper is to present an original real-time algorithm devised for detection of tsunami or tsunami-like waves we call TEDA (Tsunami Early Detection Algorithm), and to introduce a methodology to evaluate its performance. TEDA works on the sea level records of a single station and implements two distinct modules running concurrently: one to assess the presence of tsunami waves (“tsunami detection”) and the other to identify high-amplitude long waves (“secure detection”). Both detection methods are based on continuously updated time functions depending on a number of parameters that can be varied according to the application. In order to select the most adequate parameter setting for a given station, a methodology to evaluate TEDA performance has been devised, that is based on a number of indicators and that is simple to use. In this paper an example of TEDA application is given by using data from a tide gauge located at the Adak Island in Alaska, USA, that resulted in being quite suitable since it recorded several tsunamis in the last years using the sampling rate of 1 min.

1 Introduction

Tsunami Warning Systems (TWSs) are complex structures that include monitoring networks of seismic events and of sea surface elevation. In the common TWS practice, tsunami alerts follow standards that are designed specifically to provide clear information about the real tsunami threat. Tsunami alerts are initially issued only on the basis of seismic data, but, since the knowledge that can be extracted from seismic data alone on a possible tsunami generation is incomplete, it is very important for a TWS to get a quick “validation” of tsunami generation, either to confirm tsunami alerts or to

cancel them as soon as possible. Confirmation of tsunami generation comes only from the actual measure of tsunami waves, either recorded offshore or along the coast by tide gauges. The timely identification of tsunamis based on sea level records is therefore a topic of great importance in the common practice of any TWS, which was first solved only by visual inspection of the records. The problem of automatic detection was tackled relatively recently. One of the first examples of a tsunami detection algorithm was published by McGehee and McKinney in 1997, who proposed to estimate the tsunami amplitude by multiplying a short average of the sea-level record slope by the factor $T/2\pi$ where T is the expected predominant tsunami period. In the same year, Mofjeld (1997) from US NOAA (National Oceanic and Atmospheric Administration) devised a real-time algorithm to discriminate anomalous waves recorded by offshore buoys based on the deviation of the signal from the predicted tide. For coastal tide gauges, an automatic detection algorithm was in operation (Mero, 1998) till the DART buoy system was developed and activated. In British Columbia, Canada, tide gauges installed for tsunami recording are provided with a real-time algorithm, whose detections are used to warn responsible personnel in order to further investigate the tsunami event (Rabinovich and Stephenson, 2004). Recently, after the 2004 Indian Ocean tsunami, big efforts have been put worldwide to install new TWSs in the basins that were unprotected (e.g. Indian Ocean, Mediterranean Sea, Caribbean Sea) and to enhance the existing TWSs (e.g. the Pacific TWS). In parallel, the interest on real-time tsunami detection has increased in the scientific community (see e.g. recent contributions by Beltrami, 2008; Bellotti et al., 2009; Vela and Pérez, 2009; Wijeratne and Woodworth, 2009; Il-ligner and Schöne, 2009; Tolikova, 2009, 2010).

The present TWSs handle only tsunamis induced by earthquakes. However, tsunamis caused by submarine or coastal landslides can be quite disastrous and are not too rare, being in fact the second category of tsunamis in the tsunami



Correspondence to: L. Bressan
(lidia.bressan@unibo.it)

statistics as regards frequency. For such tsunamis, a warning based on seismic data is nowadays impracticable. Rapid detection at the coastline implemented on the tide gauge records might be the fastest way to release a warning for the population, though the time left to react can be quite short.

This paper presents the algorithm TEDA (Tsunami Early Detection Algorithm) and a method of performance evaluation that can be used to optimise the setting of TEDA parameters for a better adaptation to local ambient conditions and to user needs. TEDA is designed to detect the arrival of a tsunami as soon as possible or to sense hazardous sea level oscillations like big seiches or infragravity waves in coastal basins. It is based only on the analysis of sea-level data of a single station, and it is therefore independent from other data that could help tsunami identification, such as the time of the tsunamigenic earthquake occurrence, or the estimated tsunami arrival time. In this paper, the algorithm is described and assessed on sea-level time series recorded by a tide gauge installed in the harbour of Adak, the Adak Island, Alaska, USA. The harbour of Adak (51.872° N, 176.636° W) is found on the north-eastern coast of the island and is oriented EW, opening on the eastern side to a bay. This bay is in turn protected from the open ocean on the southern, western, and eastern side by the Adak Island and by other islands of the Andreanof Islands group. The tide gauge data are collected by NOS/CO-OPS of NOAA and are available online at a 1-min sampling rate at <http://tidesandcurrents.noaa.gov/tsunami/>.

The station selected to show how TEDA works is intentionally a coastal rather than an offshore station, since we want to stress the importance of coastal tide gauge records in tsunami detection. It is a frequent argument against coastal stations that they would allow no time for warning, since the tsunami would be identified when it is already attacking the coastal zone where the stations are installed. However, there is a number of reasons justifying the application of tsunami detection algorithms based on coastal sea level records. The first is that a tsunami could be identified before it floods the coast, if the tsunami leading front is negative and/or if flooding is not caused by the first waves. Second, one should bear in mind that detecting a tsunami at one coastal location is useful to all other locations that are more remote and are reached later. Third, the existing coastal tide gauge stations are more numerous than offshore tsunameters, and much cheaper to install and to maintain (and even to upgrade, if needed, to make them suitable for tsunami recording). Fourth, very many tide gauge stations are today integrated in the monitoring network of TWSs and equipped with real time data transmission facilities, so that implementing a real time detection algorithm seems to be feasible with limited economic efforts.

2 Description of TEDA and main principles

TEDA has the goal to detect tsunami waves or high-amplitude long-period waves that can be dangerous to people, harbours, and properties. It is composed of two parallel algorithms: the TEDA “tsunami detection”, that is focused on the identification of tsunami waves and is based on a signal slope discriminating approach; and the TEDA “secure detection”, which is based on incremental sea level estimation. Both algorithms work at station level, which has the advantage to use functions updated at every new data acquisition, without the need of waiting for a great load of data to make a decision on a possible alert.

2.1 TEDA tsunami detection

TEDA tsunami detection is based on the hypothesis that tsunami waves leave a signature in the sea level time series, and at any time (hereafter called the actual time t) it compares two time functions representing the Instantaneous Slope signal $IS(t)$ and the Background Slope signal $BS(t)$. Implicitly, TEDA assumes that a tsunami signal starts impulsively, with a break of the preceding signal that may be better seen by analysing variations of sea level rather than the sea level itself. More specifically, TEDA makes use of the average slope of the sea-level record corrected for tide, so that both functions $IS(t)$ and $BS(t)$ have the dimension of a sea level time change (cm min^{-1}). A sketch of all functions and time intervals used in TEDA is provided in Fig. 1.

To compute $IS(t)$, TEDA first calculates the average slope $IS_T(t)$ over the time interval I_{IS} of the most recent data, of length t_{IS} , going back from the actual time t (i.e. $I_{IS}(t) = [t - t_{IS}, t]$). Actually, $IS_T(t)$ is the slope of the straight line least-square fitting the sea level record within I_{IS} . $IS(t)$ is the detided version of $IS_T(t)$, which is obtained by subtracting from $IS_T(t)$ the estimated slope of the tide, say $Tide(t)$, i.e. $IS(t) = IS_T(t) - Tide(t)$. The function $IS(t)$ therefore represents the average slope of the sea level signal without the influence of the tide, and can be considered instantaneous if the duration t_{IS} of the interval I_{IS} is short enough. How TEDA computes the time function $Tide(t)$ will be explained later.

The TEDA Background Slope function $BS(t)$ represents the detided sea-level slope over a longer time interval $I_{BS}(t) = [t - t_G - t_{BS}, t - t_G]$, of duration t_{BS} , starting from time $t - t_G - t_{BS}$, with t_{BS} much larger than t_{IS} . The delay time t_G is a gap that is introduced to reduce the correlation between $IS(t)$ and $BS(t)$. Three ways of computing $BS(t)$ are possible for TEDA, which correspond to three methods designated hereafter as A1, A2, and A3. The corresponding functions are denoted by $BS1(t)$, $BS2(t)$, and $BS3(t)$. As is clear from the definitions given below, the Background Slope is computed by making use of the previous values of the Instantaneous Slope $IS(t')$, with t' belonging to the interval $I_{BS}(t)$:

$$BS1(t) = [\max(IS(t')) - \min(IS(t'))]/2; \quad t' \in I_{BS}(t) \quad (1a)$$

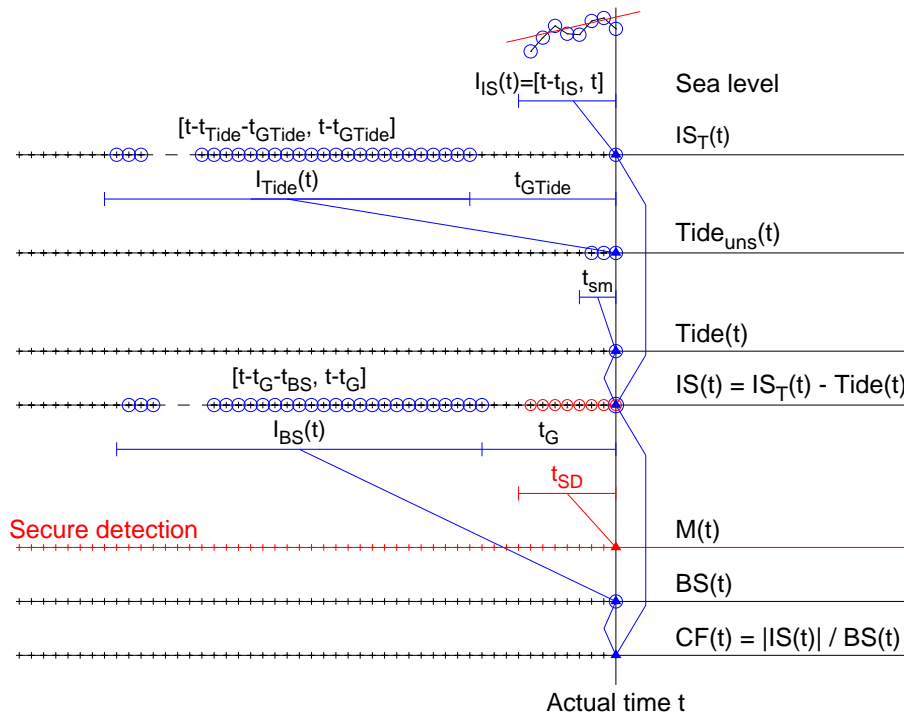


Fig. 1. Scheme of TEDA functions and of the TEDA time intervals described in the text.

$$BS2(t) = \text{standard deviation of } IS(t') \cdot \sqrt{2}; \quad t' \in I_{BS} \quad (1b)$$

$$BS3(t) = \max(|IS(t')|); \quad t' \in I_{BS} \quad (1c)$$

The Instantaneous Slope $IS(t)$ is compared to the Background Slope $BS(t)$ through the Control Function $CF(t)$, defined as the ratio $CF(t) = |IS(t)|/BS(t)$. TEDA assumes that a tsunami detection occurs at the actual time t when both $CF(t)$ and $|IS(t)|$ exceed a given threshold, that is:

$$|IS(t)| \geq \lambda_{IS} \quad (2a)$$

$$CF(t) \geq \lambda_{CF} \quad \text{i.e.} \quad |IS(t)| \geq \lambda_{CF} BS(t) \quad (2b)$$

Notice that λ_{IS} has the dimension of a slope and will be given in cm min^{-1} , while λ_{CF} is a number. Once a detection is triggered, a tsunami state condition starts, during which tsunami detection is suspended. During the tsunami, $BS(t)$ is expected to grow to values higher than normal, and therefore the tsunami state condition is assumed to last until $BS(t)$ reverts back to (or below) the value it had at the detection time. An example of how TEDA works is given in Fig. 2, where the application to a record of the 2006 Kuril Islands tsunami (later named E3 in this paper) is shown. Here, the original record and the time functions IS , BS , and CF are plotted. Notice how low the functions IS and BS are before the tsunami arrives. The arrival of the first tsunami wave causes IS to increase, together with CF , while BS increases with a delay equal to t_G . For this case, IS and CF pass the respective thresholds, therefore triggering a TEDA detection. During the tsunami, while the tsunami state is on, BS increases

much higher than the pre-tsunami level. In this example (see Fig. 2), soon after the tsunami state ends a new detection is triggered and the corresponding tsunami state lasts till the end of the ringing of the tsunami oscillations.

2.2 TEDA secure detection

TEDA secure detection has the goal to identify long period waves, such as tsunami or seiches, that are large enough to produce strong currents that might damage boats at moorings. For this purpose we have introduced the function $M(t)$, which is calculated by integrating the detected Instantaneous Slope $IS(t)$ over the interval $I_{SD}(t) = [t - t_{SD}, t]$ of duration t_{SD} , that is:

$$M(t) = \Delta t \sum IS(t'), \quad t' \in I_{SD}(t) \quad (3a)$$

where the summation is extended over all t' belonging to the interval I_{SD} and Δt is the sampling interval. TEDA secure detection triggers a warning every time the absolute value of $M(t)$ passes a given threshold λ_{SD} :

$$|M(t)| \geq \lambda_{SD} \quad (3b)$$

and starts an alert state of one hour duration. In case of consecutive exceedances, the alert state ends one hour after the last warning. The threshold λ_{SD} has the dimension of a length and will be given in cm.

We note that from the way it is computed, the function $M(t)$ is influenced by two characteristic times, namely t_{IS}

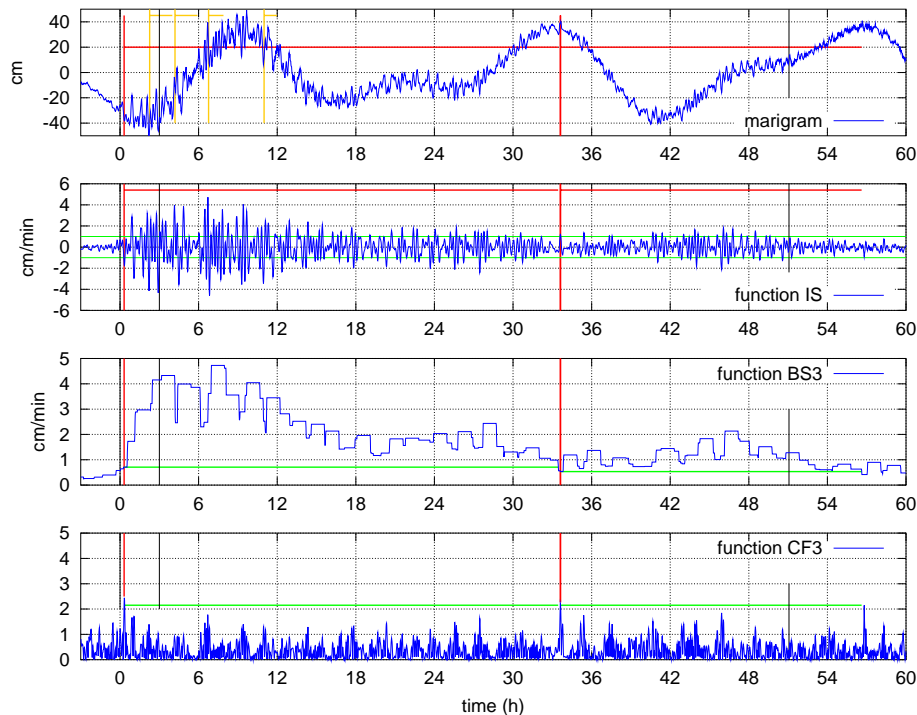


Fig. 2. Example of TEDA functions for the record of the Kuril Islands 2006 tsunami at the Adak tide gauge. Method used A3: $t_{\text{IS}} = 10$ min, $t_{\text{G}} = 15$ min, $t_{\text{BS}} = 60$ min, $\lambda_{\text{IS}} = 1$ cm min $^{-1}$, $\lambda_{\text{CF}} = 2.15$. Sampling interval is 1 min. From top to bottom: sea level original record, IS(t), BS3(t), and CF3(t). Time is counted starting from the tsunami arrival time ($t = 0$). In every panel, detections of tsunami are indicated with vertical red lines. In the first panel showing the recorded marigram, it also shows the tsunami state duration (red horizontal line), the secure detections (orange vertical lines), the secure alert duration (orange horizontal lines), and the end of the tsunami signal (black vertical lines). Notice that, soon after the first tsunami state ends (about 33.6 h), a new detection is triggered and the corresponding tsunami state lasts for about 24 h more. Horizontal green lines in the IS and CF3 function panels indicate the corresponding thresholds λ_{IS} and λ_{CF} , while in the BS3 panel they indicate the values assumed by BS3 at the detection time.

(involved in the calculation of IS) and t_{SD} . If we assume that the detided sea elevation is a pure sinusoidal signal of the type $A \sin(\omega t)$ with period $T = 2\pi/\omega$ and amplitude A and that IS(t) can be taken as the mean slope of such a signal in the interval t_{IS} , it may be easily shown that the function $M(t)$ assumes the following expression:

$$M(t; A, T, t_{\text{IS}}, t_{\text{SD}}) = A/(\omega t_{\text{IS}}) \{ [\cos(\omega(t - t_{\text{SD}})) - \cos(\omega(t - t_{\text{SD}} - t_{\text{IS}}))] - [\cos(\omega t) - \cos(\omega(t - t_{\text{IS}}))] \}$$

where the dependence of $M(t)$ from the relevant parameters of the signal (A, T) and of TEDA ($t_{\text{IS}}, t_{\text{SD}}$) has been made explicit. As expected, the function M is periodic with period T . By using trigonometric identities, it is easy to see that the above expression can be rewritten as:

$$M(t; A, T, t_{\text{IS}}, t_{\text{SD}}) = (2/(\omega t_{\text{IS}})) A \sin(\omega t_{\text{IS}}/2) \sin(\omega t_{\text{SD}}/2) \cos(\omega(t - t_{\text{SD}}/2 - t_{\text{IS}}/2))$$

or also as:

$$M(t; A, T, t_{\text{IS}}, t_{\text{SD}}) = \alpha A \sin(\omega(t - \delta))$$

where the gain factor α and the delay time δ are given respectively by:

$$\alpha = (2/\omega t_{\text{IS}}) \sin(\omega t_{\text{IS}}/2) \sin(\omega t_{\text{SD}}/2) \quad (3c)$$

$$\delta = (t_{\text{SD}} + t_{\text{IS}} - T/2)/2 \quad (3d)$$

The above analysis shows that the integral operator introduced in Eq. (3a) may be interpreted as a pass-band filter acting on the detided residuals of the sea surface elevation with lateral decay of 6 dB per octave, corresponding to an ω increase in the low-frequency range and to an ω^{-1} decrease in the high-frequency range. The gain (or transfer) function of such a filter is plotted in Fig. 3a. Examples of the function $M(t)$ are given in Fig. 3b where $M(t)$ is compared against the detided signal, which is computed by estimating the tide off-line through cubic-spline approximation. The position and width of the passing band depend on the time parameters t_{IS} and t_{SD} , both of which should be selected in such a way to filter out short-period waves and very long waves. In the examples shown in the figures, the passing band goes approximately from 15 to 50 min.

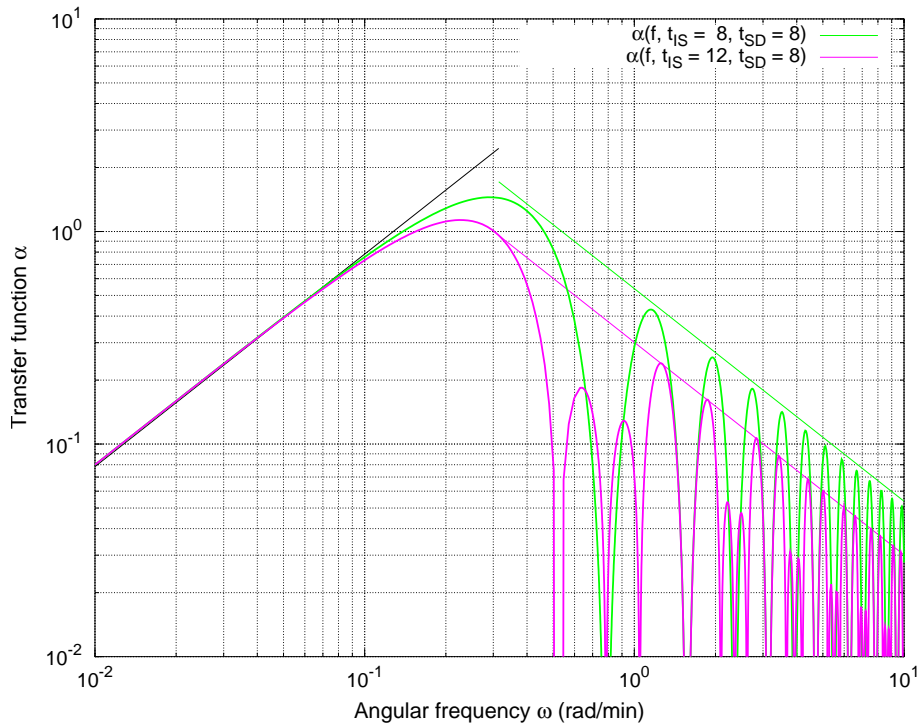


Fig. 3a. Transfer function $\alpha(\omega)$ given in Eq. (3c) for the function $M(t)$ for two different values of t_{IS} (6 min and 8 min) and for $t_{SD} = 8$ min. The transfer function decays as 6 dB per octave towards both high (see the envelope) and low frequencies. In the shown examples, the passing band covers the periods approximately between 15 and 50 min.

2.3 TEDA detiding algorithm

To detide the function $IS_T(t)$, a Tide(t) function is introduced, which is the Tide Slope Estimation. It is calculated on the basis of the previous values of the function $IS_T(t)$, and in particular on the interval $I_{Tide}(t) = [t - t_{Tide} - t_{GTide}, t - t_{GTide}]$, of length t_{Tide} , with a gap time t_{GTide} in order to make the Tide Slope Estimation independent from an incoming anomalous wave. The function Tide(t) is simply the extrapolation to the actual time t of a polynomial, least-square fitting the function $IS_T(t')$ with t' in $I_{Tide}(t)$. In this work a polynomial of degree zero has been used, so that in practice, Tide(t) is the average of the function $IS_T(t)$ over the interval $I_{Tide}(t)$. Observe that the length t_{Tide} of I_{Tide} should be appropriately adapted to the degree of the fitting polynomial. In case of degree 0, the time t_{Tide} should be short enough to allow one to approximate the tidal slope with a constant. The function Tide(t) is then averaged over an interval of length t_{sm} to further reduce oscillations of the tidal slope due to the influence of long period waves. In the application shown in this paper, the interval lengths used are $t_{Tide} = 60$ min and $t_{sm} = 6$ min. As for the interval t_{GTide} , the following relationships have been assumed: $t_{GTide} = t_G + 1$ min.

2.4 Why does TEDA use signal slopes?

Both algorithms of TEDA, i.e. the tsunami detection and the secure detection algorithms, make use of the slope of the signal rather than the signal amplitude. The main reason is related to the need for real-time detiding and the fact that real-time detided slope estimations are more accurate than estimations of real-time detided amplitudes. Indeed, it is quite difficult to get a detided signal with very small (and with a zero mean) tidal residuals through simple real-time filters (Tolkova, 2010; Kulikov, 1990). On the contrary, the detided slope signal presents smaller tidal residuals, and computing detided slopes is a more robust process. Indeed, detided slopes could be computed in the way it is performed by TEDA (i.e. by estimating the slope of the signal and of the tide, and then by subtracting the latter from the former), but also by following the alternative procedure (i) of estimating the tidal amplitude, (ii) of calculating the detided signal by difference, and (iii) of calculating its slope. It is to be stressed that even though the detided signal resulting from step (ii) may be affected by unacceptable tidal residuals, slopes resulting from step (iii) usually are not, and are quite similar to the one computed by the TEDA technique. From detided slopes, one can obtain unbiased detided functions by integration over a predetermined time interval, as is made in TEDA secure detection.

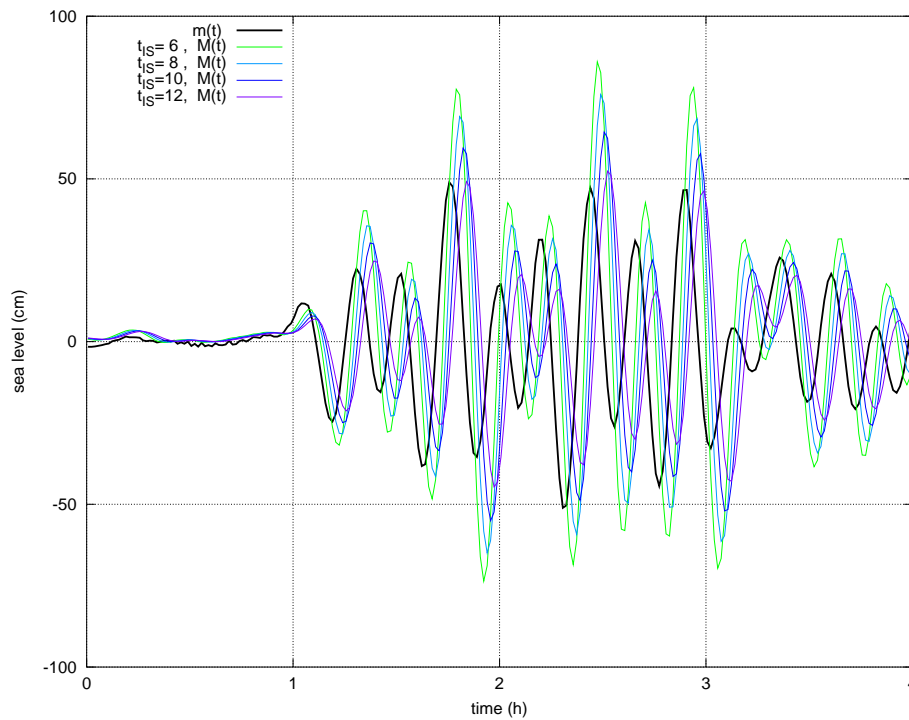


Fig. 3b. The figure shows the detided signal $m(t)$ of the Andeanov 1996 tsunami recorded at the Adak tide gauge together with the functions $M(t)$ of the TEDA secure detection computed for different t_{IS} values ($t_{IS} = 6, 8, 10, 12$ min) and for the same integration interval ($t_{SD} = 8$ min). Detiding of $m(t)$ is computed off line through the cubic splines fitting of the tidal signal and by further subtraction. For the selected combinations of parameters, it is seen that $M(t)$ is delayed with respect to $m(t)$ and amplified, with gain increasing as t_{IS} decreases.

3 TEDA performance indicators

In order to evaluate the performance of TEDA it is convenient to introduce one or more quantitative indicators that can be distinguished between elementary indicators and group indicators: in the first category we include all indicators referring to the analysis of a single event, while in the second category we put indicators concerning the analysis of groups of events.

We should point out that all indicators introduced in this section regard only the TEDA tsunami detection algorithm, i.e. how well TEDA performs in detecting tsunamis, since only in this case is it possible to express an evaluation: indeed, (1) in the case of a tsunami we can verify if TEDA tsunami detection algorithm sees it or not, and if it does, with what amount of delay and how many times, etc; and likewise (2) in case of no-tsunami, we can count possible false detections. We observe that concepts as true, missed and false detections, which are essential to build a set of performance indicators, loose meaning when they are applied to the TEDA secure detection algorithm. Both algorithms, i.e. TEDA tsunami detection and TEDA secure detection, are based on exceedance of prescribed thresholds (see Eqs. 2 and Eq. 3b), but in the first case the exceedance occurrence is assumed to reveal the occurrence of another event (the tsunami) and the goodness of this association can be judged objec-

tively, while in the second the exceedance occurrence is per se the event, and in a sense, the detection is always true and never false, by definition.

All indicators are not computed in real-time, but in an off-line mode. Here we concentrate on tsunami detections and introduce a number of elementary indicators, for which we first need to define the Tsunami Interval (TI), which is the interval where the tsunami signal is present: it starts at the tsunami arrival time and ends when the tsunami can be taken as finished. How this is identified in the actual records will be explained later. We distinguish between detections falling within the Tsunami Interval (TI) and detections falling within the first 3 h of TI that we call the Detection Window (DW). The Number of Tsunami Interval Detections is denoted as NTID, and the Number of “Acceptable” Detections is denoted as NAD. For records with no visible tsunami, it was assumed that TI and DW coincide and therefore NTID is equal to NAD. Further, we introduce the Number of False detections (NF), where a detection is considered false when it occurs outside the TI. Observe that if NAD is equal to zero this means that TEDA has failed to detect a tsunami, and this is a case of a missed detection. Another elementary performance indicator is the Delay Time (DT) of tsunami detection that is measured from the beginning of TI, and it is only used for the acceptable detections (i.e. when the $NAD \neq 0$) occurring

within the Detection Window (DW). A further elementary indicator that applies in cases of detections within TI is the Tsunami State Duration (TSP), defined as the percentage of TI that is covered by the active tsunami states triggered by all the detections falling within TI ($NTID \neq 0$). We reiterate that a tsunami state lasts from the tsunami detection instant until the function $BS(t)$ reverts to the pre-alert values. If TSP is quite low (that is much smaller than 100), it means that TEDA alerts cover only a little portion of the Tsunami Interval and that the tsunami is not recognised as such for most of the time. It might also happen that a detection occurs before the starting of TI and that the related tsunami state is active during the tsunami oscillations contributing to the determination of TSP. However, in such cases, the detection has to be considered a false detection and TSP is meaningless.

A further elementary indicator is built, starting from the values of the threshold λ_{CF} . It is clear, even with the help of Fig. 2 depicting the functions $IS(t)$, $BS(t)$, and $CF(t)$ calculated for a specific tsunami record, that changing the value of λ_{CF} influences the number of TEDA detections. If λ_{CF} is too small, then false detections are to be expected ($NF \neq 0$), while if it is too large, the tsunami can be undetected ($NAD = 0$). Therefore, the elementary indicators NF and NAD are seen to depend upon λ_{CF} and those values of λ_{CF} ensuring no false detections, and at the same time at least one acceptable detection can be considered adequate. In general, for each event two intervals can be defined on the λ_{CF} axis, descending directly from $NF(\lambda_{CF})$ and $NAD(\lambda_{CF})$: namely the No-False Detection Interval (NFI) and the Acceptable Detection Interval (ADI), whose formal respective definitions are given here below:

$$NFI = [\lambda_{CF} : NF(\lambda_{CF}) = 0] \tag{4a}$$

$$ADI = [\lambda_{CF} : NAD(\lambda_{CF}) \geq 1] \tag{4b}$$

Notice that NFI is an infinite interval since, when λ_{CF} is sufficiently large, no detections (either false or not) occur, and can be characterised by its lower end point NFI_1 that is therefore the limiting value of λ_{CF} beyond which no false detections are made by TEDA for the specific event under analysis. On the other hand, ADI is certainly a finite interval, therefore admitting a lower and an upper end point ADI_1 and ADI_2 . If an event passes undetected by TEDA, then ADI is an empty interval and introducing ADI_1 and ADI_2 in this case is useless.

In addition, we can define the Quality Detection Interval (QDI), which is the intersection of NFI and ADI and is therefore given by the following expression:

$$QDI = [\lambda_{CF} : NF(\lambda_{CF}) = 0 \quad \text{and} \quad NAD(\lambda_{CF}) \geq 1] \tag{4c}$$

Notice that QDI results to be an empty interval when the event is undetected (since ADI is empty) and when NFI_1 is larger than ADI_2 . In all other cases, QDI is not empty and its lower and upper end points can be denoted respectively as QDI_1 and QDI_2 , where $QDI_1 \equiv NFI_1$ and $QDI_2 \equiv ADI_2$.

Correspondingly, a Detection Function $DF(\lambda_{CF})$ can be introduced that is equal to 1 when λ_{CF} belongs to QDI and is zero elsewhere, that is:

$$DF(\lambda_{CF}) = 1 \quad \text{if} \quad \lambda_{CF} \in QDI \tag{5a}$$

$$DF(\lambda_{CF}) = 0 \quad \text{otherwise} \tag{5b}$$

When QDI is empty, it follows that DF is identically equal to zero. Through the indicator QDI we introduce a more stringent requirement on the mere event detection. Indeed, for all values of λ_{CF} belonging to ADI, TEDA makes a detection of the event, but it might be a “low-quality” detection since it might be accompanied by false undesired detections. Only detections occurring within QDI are high-quality, since TEDA detects the event with no false alarms. In the TEDA application shown in this paper, the concept of high-quality detections has been linked to the absence of false detections. In principle, the Number of False detections (NF) should be minimised, since measures taken in case of a wrongly declared tsunami state result in economic losses and undesired interruption of services. But the requirement can be relaxed by accepting, for example, a maximum number of false detections M_{FD} , that is, by imposing that $NF \leq M_{FD}$ instead of $NF = 0$. In the following, we will consider as detections only the high-quality detections as discussed here.

NFI , ADI , QDI , and $DF(\lambda_{CF})$ are elementary indicators since they are computed on each single tsunami event, though they are built on two more basic indicators such as $NF(\lambda_{CF})$ and $NAD(\lambda_{CF})$. On the other hand, starting from elementary indicators, other indexes can be built in TEDA that belong to the group indicator category, since they apply to more than one event and, preferably, to all the considered events. One key group indicator is built in the following manner. Let us consider a set of N tsunami events, and let ND be the number of those detected by TEDA ($0 < ND \leq N$), and let us further consider the $QDI(j)$ of the j -th detected event. Since we limit our attention to detected events, $QDI(j)$ cannot be an empty interval. We define as the Group QDI (denoted $GQDI$) the interval of the λ_{CF} , which is comprised between the maximum of the lower end points of $NFI(k)$ and the maximum of the upper end points of the $QDI(j)$, that is:

$$GQDI = [\lambda_{CF} : \max(NFI_1(k)) (k = 1, 2, \dots, N) \leq \lambda_{CF} \leq \max(QDI_2(j)) (j = 1, 2, \dots, ND)] \tag{6a}$$

It is easy to prove that the above definition is equivalent to the following one:

$$GQDI = [\lambda_{CF} : NF(\lambda_{CF}, k) = 0 \text{ for any } k \in [1, N] \text{ and } NAD(\lambda_{CF}, k) \geq 1 \text{ for at least one value of } k] \tag{6b}$$

where $NF(\lambda_{CF}, k)$ denotes the indicator $NF(\lambda_{CF})$ of the k -th event, and likewise for $NAD(\lambda_{CF}, k)$. In other words $GQDI$ is the interval of the λ_{CF} axis in which one or more of the N analysed events are high-quality detected. Correspondingly, with $GQDI$ it is convenient to define a Gain Function

Table 1. Scheme of TEDA performance indicators.

Detection Window DW (contained in TI)	Tsunami Interval TI	Background	
Number of Acceptable Detections $NAD(\lambda_{CF})$	Number of Tsunami Interval detections $NTID(\lambda_{CF})$	Number of False detections $NF(\lambda_{CF})$	
$DT(\lambda_{CF})$		Delay time (from the beginning of the TI interval)	
	$TSP(\lambda_{CF})$	Duration of the tsunami states (percentage of TI)	
Elementary performance indicators (one event)			
λ_{CF} interval definition		Missed event	Detected event
No-False Detection Interval:	$NFI = [\lambda_{CF}: NF(\lambda_{CF}) = 0]$	$[NFI_1, \infty]$	$[NFI_1, \infty]$
Acceptable Detection Interval:	$ADI = [\lambda_{CF}: NAD(\lambda_{CF}) \geq 1]$	\emptyset	$[ADI_1, ADI_2]$
Quality Detection Interval:	$QDI = [\lambda_{CF}: NF(\lambda_{CF}) = 0 \text{ and } NAD(\lambda_{CF}) \geq 1]$ $= NFI \cap ADI$	\emptyset	Quality detection: $= [QDI_1, QDI_2]$ $\equiv [NFI_1, ADI_2]$
Detection Function DF:	$DF(\lambda_{CF}) = 1, \quad \lambda_{CF} \in QDI$ $= 0, \quad \text{otherwise}$	0	$= 1, \lambda_{CF} \in QDI$ $= 0, \text{ otherwise}$
Group performance indicators (N events, of which ND detected)			
λ_{CF} interval definition			
Group Quality Detection Interval:	$GQDI = [\lambda_{CF}: (NF(\lambda_{CF}, k) = 0 \text{ for } \forall k) \text{ and}$ $(NAD(\lambda_{CF}, k) \geq 1 \text{ for at least 1 value of } k), \quad k = 1, \dots, N]$ $= (\cap_{k=1, \dots, N} NFI(k)) \cap (\cup_{k=1, \dots, ND} ADI(k))$ $= [\max(NFI_1(k)), \max(QDI_2(j))], \quad k = 1, 2, \dots, N \text{ and } j = 1, 2, \dots, ND$		
Gain function GF:	$GF(\lambda_{CF}) = \sum_{k=1, \dots, ND} DF(\lambda_{CF}, k), \quad \lambda_{CF} \in GQDI$ $= 0, \quad \text{otherwise}$		
Detection Tsunami Range:	$DTR(k) = [\lambda_{CF}: GF(\lambda_{CF}) \geq k], \quad 1 \leq k \leq ND$		

$GF(\lambda_{CF})$ that is obtained as the sum of the DF of all the ND events, that is:

$$GF(\lambda_{CF}) = \begin{cases} \sum_{k=1}^{ND} DF(\lambda_{CF}, k), & \lambda_{CF} \in GQDI \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

For each value of the detection threshold λ_{CF} , GF provides the value of the number of tsunami events that are detected by TEDA, hence assuming values that are comprised between 0 and ND. Moreover, we can also introduce the interval $DTR(k)$, what we call the Detection Tsunami Range and define as:

$$DTR(k) = [\lambda_{CF} : GF(\lambda_{CF}) \geq k], \quad 1 \leq k \leq ND \quad (8)$$

$DTR(k)$ is therefore the interval of λ_{CF} , where there are at least k event detections and no false detections.

The above indicators, when properly organised in criteria, serve to select the setting of the TEDA parameters provid-

ing the best performance of the algorithm. A scheme of the performance indicators is provided in Table 1.

4 The data

The elementary and group indicators for TEDA have been conceived as a set of tools to evaluate the performance of the algorithm. One of the main reasons to use them is to help select the TEDA parameter setting that optimises its efficiency. Indeed, before applying TEDA in a real-time mode to the records of a given coastal station, a test or learning phase should be carried out where an appropriate “learning data set”, consisting of previous records of that station, if available, or of similar stations, should be processed offline by means of different configurations of TEDA settings with the purpose to find the most adequate configuration, that is, the configuration that performs best.

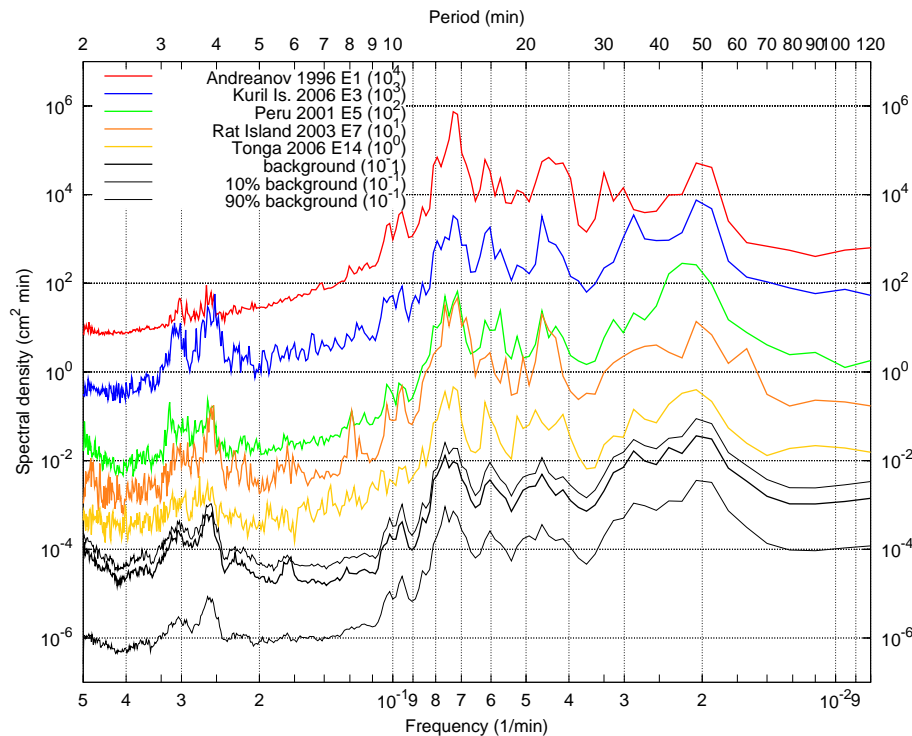


Fig. 4. Spectral analysis of the Adak Island tide gauge records. Sample interval is 1 min. The black line is the background spectral density (with no tsunami) averaged over 70 days, with the gray lines representing the corresponding 10th and 90th percentile. The lines above are average spectral densities computed for the Andreev 1996 (E1), the Kuril Islands 2006 (E3), the Peru 2001 (E5), the Rat Island 2003 (E7), and the Tonga 2006 (E14) events. Spectral densities curves are shifted through an amplification factor to make an easier inter-comparison. Spectral densities are calculated through FFT on 10.5 h moving windows over intervals of various length: for tsunamis, such intervals are adapted to the duration of the tsunami signal. Notice that all spectral densities have similar shapes, indicating that local eigenmodes are the dominant factor. The main spectral peaks are around 3–4, 10, 13–14, 16, 21–24, 35 and 48–52 min.

In order to show how TEDA works, a learning dataset is used in this paper that was made available by PMEL/NOAA and that consists of sea-level time series recorded by a tide-gauge station located in the Adak Island, Alaska, that is one of the oldest stations in the USA coastal station network, with a rich archive of records including many historical tsunami events.

The station is located inside a harbour, in a bay on the NE coast of the Adak Island. The local tide has a maximum range of about 2 m. The site is frequently windy, which is a factor favourable to the onset of local seiches that are amplified by the resonant effect of local basins (Rabinovich et al., 2006). One should bear in mind that often times, tsunamis and seiches go together because tsunamis can excite local resonances, with the consequence that tsunami spectra may be dominated by resonant peaks masking the source signature (see Honda et al., 1908; Miller et al., 1962; Miller, 1972; Sanchez and Ferreras, 1983; Van Dorn, 1984; Rabinovich, 1997; Rabinovich et al., 2006).

The Adak Island learning dataset includes about 141 days in the period 1996–2010, of which about 18 with tsunami signal and the remaining 123 with only background signal

(see Table 2). Spectral analysis has been carried out on the detided sea level records covering about 70 out of the 123 days with no tsunami events by applying the FFT technique over a moving window of 10.5 h. Only the background signal before the event, but not after, has been here considered to avoid any influence of the event on the calculated background spectral densities. One of the results of the analysis is that the average background spectral density is not so different from the ones computed in case of tsunamis. In Fig. 4, where such spectral densities are plotted, one sees that most of the spectral peaks of the (no tsunami) background persist even when a tsunami arrives. It is also relevant to stress that the background spectral density was found to remain significantly stable and that the same spectral peaks can be observed not only in the average curve, but also in all sea conditions (from calm to rough sea), which means that local oscillations are easy to excite and well persistent. The considerable number of peaks and their corresponding periods ranging from 2–3 min up to more than 30 min testify that eigenmodes of basins of various size and complicated geometry are involved (from the harbour basin where the tide gauge station is installed to the larger bay where the harbour

Table 2. Tsunami events. Date and location of the tsunami are listed, together with the magnitude of the tsunamigenic earthquakes. Event E4 is of volcanic origin. For each event, the total length (in days) of the examined record is given and it is further split into the length of the tsunami signal and the length of the background signal. Additional time series with no tsunami have also been examined, with length given in row B1. Altogether the total length of the background signal analysed is about 123 days. The event E1 is indeed a double tsunami with two earthquake sources: the main shock followed 11 h later by a smaller aftershock.

Event	Date	Earthquake magnitude	Record (days)	Event (days)	background (days)	Source location
E1	10 June 1996	7.9 7.3	6.66	0.57	6.09	Andreanof Islands, AK, USA
E2	27 February 2010	8.8	3.00	2.41	0.59	Chile Off Southern Coast
E3	15 November 2006	8.3	18.00	2.14	15.86	Kuril Islands, Russia
E4	7 August 2008	–	2.00	0.44	1.56	Kasatochi volcano, USA
E5	23 June 2001	8.4	4.98	1.06	3.92	South Peru
E6	15 August 2007	8.0	2.00	0.75	1.25	South Peru
E7	17 November 2003	7.8	10.12	0.51	9.61	Rat Island, Aleutian Islands, AK, USA
E8	26 December 2004	9.0	9.25	1.97	7.28	Indonesia off West Coast of Sumatra
E9	29 September 2009	8.0	2.00	1.80	0.20	Samoa
E10	7 October 2009	7.6	1.00	0.53	0.47	Vanuatu
E11	1 April 2007	8.1	3.00	2.33	0.67	Solomon Islands
E12	13 January 2007	8.1	2.25	0.57	1.68	Kuril Islands, Russia
E13	3 January 2009	7.6	1.00	0.57	0.43	North Coast of Papua New Guinea
E14	3 May 2006	8.0	23.00	1.70	21.30	Tonga
E15	5 December 1997	7.8	5.56	0.12	5.44	Kamchatka, Russia
E16	26 November 1999	7.5	3.33	0.12	3.21	Vanuatu
E17	25 September 2003	8.3	1.57	0.12	1.45	Hokkaido Island, Japan
B1	No tsunami	–	42.20	0.00	42.20	Background record
Total			140.92	17.71	123.21	

is placed and to channels separating the Adak Island from the other islands of the Aleutian archipelago). The most prominent peak seems to be located at 13–14 min.

5 Tsunami events

The records examined by TEDA in the application shown in this paper include 17 tsunami events, which are listed in Table 2 and that will hereafter be denoted by E_k ($k = 1, 17$). They have been ordered according to their descending magnitude in the local records, that is measured through the tsunami range (see last column of Table 3). These events were recorded by the Adak Island tide gauge from 1996 to 2010 with a sampling interval of 1 min. One event (E4) has volcanic origin, generated during the 2008 explosive eruption of the Kasatochi Volcano near the Adak Island. All the remaining events have seismic origin. The Andreanov 1996 (E1) and the Rat Island 2003 (E7) events can be considered near field, while all other earthquake-induced events are far field since they were generated along the South-American (Chile 2010 and Peru 2001, 2007), Russian (Kamchatka 1997, Kuril Islands 2006, 2007), and Japanese coasts (Hokkaido 2003); or near South Pacific Islands (Vanuatu 1999, 2009, Tonga 2006, Solomon Islands 2007, Papua New

Guinea 2009, Samoa 2009) and in the Indian Ocean (Sumatra 2004).

The spectral densities of some sea-level records are shown in Fig. 4, and, as already remarked in the previous section, they are quite similar to each other, which can be interpreted as the proof that tsunamis excite oscillation modes typical of the site. However, different tsunamis may be characterised by different source signature and so they may also exhibit distinctive peaks in the record: indeed, the E1 tsunami has a distinctive peak around 30 min and the E7 tsunami shows an additional peak with a period of 63 min.

The detided signals of the 17 tsunamis are plotted in Fig. 5. Notice that in this case, detiding was not performed by means of TEDA that always works on original sea-level records including tides and uses its own real-time detiding algorithm described in Sect. 2. Detiding was performed off-line with estimates of tide obtained by cubic splines fitting with the purpose to better show the tsunami signal in the graphs. The learning dataset includes events that occurred in different months of the year, which ensures that different climate conditions are taken into account in the application of the algorithm, which is reflected in the different magnitude of the background signal. All events have moderate amplitudes, with the largest being of about 50 cm for the tsunami E1 (see also Table 3). It is worth further noting that the event E1 is

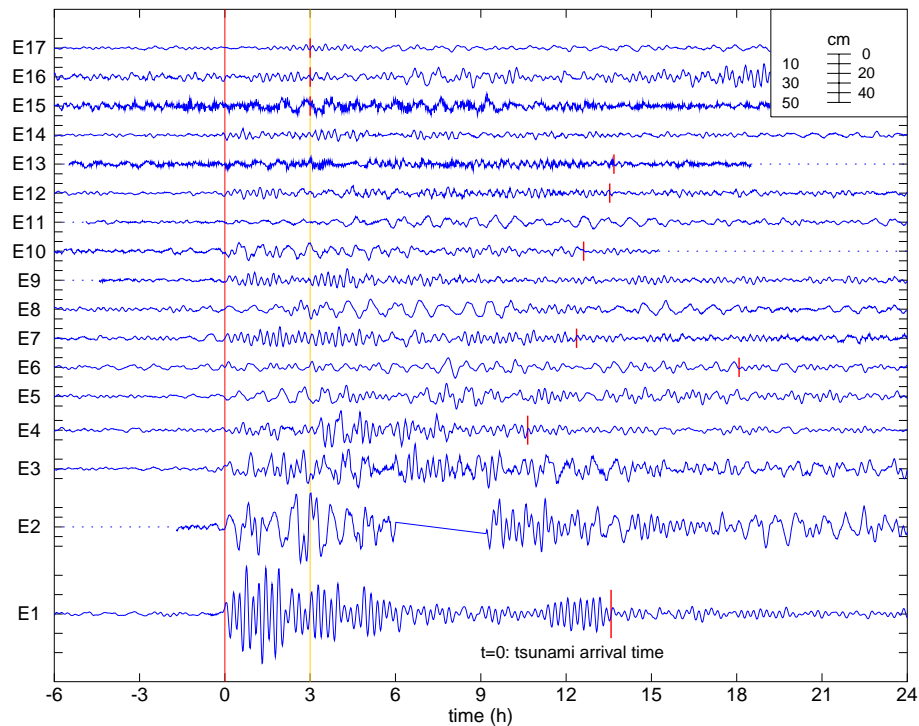


Fig. 5. Detided sea-level records of the 17 tsunami events analysed. Detiding was obtained through cubic splines fitting. The vertical red line at time 0 indicates the time of tsunami arrival, while the orange vertical line at time 3 h indicates the end of the tsunami Detection Window DW. The subsequent red vertical segments in each record are the end of the tsunami signal. The interval comprised between the beginning and the end of the tsunami is designated as the Tsunami Interval TI. For some events TI is longer than 24 h.

indeed a double one, with two tsunamis generated by two distinct earthquakes: an $M = 7.3$ aftershock following the main $M = 7.9$ shock 11 h later, which produced a second tsunami. In this work, the two tsunamis have been considered together as a single event.

In most cases the tsunami signal in the detided records has been identified manually by visual inspection, and the temporal limits of the tsunami beginning and end have been assigned correspondingly (see Fig. 5), also by taking into account data from papers (e.g. Eble et al., 1997; Wang and Liu, 2006) and from tsunami catalogues (NOAA/WDC Historical Tsunami Database at NGDC and WCATWC Tsunami Catalog), and by considering estimated arrival times from numerical simulations. We recall here that the identified interval between the start and the beginning of the tsunami has been denoted as Tsunami Interval (TI). In cases of an unclear tsunami signal, as is the case for the E15, E16 and E17 events, the tsunami start has been assumed to occur 30 min before the estimated arrival time, extrapolated from tsunami propagation maps obtained by means of Websift, NOAA, and the tsunami has hence been taken to last 3 h. In TEDA terminology, we can state that TI has been assumed to be equal to the tsunami Detection Window (DW): $TI = DW$. The assumed shifting of 30 min aims to account for uncertainty in the tsunami arrival time estimation. In Fig. 5, the tsunami

detided records are shown with indication of the identified or assumed tsunami start and end instants, while Table 3 provides information on the tsunami wave amplitudes and heights during the initial phase (first wave) as well as within the Detection Window DW and the Tsunami Interval TI.

6 Assessment of the tsunami detection algorithm performance

In the application shown in this paper, the performance of TEDA has been evaluated by using only a subset of the possible options of TEDA, namely by using all the methods A1, A2, and A3 to compute the function $BS(t)$ (see Eqs. 1a–c), and by varying the temporal parameters t_{IS} , t_G , and t_{Gide} , that enter in the definition of the basic intervals $I_{IS}(t)$ and $I_{BS}(t)$, and also in the detiding algorithm. Seven different combinations of the temporal parameters have been tried, which are named C_n ($n = 1, 7$) with association explained in Table 4. In total, therefore, as many as 21 TEDA configurations have been tested. The parameters that have been assumed as invariant in this application are $t_{BS} = 60$ min, $t_{Tide} = 60$ min, and $t_{sm} = 6$ min. Even the threshold λ_{IS} for $|IS|$ has been kept constant and equal to 1 cm min^{-1} . For each event (E_k) and each configuration C_n , TEDA tsunami detection has been

Table 3. Tsunami amplitudes and heights. The minimum and maximum sea elevations are denoted by m and M respectively: they refer to the first wave, to the Detection Window DW (i.e. the 3 h long interval within which a TEDA detection is considered successful), and the Tsunami Interval TI (i.e. the interval where tsunami signal is present). In particular, in the various columns the following quantities are given: the 1st half-wave amplitude (positive for a crest and negative for a trough); the 1st wave height; the largest peak in the DW; the largest peak in the whole TI; and the tsunami range, i.e. the difference between the maximum and the minimum in the DW and in the TI. Notice that events have been numbered according to the descending range of the tsunami waves within TI (last column). Given the difficulties in identifying the first wave for events E15, E16 and E17, the respective information is not given.

Event	1st half-wave amplitude:	1st wave height:	DW wave peak:	TI wave peak:	DW tsunami range:	TI tsunami range:
	m or M (cm)	$M-m$ (cm)	m or M (cm)	m or M (cm)	$M-m$ (cm)	$M-m$ (cm)
E1	12	36	-51	-51	100	100
E2	12	36	-37	-37	71	72
E3	8	15	19	26	35	45
E4	-4	12	-10	20	18	37
E5	-4	9	10	13	16	27
E6	6	9	6	-11	10	21
E7	5	9	12	12	20	21
E8	3	7	9	10	16	21
E9	5	6	8	12	14	20
E10	-6	15	-10	-10	19	19
E11	-2	4	-3	7	6	14
E12	2	5	-6	8	12	14
E13	-5	9	-6	-7	11	13
E14	-5	8	7	7	12	14
E15			-9	-9	18	18
E16			-5	-5	10	10
E17			3	3	6	6

Table 4. Combinations of the temporal parameters t_{IS} , t_G , and T_{GTide} . We recall that in this work, t_{BS} , t_{Tide} , and t_{sm} are kept constant and equal to $t_{BS} = t_{Tide} = 60$ min and $t_{sm} = 6$ min.

Combination	t_{IS} (min)	t_G (min)	T_{GTide} (min)
C1	6	16	17
C2	6	11	12
C3	8	16	17
C4	8	11	12
C5	10	16	17
C6	10	11	12
C7	12	16	17

tested with the CF threshold λ_{CF} ranging from 1.0 to 5.0 at steps of 0.05.

Table 5 shows the events that are detected. We recall that the learning dataset on which TEDA has been applied includes about 141 days of data and 17 tsunami events (see Table 2). We observe further that imposing no false detection for TEDA on the learning dataset, which includes about 123 no-tsunami days, does not guarantee that TEDA will always avoid false detections. However, this strict requirement will

certainly limit the number of false detections when TEDA is applied real-time in the operational routine. This topic will be touched upon later on in the paper.

We further recall that an event is detected by TEDA if the corresponding Detection Function DF in GQDI is not identically zero, i.e. if TEDA shows no false detection and $NAD > 0$ for at least one value of λ_{CF} (see Eqs. 4c and 5). The first immediate observation is that most of the tested configurations detect no events, or at most they detect only one (namely E1, which is the tsunami showing the largest oscillations – see Table 3). There are as many as 11 events missed by all configurations (namely E5, E6, E7, E8, E9, E11, E12, E13, E15, E16, E17). All the missed events are small and some of them (E15, E16, E17) cannot be recognised in the records, even in post-processing.

Looking at the results, it is evident that method A3 is the most efficient, while method A2 detects only the E1 event and only with combination C7. Therefore, method A2 turns out to be quite inefficient. This is a result to be stressed, since the way BS2 is computed, which is based on the calculation of the standard deviation of the background noise (see Eq. 1b), is one of the most common ways to characterise the background in signal detection techniques (see e.g. automatic seismic picking algorithms: Allen, 1978; Earle and Shearer, 1994), comparing short-term to long-term averages.

Table 5. Events detected (that is with $DF(\lambda_{CF})$ in GQDI not identically equal to zero) by all tested TEDA configurations, that are obtained by applying methods A1, A2 and A3 with different combinations of temporal parameters (see Table 3 for event codes and Table 4 for combinations).

Combination	Method A1	Method A2	Method A3
C1	–	–	E1
C2	E1	–	E1
C3	E1	–	E1
C4	E1	–	E1, E3
C5	E1, E14	–	E1, E2, E3, E4, E10, E14
C6	E1, E14	–	E1, E2, E3, E14
C7	E1, E3, E10	E1	E1, E2, E3, E4, E10

Table 6. Detection results for the configurations A3C5 and A3C7 for all events detected within the respective interval $DTR(5)$. $DTR_{A3C5}(5)$ ranges from 2.15–2.35, while $DTR_{A3C7}(5)$ ranges from 2.05–2.15 (see Fig. 6). The value at the detection time of functions IS, BS3 and CF3 is also given together with the Delay Time (DT), the Tsunami State Duration (TSP) and the Number of Tsunami Interval Detections (NTID) falling within the Tsunami Interval (TI), and the Number of Acceptable Detections (NAD) falling within the Detection Window (DW). To reiterate, the only detections within the DW are TEDA detections, though the NTIDs are not considered false detections and contribute to TSP.

Method	Event	$DTR(5)$	DT (min)	TSP %	NTID	NAD	IS(DT) (cm min^{-1})	BS3(DT) (cm min^{-1})	CF(DT)
A3C5	E1	2.15–2.35	2	99.7	1	1	1.30	0.48	2.72
A3C5	E2	2.15–2.20	159	4.0	1	1	–9.13	4.07	2.24
A3C5	E3	2.15	18	99.0	3	1	–1.60	0.71	2.26
		2.20	18	64.9	2	1	–1.60	0.71	2.26
		2.25	18	64.9	1	1	–1.60	0.71	2.26
		2.30–2.35	19	64.9	1	1	–1.73	0.71	2.44
A3C5	E4	2.15–2.35	38	94.1	1	1	1.35	0.57	2.36
A3C5	E10	2.15	27	58.7	1	1	1.25	0.57	2.19
		2.20–2.35	28	58.5	1	1	1.49	0.57	2.62
A3C5	E14	2.15	12	22.0	2	1	1.09	0.42	2.57
		2.20–2.35	12	15.2	1	1	1.09	0.42	2.57
A3C7	E1	2.05–2.15	2	99.7	1	1	1.01	0.44	2.27
A3C7	E2	2.05	28	9.6	1	1	–3.05	1.46	2.09
		2.10	29	9.6	1	1	–3.09	1.46	2.11
		2.15	160	4.0	1	1	–7.85	3.59	2.19
A3C7	E3	2.05	19	99.0	2	1	–1.40	0.59	2.38
		2.10–2.15	19	64.9	1	1	–1.40	0.59	2.38
A3C7	E4	2.05–2.15	39	93.9	1	1	1.13	0.47	2.39
A3C7	E10	2.05–2.10	28	58.8	1	1	1.08	0.51	2.12
		2.15	29	58.7	1	1	1.30	0.51	2.55

Methods A1 and A3 work better with almost all parameter combinations, detecting at least one tsunami event, with the exception of A1C1 that misses all tsunamis. The maximum number of events detected is six, which was by method A3C5, followed by five events detected by A3C7, four events detected by A3C6, and three by A1C7.

The performance of the various TEDA configurations can be further evaluated by means of the Group Quality Detection Intervals, GQDIs, defined by Eq. (6a), which are plotted in Fig. 6, and from which one can also easily deduce the resulting Gain Function $(GF)(\lambda_{CF})$ and the corresponding Detection Tsunami Ranges $DTR(k)$. In terms of $GF(\lambda_{CF})$, we

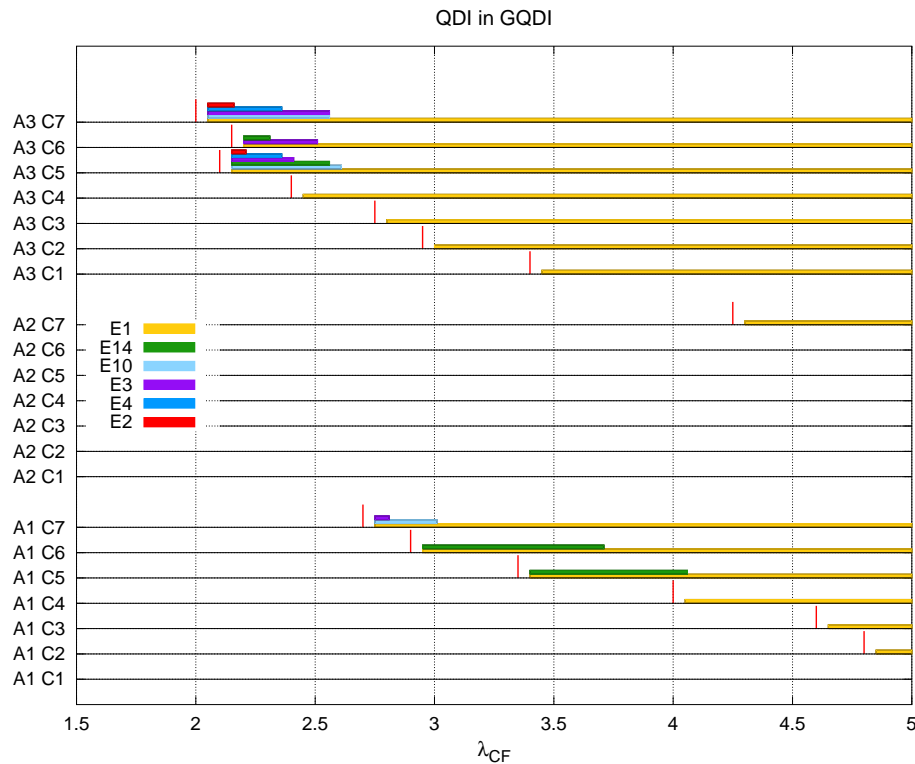


Fig. 6. For every TEDA configuration (i.e. for each method and parameter combinations), the Group Quality Detection Intervals (GQDI, see Eq. 6a) are given (in orange). In addition the intervals QDI of the detected events are plotted but only when they overlap with the GQDI. We observe that the QDI plot of the event E1 would coincide identically with the GQDI for all configurations. From this figure one can deduce easily both the gain function $GF(\lambda_{CF})$ as well as the Detection Tsunami Ranges DTR. In the text particular attention is given to the range $DTR(5)$. Vertical red bars indicate the last value of the tested λ_{CF} for which $NF \neq 0$.

can rephrase the above statements by stating that the maximum value of GF is 6 for the configuration A3C5; and is 5 for the configuration A3C7. These configurations turn out to be the best performing and are therefore the ones on which the attention will be focused in the following.

In order to compare the detection results obtained by A3C5 and A3C7, we first take into account the interval $DTR(5)$ as defined through the expression (8), which is the interval of λ_{CF} where both reveal at least 5 events: $DTR_{A3C5}(5) = [2.15, 2.35]$ and $DTR_{A3C7}(5) = [2.05, 2.15]$. Results are shown in Table 6. We observe that both intervals are quite short. We further observe that the analysis of such TEDA indicators cannot help give an univocal evaluation. Indeed, we note that the length of the $DTR(5)$ range is longer for A3C5, which is an element in favour of A3C5, since a longer DTR suggests that a tsunami can trigger a detection alert for a wider range of threshold values and therefore allows more flexibility for TEDA operational procedures. On the other hand, $DTR_{A3C7}(5)$ contains smaller values than $DTR_{A3C5}(5)$, which is in favour of A3C7, since lower thresholds imply that the corresponding configuration would have the potential to trigger a detection for smaller tsunami. The elementary indicators DT and TSP are given in

Table 6 and are also compared in Fig. 7, while Fig. 8 shows the DTs graphically marked on the detected records of the detected events. From the upper panel of Fig. 7 regarding DT, it can be seen that there are no substantial differences between the Delay Times obtained by the two configurations for four events, namely E1, E3, E10, and E4, listed here in ascending order of DT. The main differences regard events E14 and E2. Event E14 is seen quite soon ($DT = 12$ min) and only by A3C5. In contrast, event E2 is detected by both configurations: it is seen by A3C5 at time $DT = 159$ min ($\lambda_{CF} = 2.15, 2.20$), out of scale in Fig. 7; while A3C7 gives much faster detections only for thresholds $\lambda_{CF} = 2.05, 2.10$, with Delay Times respectively $DT = 29, 30$ min. Notice that raising the threshold to $\lambda_{CF} = 2.15$ causes the Delay Time to increase to $DT = 160$ min.

The comparison of TSP (Table 6 and Fig. 7) confirms that values resulting from the two configurations are quite similar for the group of four events also having similar DTs. Notice further that these are the only ones possessing TSP values about or larger than 60%, with a consequent tsunami state that covers at least 50% of the tsunami duration. Some differences can be observed for the other events for which, however, TSP is too low (less than 25%).

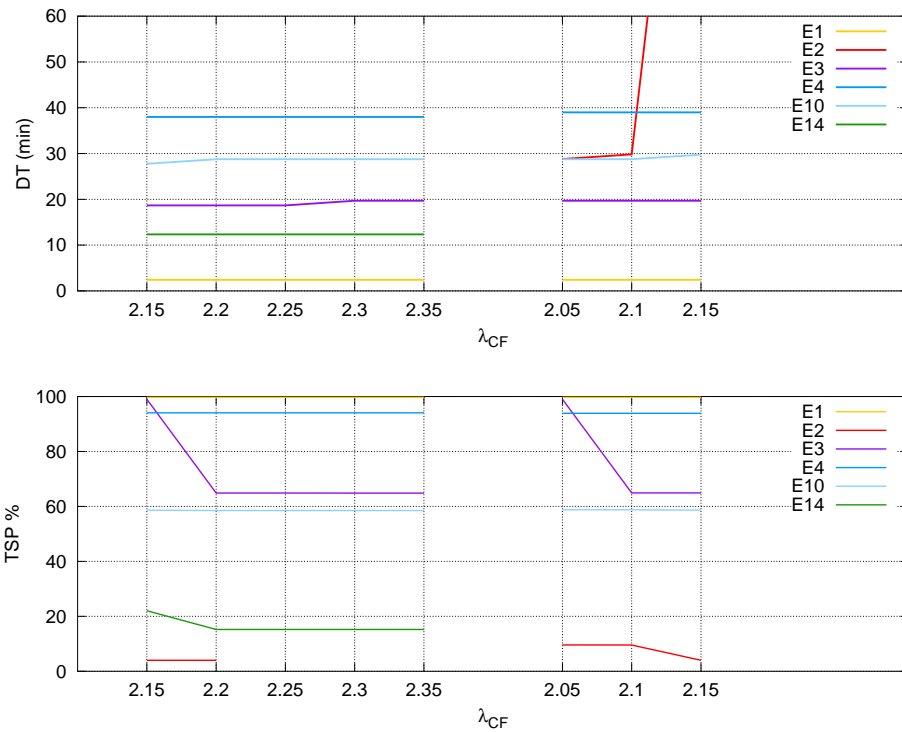


Fig. 7. Comparison of Delay Time (DT) and Tsunami State Duration (TSP) for configurations A3C5 (on the left) and A3C7 (on the right) within the Detection Tsunami Range (DTR(5)). Every event can be identified by a different color.

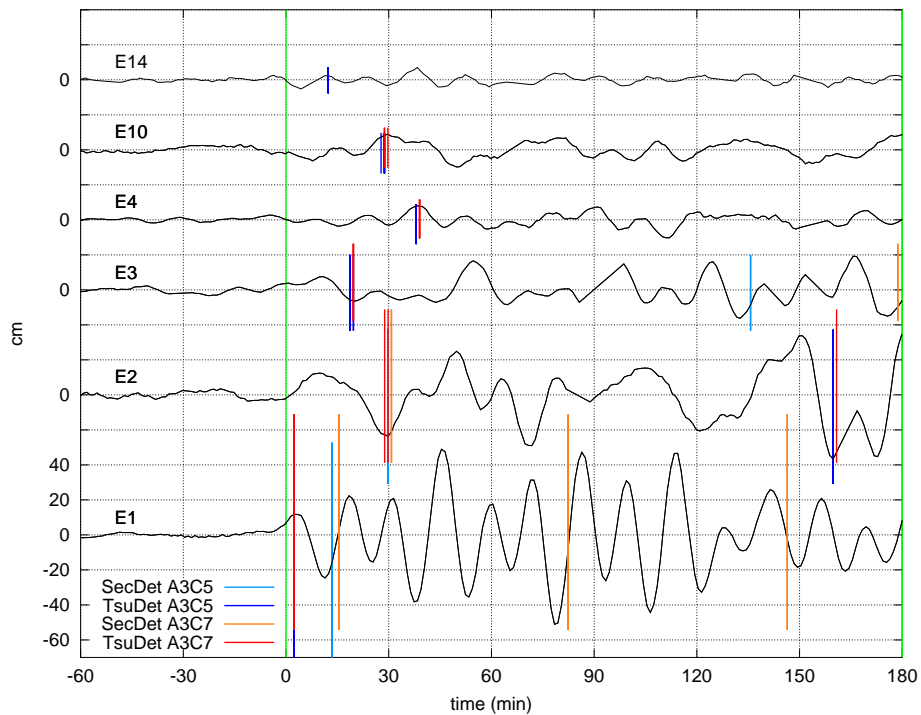


Fig. 8. Comparison of TEDA detections resulting from configurations A3C5 and A3C7 by applying the tsunami detection and the secure detection procedures. All thresholds λ_{CF} belonging to the respective DTR(5) have been considered. Delay Times (DT) (vertical lines) are shown for both TEDA tsunami detections and for TEDA secure detections on the detided records for the six events that are seen by TEDA. The vertical scale is the same in all graphs. Records are plotted for a 4 h long interval, including the Detection Window (DW).

Table 7. Values of TEDA functions IS and BS3 within the tsunami Detection Window (DW) and within the preceding time interval of 20 min length for both configurations A3C5 and A3C7. Time is measured from the beginning of the Tsunami Interval (TI). It is interesting to compare the maximum value of BS3 in the 20 min long interval with the maximum IS value of the Detection Window. Notice that for events E6, E11, E13, E17, the function $|IS|$ remains below the threshold $\lambda_{IS} = 1 \text{ cm min}^{-1}$ within the entire Detection Window.

Event	Time interval (min)	A3C5			A3C7		
		max(IS) (cm min^{-1})	max(BS3) (cm min^{-1})	mean(BS3) (cm min^{-1})	max(IS) (cm min^{-1})	max(BS3) (cm min^{-1})	mean(BS3) (cm min^{-1})
E1	[-20,0]	0.5	0.56	0.49	0.42	0.46	0.45
E1	[0,180]	13.17	13.17	9.6	10.51	10.51	7.66
E2	[-20,0]	0.97	0.52	0.52	0.83	0.44	0.44
E2	[0,180]	9.13	9.13	3.98	7.85	7.85	3.36
E3	[-20,0]	0.67	0.58	0.56	0.54	0.48	0.46
E3	[0,180]	4.33	4.15	2.52	3.65	3.41	2.14
E4	[-20,0]	0.25	0.44	0.44	0.26	0.37	0.37
E4	[0,180]	2	2	1.41	1.72	1.72	1.17
E5	[-20,0]	0.28	0.33	0.33	0.28	0.3	0.3
E5	[0,180]	1.65	1.34	0.93	1.5	1.23	0.85
E7	[-20,0]	0.9	0.9	0.6	0.81	0.8	0.55
E7	[0,180]	2.67	2.67	1.79	1.99	1.99	1.4
E8	[-20,0]	0.4	0.89	0.74	0.44	0.75	0.66
E8	[0,180]	1.47	1.47	0.67	1.19	1.19	0.59
E9	[-20,0]	0.36	0.36	0.32	0.31	0.31	0.28
E9	[0,180]	1.85	1.85	1.41	1.36	1.36	1.08
E10	[-20,0]	0.57	0.5	0.5	0.51	0.46	0.46
E10	[0,180]	1.82	1.82	1.41	1.6	1.6	1.2
E12	[-20,0]	0.33	0.42	0.42	0.27	0.35	0.35
E12	[0,180]	1.58	1.58	1.19	1.29	1.29	0.97
E14	[-20,0]	0.47	0.44	0.44	0.37	0.41	0.41
E14	[0,180]	1.42	1.42	1.07	1.17	1.17	0.89
E15	[-20,0]	0.73	1.04	1.04	0.58	0.77	0.77
E15	[0,180]	1.45	1.45	1	1.32	1.32	0.83
E16	[-20,0]	0.69	0.92	0.92	0.53	0.8	0.8
E16	[0,180]	1.32	1.32	1.04	1.1	1.1	0.85
E6	[0,180]	0.91	0.91	0.84	0.77	0.77	0.71
E11	[0,180]	0.45	0.45	0.38	0.39	0.39	0.34
E13	[0,180]	0.92	0.92	0.57	0.81	0.81	0.52
E17	[0,180]	0.8	0.65	0.39	0.59	0.5	0.32

7 Discussion of the tsunami detection results

In general, the number of events detected for each configuration is quite low. The only events detected are E1, E2, E3, E4, E10, and E14, while all other events are missed with every method and parameter combination. The Andreanov 1996 tsunami (E1) is the one with the highest number of detections, followed by the Kuril Island 2006 tsunami (E3).

We point out, however, that all the events considered in our database of the Adak Island tide-gauge records are modest with no large amplitudes (see Table 3): in particular, the events E6, E11, E13, and E17 do not meet even the first condition for detection (see Eq. 2a), since for them $|IS(t)|$ never exceeds the assumed threshold $\lambda_{IS} = 1 \text{ cm min}^{-1}$. This can be seen in Table 7, where, for each event one can find the

maximum values that are taken by the time functions $|IS(t)|$ and $BS3(t)$, as well as by the average of $BS3(t)$, separately within a 20-min long window just preceding the event and within the Detection Window (DW). Notice further that the events E15, E16, and E17 are too small to be distinguished from the background noise (see Table 3), and missing detection is not a serious problem.

As regards the six events detected, four (E1, E2, E3, E4) are characterised by the largest wave amplitude and height, while two (E10, E14) are quite small. The events E1, E2, E3, and E4 are detected by A3C5 and A3C7 all within the first 40 min, with the exception of the event E2 (see Table 6). The event E2 is characterised by very long periods (about 40 min) and by increasing amplitudes, which makes the tsunami detection problematic. The detection results for this event are

indeed quite bad, with Delay Times $DT = 159$ and 160 min, and as a consequence a very low TSP ($TSP = 4\%$), with the exclusion of the lowest thresholds for the configuration A3C7, which gives a reasonable detection ($DT = 28, 29$ min), but a low value of TSP ($TSP = 9.6\%$).

The detection of event E10 is an example of how TEDA may detect small events, since the E10 situation meets the hypothesis at the base of the algorithm, that is this event arrives with a quite high first wave on low background. As for event E14, that event is also quite small, noting that its detection is exceptional and fortuitous, since the event occurred in a very low background condition and with the first wave just passing the assumed threshold of λ_{IS} : $|IS| = 1.09 \geq 1$ (cm min^{-1}).

As for the other cases, i.e. E5, E7, E8, E9, E12 (and E14), some hypotheses about the missed detections can be made. The Peru 2001 (E5) and the Sumatra 2004 (E8) events are far-field tsunamis and reach the tide gauge with a train of waves of slowly increasing amplitude. TEDA tsunami detection, however, is conceived to work with impulsive signals. Indeed, a slowly increasing signal, which is also typical of phenomena of atmospheric origin like seiches, is not supposed to trigger a detection because of the simultaneous increasing of the slope functions IS and BS. In addition, the E8 tsunami arrives when the function BS is too high relative to the size of the tsunami (see Table 7 and compare $\max(BS3) = 0.75 \text{ cm min}^{-1}$ against $\max(|IS|) = 1.19 \text{ cm min}^{-1}$). A possible reason why E7 is missed by all configurations can be the presence of a big oscillation, well visible in the record (see Fig. 5) right before the starting of the tsunami event. This increases the function BS3 and at the same time keeps CF at about the same level, hence masking the following tsunami waves. As for the missed detections of the events E9 and E12, it happens that the first incoming wave for both events is too low to pass the assumed threshold λ_{IS} . When the first wave is not recognised, it is incorporated in the background by TEDA: therefore the function BS increases, which as a consequence prevents the function CF to rise. We also note, however, that these two events are quite small, as they both reach a maximum value of $|IS|$ about 1.3 cm min^{-1} (see Table 7).

The above considerations and the analysis of the performance indicators allows us to judge which ones of the 21 explored configurations are preferred for the application of TEDA to the Adak Island tide gauge station. The choice is restricted between configurations A3C5 and A3C7, yet we stress that the analysis of the number of detections, of the Delay Time (DT), of the Tsunami State Duration (TSP), and of the Detection Tsunami Range (DTR(5)) has not provided us unambiguous elements to make a decision. In the next section we will add a further element for the evaluation.

8 Analysis of the background signal and implication on TEDA performance

In Sect. 4 the background signal has been analysed to identify the main spectral components, finding that spectral densities are quite stable, exhibiting a series of peaks (the dominant being around 13–14 min) that persist even in the case of a tsunami. In this section we examine the background signal in view of establishing if the site conditions are favourable to the application of TEDA, and also in the attempt to distinguish between the performance of the configurations A3C5 and A3C7.

Since we expect that sea surface oscillations, once detided, depend on weather conditions, we also expect that they show a seasonal variability that predominates inter-annual changes, that is predominates changes from one year to the other. Under this assumption, since analysing one year or another is equivalent, we explored the Adak Island tide gauge dataset at a 1 min sampling rate to find a concatenated sequence of one year of data in the period between July 2005 and June 2010. Since this was not possible due to quite frequent intervals of missing data, we have built a virtual year of data on a day-of-year basis, grouping together the available days from 1 to 365 of different years. For example, for the first day of the virtual year (1 January) only data from years 2009 and 2010 could be used, since they have a continuous sea level record, while the remaining years have too many gaps in data. Likewise for the second day (2 January) only data from years 2006, 2008, 2009 and 2010 have been used, and so on. Indeed, the main problem derives from frequent gaps in data. If gaps are too long or too frequent in one day, that day is considered inadequate to build the virtual sample. In the other cases, gaps of missing data have been filled by interpolation according to a two-step procedure. First, only gaps not longer than 12 min and preceded and followed by at least 4 min of data have been interpolated. Then in the second round, the linear interpolation regarded longer gaps, but not longer than 18 min and provided that they were surrounded by at least 90 min of data before and after. Considering that the predominant period of the Adak Islands tide gauge records is about 13–14 min (see Fig. 4), the linear interpolation can involve an entire wave period. With the above procedure we have been able to build a full 365-day long virtual year of no-tsunami data, i.e. a set of 365 intervals 24-h long referring to consecutive days from the 1 January to the 31 December.

The background signal $BS(t)$ has been computed for the entire virtual-year data set at the time step of 1 min. Since in Sect. 6 configurations A3C5 and A3C7 have proven to perform better than the others, the corresponding settings have been used to compute $BS(t)$, or more precisely $BS3(t)$. The two yearly normalised frequency distributions are shown in Fig. 9a. The computed values of BS3 have been grouped in semi-open bins (that include the lower endpoint and exclude the upper one), with width of 0.1 cm min^{-1} . For each of such

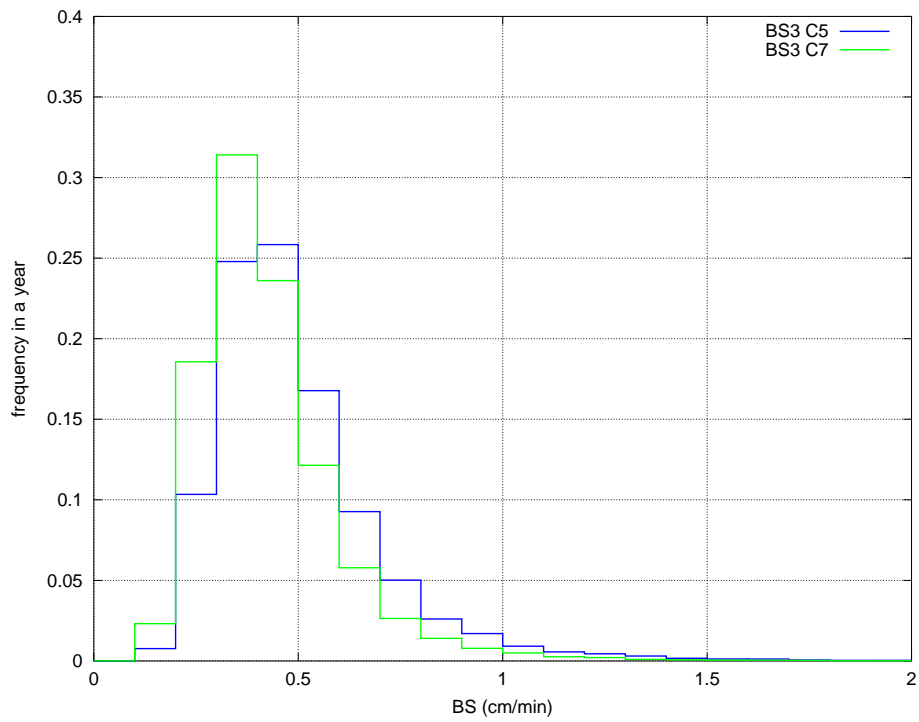


Fig. 9a. Normalised frequency distribution of BS3 for the configurations A3C5 and A3C7. Notice that most of the BS3 values are below 1 cm min^{-1} .

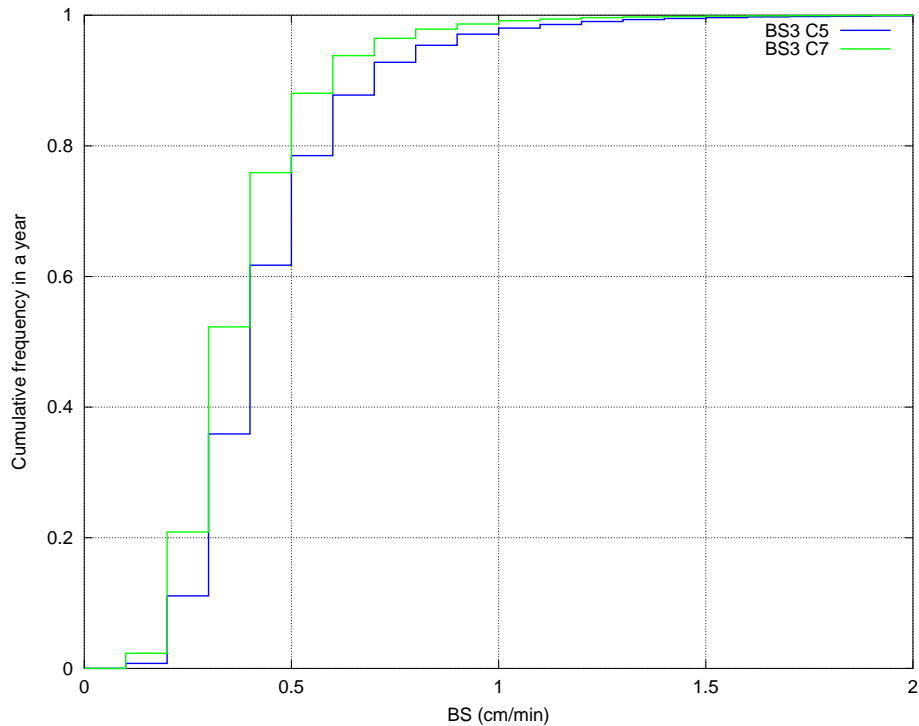


Fig. 9b. Cumulative distribution of BS3 for the configurations A3C5 and A3C7.

bins, all BS3 values included in the interval are counted with a weight of $1/(365n)$, with n being equal to the number of years available for the given day, in order to get a total weight of $1/365$ for every day of the year. In the following, each bin will be referred to by means of its central point. The corresponding cumulative distributions are displayed in Fig. 9b. The two resulting distributions have similar shape, but with some clear difference. Frequencies of the A3C5 curve compared to the ones of the A3C7 curve tend to be smaller for lower values of BS3 and higher for larger values, the discriminant being 0.40 cm min^{-1} . They are both unimodal and asymmetrical with mode in the bin 0.35 cm min^{-1} for the configuration A3C7 and in the bin 0.45 cm min^{-1} for A3C5. Furthermore, they are both positive skew with a long right tail, but the A3C7 distribution has a visibly higher skewness than the A3C5 distribution. From the cumulative distributions we see that the 60th percentile falls between bins $0.35\text{--}0.45 \text{ cm min}^{-1}$; the 80th percentile falls between bins $0.45\text{--}0.55 \text{ cm min}^{-1}$ for A3C7 and between the following bins $0.55\text{--}0.65 \text{ cm min}^{-1}$ for A3C5. Similarly, the 90th percentile is shifted by one bin and passes from between $0.55\text{--}0.65 \text{ cm min}^{-1}$ for A3C7 to $0.65\text{--}0.75 \text{ cm min}^{-1}$ for A3C5.

What is relevant in terms of TEDA tsunami detection algorithm is the value of the threshold λ_{IS} , since when the function $|IS(t)|$ is below λ_{IS} , no tsunami detection can occur (see Eq. 2a). Since BS3 is computed on the basis of the $\max(|IS_{tsu}(t)|)$ within the interval I_{BS} (see Eq. 1c), then $BS3(t) < \lambda_{IS}$ implies that $|IS(t)| < \lambda_{IS}$. Because in the present application it is assumed that $\lambda_{IS} = 1 \text{ cm min}^{-1}$, it is relevant to observe that the value $BS3 = 1 \text{ cm min}^{-1}$ is around the 96th (98th) percentile of the A3C5 (A3C7) frequency distribution, which means that TEDA is expected to give no false detections in at least the 96% (98%) of the cases, or, in other words, in most of the days of the year. This consideration gives us a two-percent element in favour of configuration A3C7 over A3C5. For the rest of the days, a false detection may occur if the second criterion (Eq. 2b) is matched, but nothing can be decided on the basis of the BS frequency distribution alone. We recall, however, following the performance analysis on the TEDA tsunami detection algorithm carried out in Sect. 6, that we have set up the threshold λ_{CF} in such a way to get no false detections in the learning data set, which ensures that false detections are unlikely.

In addition to false detections, a further relevant issue is that of missed detections. So one question we can pose is how big a tsunami has to be in order to be detected, given the distribution of the background signal we have found. To answer, let us assume that the signal $m(t)$ of a marigram with a tsunami can be functionally seen as the superposition of the background $b(t)$ and of the tsunami signal $f(t)$, so that the total signal $m(t)$ is equal to the sum $m(t) = b(t) + f(t)$. With this assumption, the sea level slope $m'(t)$ is given by $m'(t) = b'(t) + f'(t)$. In terms of TEDA functions, taking into account the definition (1c), we can consider $|IS(t)|$ as an estimation of $|m'(t)|$, and $BS3(t)$ as an estimation of $|b'(t)|$.

Analogously we can introduce $|IS_{tsu}(t)|$ as the estimate of the tsunami slope $|f'(t)|$. It is then possible to maximise $|IS(t)|$ in the following way:

$$|IS(t)| \leq |IS_{tsu}(t)| + BS3(t) \leq \max(|IS_{tsu}(t)|) + BS3(t) \quad (9)$$

In view of the conditions (2a) and (2b), a tsunami cannot be detected if either $|IS| < \lambda_{IS}$ or $|IS| < \lambda_{CF}BS3$, which implies that it is even more so if either

$$\max(|IS_{tsu}(t)|) + BS3 < \lambda_{IS} \quad (10a)$$

or

$$\max(|IS_{tsu}(t)|) + BS3 < \lambda_{CF}BS3, \quad (10b)$$

The above inequalities can be transformed into no-detection conditions in terms of BS3:

$$BS3 < \lambda_{IS} - \max(|IS_{tsu}(t)|) = IS1 \quad (11a)$$

$$BS3 > \max(|IS_{tsu}(t)|) / (\lambda_{CF} - 1) = IS2 \quad (11b)$$

where IS1 and IS2 are implicitly defined and the second inequality holds if $\lambda_{CF} > 1$. Let us try an interpretation of such inequalities. On the BS3 axis, they define two intervals of BS3, where a tsunami of a given maximum value $\max(|IS_{tsu}(t)|)$ is undetected. More precisely, when the conditions (11a) and (11b) are true, then the tsunami is missed, but it can also be missed even if the conditions are not met (for instance, if detection does not occur within the Detection Window DW). Therefore, they define the two minimum intervals of no-detection. Consider that for those values of $\max(|IS_{tsu}(t)|)$ for which $IS1 < IS2$, the two intervals do not overlap. On the other hand, when $IS1 \geq IS2$, they overlap and consequently any value of BS3 satisfies the system of inequalities (11) and the tsunami is therefore undetected. The frequency distribution of undetected tsunamis can be derived from the cumulative frequency distribution of BS3 given in Fig. 9b and expressed as a function of $\max(|IS_{tsu}(t)|)$. In Fig. 10, this type of frequency distribution is displayed for configurations A3C5 and A3C7 (blue and green curves respectively), together with the corresponding complementary distributions (red and orange curves). To be more precise, in Fig. 10 a number of distributions are plotted corresponding to all the values of the threshold λ_{CF} that are included in the intervals $DTR_{A3C5}(5)$ and $DTR_{A3C7}(5)$. If we interpret normalised frequency as probabilities, these curves provide a lower limit to the probability that a tsunami of given $\max(|IS_{tsu}(t)|)$ is missed, passing as undetected, while the complementary curves provide an upper limit to the probability that such a tsunami will be detected. On comparing such curves, it is clear that configuration A3C7 is more adequate than A3C5 for all the used values of λ_{CF} , since it guarantees a smaller lower limit for no-detection probabilities and a larger upper limit for detection probabilities over the whole range of $\max(|IS_{tsu}(t)|)$.

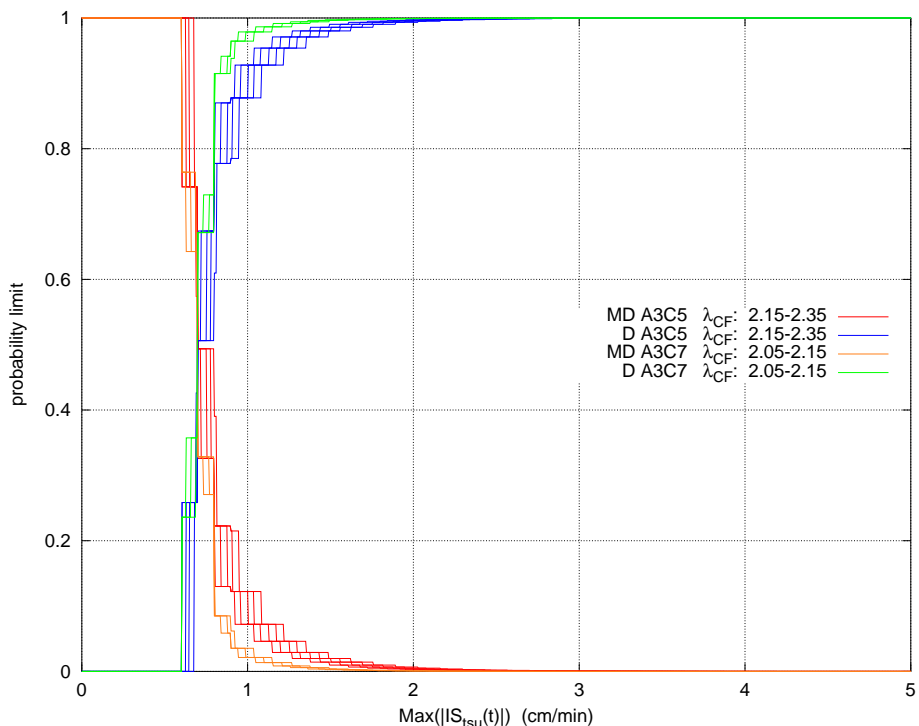


Fig. 10. Lower limit of probability for missed tsunami detection (red and orange) and upper limit of probability for tsunami detection (blue and green) derived from cumulative distributions of BS for both configurations A3C5 and A3C7 and computed for λ_{CF} values taken from the intervals $DTR_{A3C5}(5)$ and $DTR_{A3C7}(5)$.

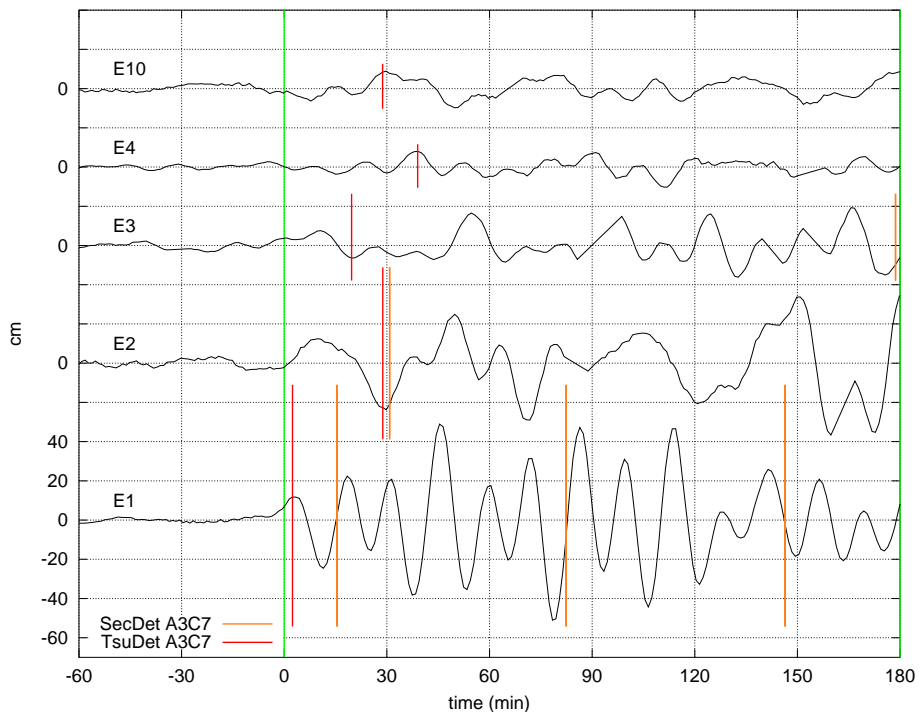


Fig. 11. TEDA detections shown on the detided records of the E1, E2, E3, E4 and E10 tsunamis, with the optimal configuration A3C7. Tsunami arrival time is at time $t = 0$. TEDA detection time (vertical lines) are shown both for TEDA tsunami detection and for TEDA secure detection. Method A3C7 corresponds to $t_{IS} = 12$ min, $t_G = 16$ min, $t_{GTide} = 17$ min, $t_{BS} = t_{Tide} = 60$ min, and CF threshold equalling 2.05.

Finally, we remark that tsunami size is given here through $\max(|IS_{tsu}(t)|)$, which is the maximum slope expressed in absolute value of the tsunami waves on the tide gauge records. It is possible to make an estimate of the corresponding tsunami height by considering the typical tsunami spectral density portrayed in Fig. 4, where one can see a number of resonant peaks. Since the predominant peak is located around the period $T_{peak} = 13\text{--}14$ min, one can translate $\max(|IS_{tsu}(t)|)$ into wave height H by simply multiplying it by the factor T_{peak}/π , that is $H = \max(|IS_{tsu}(t)|) T_{peak}/\pi$ (see McGehee and McKinney, 1997).

9 The best TEDA setting for the tsunami detection algorithm

From the analyses performed in the previous sections, the best choice for TEDA appears to be the settings for configuration A3C7 and, considering that $DTR_{A3C7}(5) = [2.05, 2.15]$, the best choice for the threshold λ_{CF} turns out to be $\lambda_{CF} = 2.05$, which is the minimum threshold to avoid false detections.

With this parameter setting, TEDA detects the Andreanov 1996 (E1), the Chile 2010 (E2), the Kuril Island 2006 (E3), the Kasatochi 2008 (E4), and the Vanuatu 2009 (E10) tsunamis, respectively, with Delay Time DT of 2 min, 28 min, 19 min, 39 min, and 28 min from the estimated arrival time. The corresponding Tsunami State Duration (TSP) percentages are respectively of about 100% for E1, 10% for E2, 99% for E3, 94% for E4 and 59% for E10 event (see Table 6). Delay Times for the selected setting of A3C7 are shown in Fig. 11. Three of the well detected events are detected at the first tsunami wave. The Andreanov 1996 (E1) tsunami is detected at the first maximum of the first wave (that is a positive one), the Chile 2010 (E2) and the Kuril Island 2006 (E3) tsunamis are detected at the first minimum (which is the trough of the first wave). In contrast, the Kasatochi 2008 (E4) and the Vanuatu 2009 (E10) tsunamis are detected at the highest maximum of the leading train of waves.

10 Results of the tsunami secure detection algorithm

The TEDA secure detection algorithm has been tested with different values of parameters, in analogy with the tsunami detection algorithm. More specifically, the duration t_{SD} of the integration interval I_{SD} (see Sect. 2) has been varied in the range from 1–60 min and applied to the subset of events composed of E1, E3, E5, E7, and E14 by using all combinations of Table 4 in order to select the most suitable value. The criterion used for the selection has been the maximisation of the integral function $M(t)$ given by the Eq. (3a), since this ensures a higher sensitivity of the algorithm. The value of t_{SD} which provides the highest values turns out to be the same for all combinations and is $t_{SD} = 8$ min. With this value of t_{SD} ,

Table 8. TEDA secure detection results. All parameter combinations detect the same three events E1, E2 and E3, which are the ones with the largest amplitudes. The Delay Time DT, measured from tsunami arrival, seems to increase along with t_{IS} and to increase as tsunami size decreases (see Table 3 to compare amplitudes, heights, and ranges of the three events).

Cn	t_{IS} (min)	t_G (min)	DT: E1 (min)	DT: E2 (min)	DT: E3 (min)
1	6	15	12	28	56
2	6	10	12	28	56
3	8	15	13	29	134
4	8	10	13	29	134
5	10	15	14	30	136
6	10	10	14	30	136
7	12	15	16, 83, 147	31	179

the algorithm has been applied for all the combinations to all the Adak Islands records included in the learning database (see Table 2), including tsunami events and a total of about 123 days of background. The reason is that the algorithm is supposed to reveal an event with amplitude high enough to be potentially dangerous, no matters whether it is a tsunami or not. The result is that no detections occurred during the background days, and the only detections were seen in correspondence to tsunami events. We recall here that in TEDA terminology, a tsunami is the part of the record that is contained in the Tsunami Interval (TI) according to the definition given in Sect. 5. For all combinations Cn, TEDA detection is triggered only by the events E1, E2, and E3, which are the ones with the highest amplitude: the maximum excursion of the detided sea level oscillations given in Table 3 are larger than 25 cm. For all the remaining events, the threshold λ_{SD} is not passed by the absolute value of the function $M(t)$.

Since the secure detection algorithm is also triggered by tsunamis, we can use, even in this case, the Delay Time DT, which is one of the event indicators introduced for the tsunami detection algorithm. The values of such DTs for the events E1, E2, and E3 are given in Table 8 for all combinations used. We notice that DT is higher for higher values of the parameter t_{IS} and increases as the tsunami amplitude decreases. Further, in Fig. 8 the DTs resulting from the application of the TEDA tsunami detection algorithm are compared with the ones resulting from the TEDA secure detection for both configurations A3C5 and A3C7. It appears that in most instances, the DTs of the secure detection are larger than the corresponding DT of the tsunami detection, the only exception being the detection of E2 by A3C5.

11 Conclusions

A performance evaluation and optimal parameters setting of TEDA has been carried out for the Adak Island tide gauge by using seventeen event records. Three methods of computing the BS functions have been tested using different values of the temporal parameters t_{IS} , t_G , and t_{BS} , with CF threshold varying from 1 to 5 at steps of 0.05. A strong criterion to select the optimal setting has been that of avoiding false detections, which has influence on the minimum value of CF threshold admissible. Other criteria used have been the maximisation of the number of events detected, the minimisation of the detection delay time, and the maximisation of the tsunami state interval coverage. These criteria altogether show that the best configurations for the Adak Island station are A3C5 and A3C7, but cannot lead us further since results are quite similar. A powerful means to discriminate between the two has been the analysis of the background signal that was carried out using a virtual year database, showing that the frequency distribution of BS and of IS_{tsu} for A3C7 is more adequate, being less prone to false detections and also providing better probability limits for missed and true detections.

The threshold range taken into account for the configuration A3C7 is $DTR_{A3C7}(5) = [2.05, 2.15]$. The background analysis confirms that a lower threshold provides slightly better results. Setting the threshold at 2.05 is for this analysis the best choice, but it is important to stress that this setting could be further refined through the analysis of additional data and through a further performance evaluation. This setting allows the detection of the major tsunamis of the analysed dataset, yet at the same time minimises the number of false detections. The condition of avoiding false detections brings as a direct consequence the impossibility of detecting small events, which is an acceptable drawback in the optics of Tsunami Warning Systems, whose main focus is the timely detection of dangerous tsunamis. The test of our TEDA algorithm showed that a big tsunami event would be detected fast (small DT), and that the tsunami state is adequate (large TSP). However, in some cases, tsunamis might exhibit slowly increasing amplitudes and their arrival at the coastline might not fit the hypothesis made of a sudden change in the sea-level. But in such cases, the secure detection criterion assures that a warning is going to be issued for high amplitude dangerous waves.

It is important to stress that these results are valid only for the Adak sea level station, and that the calibration of TEDA should be performed case by case, since its performance might vary according to the local conditions.

Acknowledgements. The development and testing of TEDA has been undertaken within the framework of the 3-year project TRANSFER (2006–2009), funded by the European Commission and within the framework of the Italian Project DPC-INGV S3. Testing of the algorithm has been possible thanks to a collabora-

tion with NCTR of PMEL/NOAA, which made available original Adak Island tide-gauge data at a 1 min sampling rate.

The authors are indebted to the anonymous referees who evaluated the paper for their valuable suggestions.

Edited by: I. Didenkulova

Reviewed by: two anonymous referees

References

- Allen, R. V.: Automatic earthquake recognition and timing from single traces, *B. Seismol. Soc. Am.*, 68, 1521–1532, 1978.
- Bellotti, G., Di Risio, M., and De Girolamo, P.: Feasibility of Tsunami Early Warning Systems for small volcanic islands, *Nat. Hazards Earth Syst. Sci.*, 9, 1911–1919, doi:10.5194/nhess-9-1911-2009, 2009.
- Beltrami, G. M.: An ANN algorithm for automatic, real-time tsunami detection in deep-sea level measurements, *Ocean Eng.*, 35, 572–587, 2008.
- Earle, P. S. and Shearer, P. M.: Characterization of global seismograms using an automatic-picking algorithm, *B. Seismol. Soc. Am.*, 84, 366–376, 1994.
- Eble, M. C., Newman, J., Wendland, J., Kilonsky, B., Luther, D., Tanioka, Y., Okada, M., and Gonzalez, F. I.: The 10 June 1996 Andreanov Tsunami Database, NOAA Data Report ERL PMEL, 1997.
- Historical Tsunami Database at NGDC Tsunami Database, http://www.ngdc.noaa.gov/hazard/tsu_db.shtml, 2010.
- Honda, K., Terada, T., Yoshida, Y., and Isitani, D.: An investigation on the secondary undulations of oceanic tides, *J. College Sci., Imper Univ., Tokyo*, 1908.
- Illigner, J. and Schöne, T.: TRANSFER Project: GFZ software, Software for Automatic Detection, Analysis and Quality Control of high-frequency and low-latency sea level signals (Deliverable 4.3.4), 22–29, 2009.
- Kulikov, E. A.: Sea level measurements and tsunami forecasting, *Sov. Meteorol. Hydrol.*, 6, 61–68, 1990.
- McGehee, D. D. and McKinney, J. P.: Tsunami detection and warning capability using nearshore submerged pressure transducers – Case study of the 4 October 1994 Shikotan tsunami, in: *Perspectives on Tsunami Hazard Reduction*, edited by: Hebenstreit, G., Kluwer, Norwell, MA, 133–143, 1997.
- Mero, T. N.: NOAA/National Ocean Service Application of Real-Time Water Levels, in *Proceedings, Volume 2, Ocean Community Conference, The Marine Technology Society Annual Conference*, 16–19 November 1998, 1036–1039, 1998.
- Miller, G. R.: Relative spectra of tsunamis, *Hawaii Inst Geophys HIG-72-8*, 1972.
- Miller, G. R., Munk, W. H., and Snodgrass, F. E.: Long-period waves over California's continental borderland, II, *Tsunamis*, *J. Mar. Res.*, 20, 31–41, 1962.
- Mofjeld, H. O.: Tsunami detection algorithm, not published paper, http://nctr.pmel.noaa.gov/tda_documentation.html, 1997.
- Rabinovich, A. B.: Spectral analysis of tsunami waves: separation of source and topography effects., *J. Geophys. Res.*, 102(C6), 12663–12676, 1997.
- Rabinovich, A. B. and Stephenson, F. E.: Longwave Measurements for the Coast of British Columbia and Improvements to the

- Tsunami Warning Capability, *Nat. Hazards*, 32, 313–343, 2004.
- Rabinovich, A. B., Thomson, R. E., and Stephenson, F. E.: The Sumatra tsunami of 26 December 2004 as observed in the North Pacific and North Atlantic oceans, *Surv. Geophys.*, 27, 647–677, 2006.
- Sanchez, A. J. and Farreras, S. F.: Maximum entropy spectral analysis of tsunamis along the Mexican coast, 1957–1979, in: *Tsunamis: their science and engineering*, edited by: Iida, K. and Iwasaki, T., Terra Sci, Tokyo, 147–159, 1983.
- Tolkova, E.: Principal component analysis of tsunami buoy record: Tide prediction and removal, *Dynam. Atmos. Oceans*, 46, 62–82, 2009.
- Tolkova, E.: EOF analysis of a time series with application to tsunami detection, *Dynam. Atmos. Oceans*, 50-1, 35–54, 2010.
- Van Dorn, W. G.: Some tsunami characteristics deducible from tide records, *J. Phys. Oceanogr.*, 14, 353–363, 1984.
- Vela, J. and Pérez, B.: TRANSFER Project: Puertos del Estado (OPPE) software, Software for Automatic Detection, Analysis and Quality Control of high-frequency and low-latency sea level signals (Deliverable 4.3.4), 4–8, 2009.
- Wang, X. and Liu, P. L. F.: Numerical Simulation of Nov 15 2006 Central Kuril Islands Tsunami, School of Civil and Environmental Engineering, Cornell University, http://ceeserver.cee.cornell.edu/pll-group/doc/2006_%20Kuril_tsunami.pdf, unpublished, 2006.
- WCATWC Tsunami Catalog, <http://wcatwc.arh.noaa.gov/about/tsunamimain.php>, 2010.
- Wijeratne, E. M. S. and Woodworth, P. L.: TRANSFER Project: NERC (POL) software, Software for Automatic Detection, Analysis and Quality Control of high-frequency and low-latency sea level signals (Deliverable 4.3.4), 8–14, 2009.