

# Tsunami waveform inversion by numerical finite-elements Green's functions

A. Piatanesi, S. Tinti, and G. Pagnoni

Dipartimento di Fisica, Università di Bologna, Bologna, Italy

Received: 06 August 2001 – Accepted: 16 November 2001

**Abstract.** During the last few years, the steady increase in the quantity and quality of the data concerning tsunamis has led to an increasing interest in the inversion problem for tsunami data. This work addresses the usually ill-posed problem of the hydrodynamical inversion of tsunami tide-gage records to infer the initial sea perturbation. We use an inversion method for which the data space consists of a given number of waveforms and the model parameter space is represented by the values of the initial water elevation field at a given number of points. The forward model, i.e. the calculation of the synthetic tide-gage records from an initial water elevation field, is based on the linear shallow water equations and is simply solved by applying the appropriate Green's functions to the known initial state. The inversion of tide-gage records to determine the initial state results in the least square inversion of a rectangular system of linear equations. When the inversions are unconstrained, we found that in order to attain good results, the dimension of the data space has to be much larger than that of the model space parameter. We also show that a large number of waveforms is not sufficient to ensure a good inversion if the corresponding stations do not have a good azimuthal coverage with respect to source directivity. To improve the inversions we use the available *a priori* information on the source, generally coming from the inversion of seismological data. In this paper we show how to implement very common information about a tsunamigenic seismic source, i.e. the earthquake source region, as a set of spatial constraints. The results are very satisfactory, since even a rough localisation of the source enables us to invert correctly the initial elevation field.

## 1 Introduction

In the last few years, a great effort has been made to improve the quantity and the quality of the collected data concerning tsunamis. An example is the Deep-ocean Assessment and

Reporting of Tsunami (DART) system for tsunami detection and forecasting, build-up by the Pacific Marine Environmental Laboratory/NOAA (USA) (Titov et al., 1999). Today, any relevant tsunami occurring in the Pacific area is recorded at several tide-gage stations distributed along the coastline of many countries facing the Pacific Ocean. Just to give an example we can mention the last tsunami generated near the coast of Peru on 23 June 2001, for which some tens of good tide-gage and ocean bottom pressure gage records have been made available to the scientific community with a time lag of only one day. For those tsunamis that cause deaths and large coastal inundation, in addition to these “far-field” data, local data are collected: actually, it is now a standard practice to organize a post-event field survey with the purpose of measuring run-up heights along the most affected segments of coast, in order to obtain a detailed picture of the inundation scenario.

Due to this steady increase in the quantity and quality of the available data, the interest in the inversion problem for tsunami data has increased as well. Satake (1987, 1989) shows, for the first time, that the inversion of tide-gage records can be used to retrieve some information about the tsunamigenic source mechanism. In Satake's approach, it is necessary to preliminarily assume a fault plane solution, usually on the basis of seismological data. Then, after segmenting the fault plane into several sub-faults, this method allows for the estimation of the slip distribution along the fault. A very similar approach to retrieve the slip distribution along the tsunamigenic fault is proposed by Piatanesi et al. (1996) for the case of the 1992 Nicaragua tsunami earthquake. In that paper they show that in the case where tide-gage records are not available or are not in a sufficient quantity and/or quality to perform an inversion, the local run-up heights, collected during the post-event field surveys, can be used for this purpose. It is now quite a common practice to determine the slip distribution along the fault of tsunamigenic earthquakes on the basis of tsunami data, eventually using joint inversion of tsunami and geodetic data (Johnson et al., 1996; Johnson, 1999).

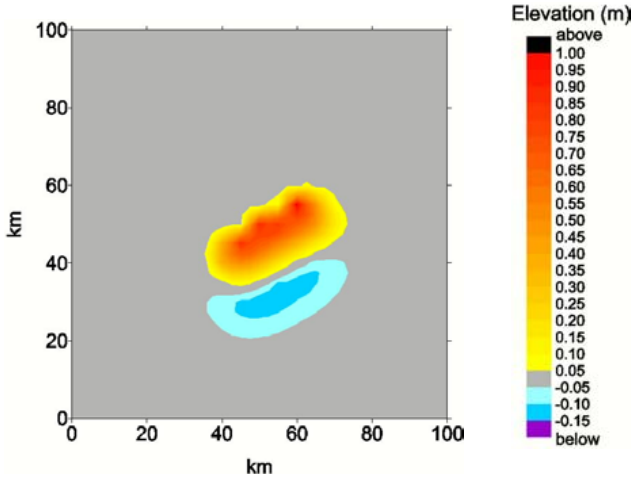


Fig. 1. Water elevation field representing the initial condition  $\zeta(t_0)$ .

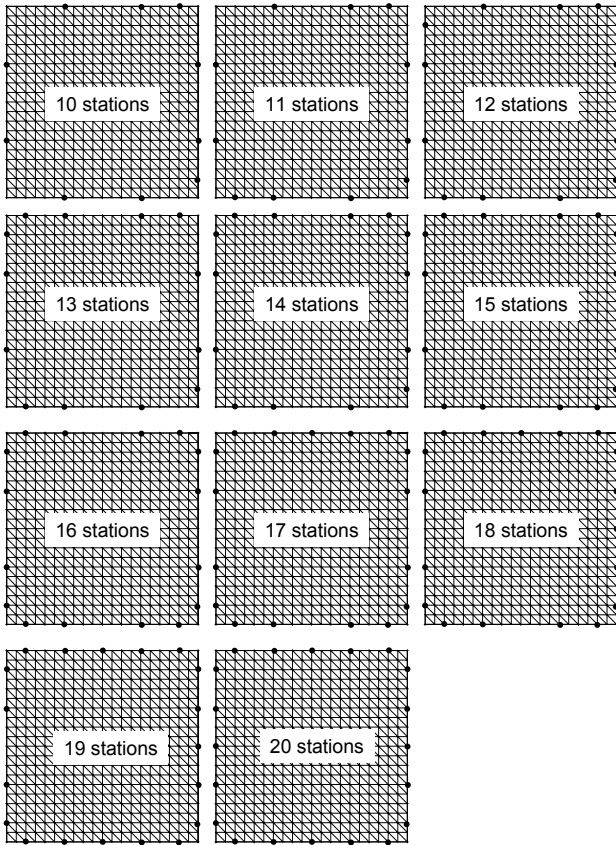


Fig. 2. Station distributions relative to the inversion experiments where various number of waveforms are used.

Very recently, Pires and Miranda (2001) proposed an adjoint method for tsunami waveform inversion, as an alternative to the technique based on Green's functions of the linear long wave model (Satake, 1987). In the present paper we use an inversion method, already described in its fundamentals in a previous paper (Tinti et al., 1996), for which the data space consists of a given number of tide-gage records and

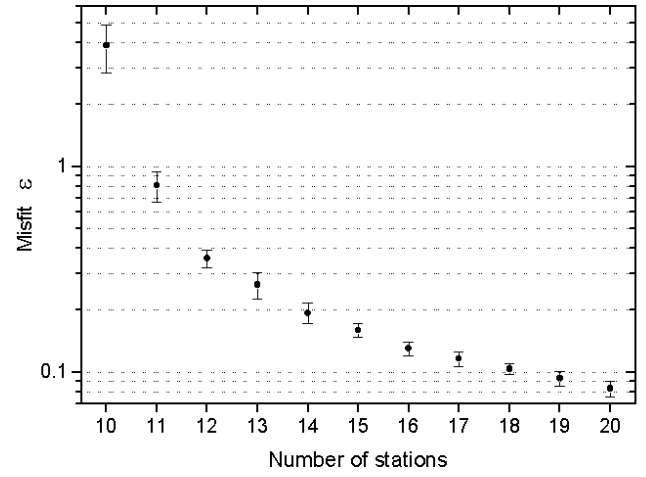


Fig. 3. Dependence of the misfit on the number of waveforms considered in the inversions. The corresponding station distributions are sketched in Fig. 2.

the model parameter space is represented by the values of the initial water elevation field at a given number of points. One of the main advantages of this method is that it does not require an *a priori* assumption of a fault plane solution: actually, as we will see in the following section, this method is completely independent of any particular source model.

## 2 Inversion method

The forward model for the tsunami propagation, i.e. for the calculation of the synthetic tide-gage records starting from an initial water elevation field, is based on the linear shallow water equations:

$$\begin{aligned} \partial_t \zeta &= -\nabla \cdot (h\mathbf{v}) \\ \partial_t \mathbf{v} &= -g\nabla \zeta, \end{aligned} \quad (1)$$

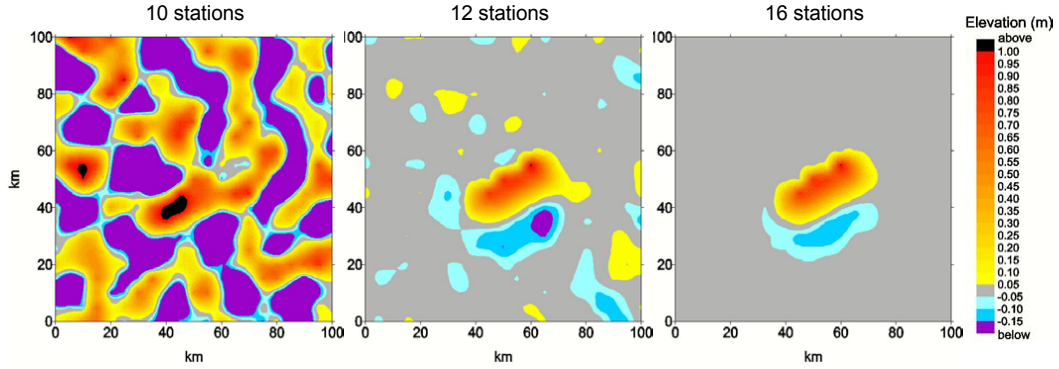
completed by the following boundary conditions:

$$\mathbf{v} \cdot \mathbf{n} = \frac{g}{c} \zeta \quad \text{on the open boundary} \quad (2a)$$

$$\mathbf{v} \cdot \mathbf{n} = 0 \quad \text{on the solid boundary.} \quad (2b)$$

In the above equations  $\zeta$  is the water elevation above the mean sea level,  $\mathbf{v} = (u, v)$  is the horizontal fluid velocity vector whose  $x$ - and  $y$ -components are, respectively,  $u$  and  $v$ ,  $h$  is the basin depth,  $g$  is the gravity acceleration,  $c = (gh)^{1/2}$  is the wave phase speed and  $\mathbf{n}$  is the unit vector, outwardly directed, normal to the boundary. To solve Eqs. (1) and (2) we use a finite-element technique: as shown in Tinti and Piatanesi (1995), finite-element spatial discretisation transforms Eqs. (1) and (2) into a linear set of ordinary differential equations that are first order in time and that can be put in the following compact form:

$$\frac{d}{dt} \xi(t) = \mathbf{A} \xi(t). \quad (3)$$



**Fig. 4.** Initial fields inverted using 10, 12 and 16 waveforms.

**Table 1.** Fault parameters of the seismic source used to compute the initial condition

$x_0$ km	$y_0$ km	Depth km	Length km	Width km	Slip m	Strike deg	Dip deg	Rake deg
50	50	1	30	15	2	60	30	70

Here,  $\xi = (\zeta, u, v)$  is a  $3N$ -components vector representing the value of the unknown fields on the nodes of the finite-element grid consisting of  $N$  nodes and  $\mathbf{A}$  is a matrix of constant coefficients that also includes the boundary conditions (see Tinti and Piatanesi, 1995). The classical theory of the linear set of differential equations (e.g. Arnold, 1978) provides a formal solution for the unknown vector  $\xi(t)$  of Eq. (3) through a spectral decomposition of the matrix  $\mathbf{A}$ :

$$\xi(t_k) = \exp[\mathbf{A}(t_k - t_0)] \xi(t_0) = \mathbf{E} \exp[\Lambda(t_k - t_0)] \mathbf{E}^{-1} \xi(t_0), \quad (4)$$

where  $\xi(t_k)$  is the unknown vector computed at the time  $t_k$ ,  $\mathbf{E}$  and  $\mathbf{E}^{-1}$  are, respectively, the eigenvectors matrix and its inverse, whereas  $\Lambda$  is the diagonal eigenvalues matrix and  $\xi(t_0)$  is the initial condition. From Eq. (4) we can define the Green's function of the problem as:

$$G_{kj}^i = E_{in} \exp[\Lambda_{nm}(t_k - t_0)] E_{mj}^{-1}, \quad (5)$$

where the rule of summation on the repeated index is adopted. If we restrict the problem to the case of a static initial condition and to the calculation of the water elevation  $\zeta(t)$  solely, the forward problem can be written in terms of the Green's functions as:

$$\zeta_i(t_k) = G_{kj}^i \zeta_j(t_0). \quad (6)$$

In Eq. (6),  $\zeta_i(t_k)$  is the elevation computed at the  $i$ -th node at the time  $t_k$  and the Green's function  $G_{kj}^i$  has the usual interpretation as the elevation on the node  $i$  at the time  $t_k$ , produced by a unitary  $\delta$ -shape  $\zeta$  pulse applied on the node  $j$  at the time  $t_0$ .

As already stated, our inverse problem consists of providing a given number of tide-gage records as data input to obtain the initial water elevation field  $\zeta_j(t_0)$  as output. Let us

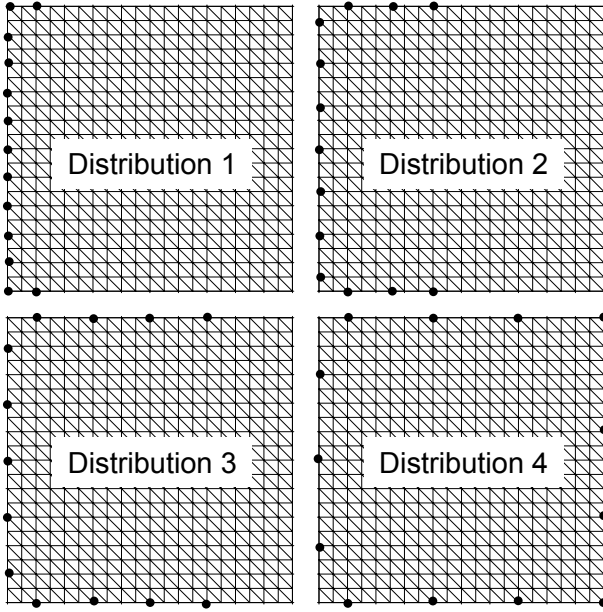
denote by  $N$  the number of grid nodes, by  $R$  the number of the available tide-gage records and by  $P$  the number of data points on each record: the system consisting of  $R \times P$  equations and the  $N$  unknowns  $\zeta_j(t_0)$  ( $j = 1, \dots, N$ ) that mathematically represents our inverse problem can be written in the following way:

$$\begin{aligned} \zeta_1(t_1) &= G_{1j}^1 \zeta_j(t_0) \\ &\vdots \\ \zeta_1(t_P) &= G_{Pj}^1 \zeta_j(t_0) \\ &\vdots \\ \zeta_N(t_1) &= G_{1j}^N \zeta_j(t_0) \\ &\vdots \\ \zeta_1(t_P) &= G_{Pj}^N \zeta_j(t_0) \end{aligned} \quad (7)$$

Since the number of rows is greater or equal to the number of columns ( $R \times P \geq N$ ), the solution  $\zeta_j(t_0)$  will be computed through a least square inversion of the system (7). In the following sections we will illustrate a series of numerical experiments performed on synthetic tide-gage records that will enable us to point out the main features of the proposed inversion method.

### 3 Inversion experiment set-up

The computational domain for all the numerical inversion experiments shown in this paper is a squared basin, with sides of 100 km of length along which open boundary conditions are applied, characterised by a flat bathymetry 200 m deep. This domain is spatially discretised by means of a mesh, consisting of 800 triangular elements and 441 nodes. On the basis of the theory exposed in the previous section and by means of Eq. (5), we compute the numerical Green's functions  $G_{kj}^i$ , where the index  $i$  (stations) spans over the nodes belonging to the boundary of the domain and  $j = 1, \dots, N$  (sources). A comment is necessary about the index  $k = 1, \dots, P$ , spanning the time along a synthetic record: if we denote by  $T_{end}$  the final computation time and



**Fig. 5.** Station distributions with different azimuthal coverage of the domain.

by  $\Delta t$  the sampling interval on a record,  $P$  can be obviously expressed as:

$$P = \frac{T_{end}}{\Delta t}. \quad (8)$$

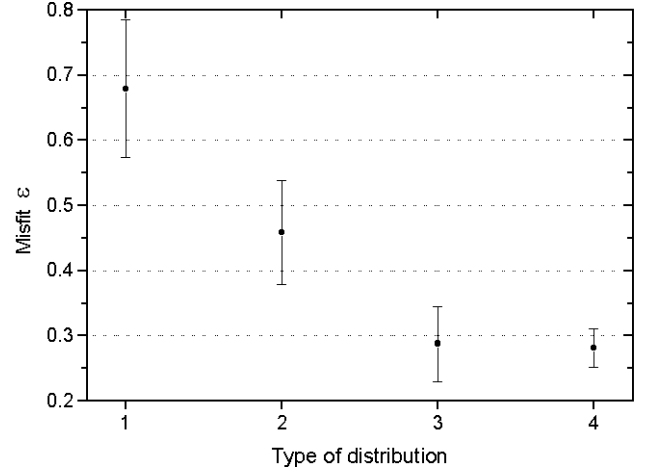
We choose a sampling interval  $\Delta t = 60$  s that represents a common value for most tide-gages installed to record tsunami waveforms; since in our simulations we consider  $T_{end} = 6000$  s, each synthetic record will have a number of points  $P = 100$ .

As the initial condition  $\zeta(t_0)$ , we consider the coseismic vertical displacement of the sea bottom produced by a seismic fault that we compute through Okada's analytical model (Okada, 1992): the fault parameters are listed in Table 1, while the corresponding water elevation field is shown in Fig. 1.

With the above initial condition and using the computed Green's functions, we build up the data space, consisting of the synthetic tide-gage records computed by means of Eq. (6) at virtual stations located on the nodes belonging to the boundary. To estimate the goodness of an inversion experiment we use an  $L_2$ -norm misfit parameter that represents the squared averaged difference between the theoretical initial field  $\zeta^{the}$  and the inverted one  $\zeta^{inv}$ :

$$\varepsilon = \left[ \frac{\sum_i (\zeta_i^{the} - \zeta_i^{inv})^2}{\sum_i (\zeta_i^{the})^2} \right]^{\frac{1}{2}} \quad i = 1, \dots, N. \quad (9)$$

Since real tsunami tide-gage records are always affected by errors, we take into account such uncertainty in the data by injecting a Gaussian random noise of 10% in magnitude into the synthetic waveforms. As we will see in the next section, where the results of the inversions experiments are presented,



**Fig. 6.** Dependence of the misfit on the type of stations distribution (sketched in Fig. 5).

this will give us the opportunity to estimate the robustness of the inversion with respect to fluctuations in the data, in terms of the variance reduction of the misfit parameter.

#### 4 Unconstrained inversion experiments

The first series of inversion experiments studies the dependence of the inverted initial field on the number of stations used in the inversion procedure: we perform the inversions using a number of records that ranges from 10 to 20. Figure 2 shows how the stations are distributed along the boundary, for each of the cases studied. In order to avoid the effects due to insufficient azimuthal coverage in this kind of experiment, the stations are uniformly distributed as much as possible along the four sides of the basin. Figure 3 shows the misfit parameter  $\varepsilon$  plotted as a function of the number of waveforms used in the inversion: what we found is that adding records in the inversion procedure leads to results that are better and better. This is particularly evident passing from 10 to 11 and from 11 to 12 stations, for which we have a misfit reduction by a factor of the order of 5 and 2, respectively. Then, as far as new records are added, the misfit continues to reduce, even if in a less relevant way. Actually, each value of the misfit, plotted in Fig. 3, is the mean of 10 numerical inversion experiments run with a different seed for the generator of the Gaussian random noise that perturbs the records. This enables us to associate to each value of the misfit, an error bar representing the standard deviation (root mean square) of the 10 values of the misfit. We found that adding stations in the inversion procedure leads not only to a reduction of the misfit, but also to a reduction of the standard deviation. This means that as long as new records are added in the data space, random fluctuations in the data become less and less important, i.e. the inversions become more and more robust. To give a better idea of the progressive inversion improvement as new waveforms are added, in Fig. 4 we show three initial

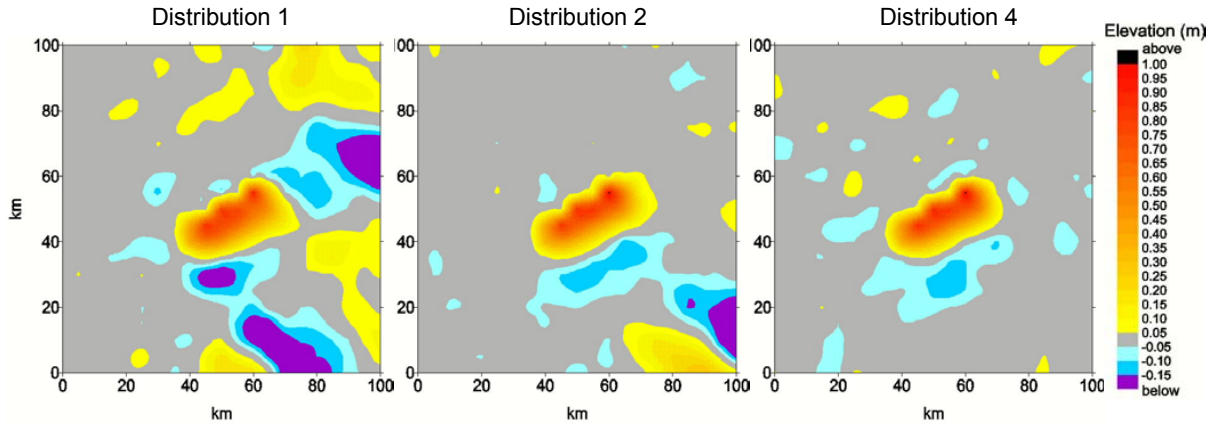


Fig. 7. Representative initial fields inverted using different stations distributions (shown in Fig. 5).

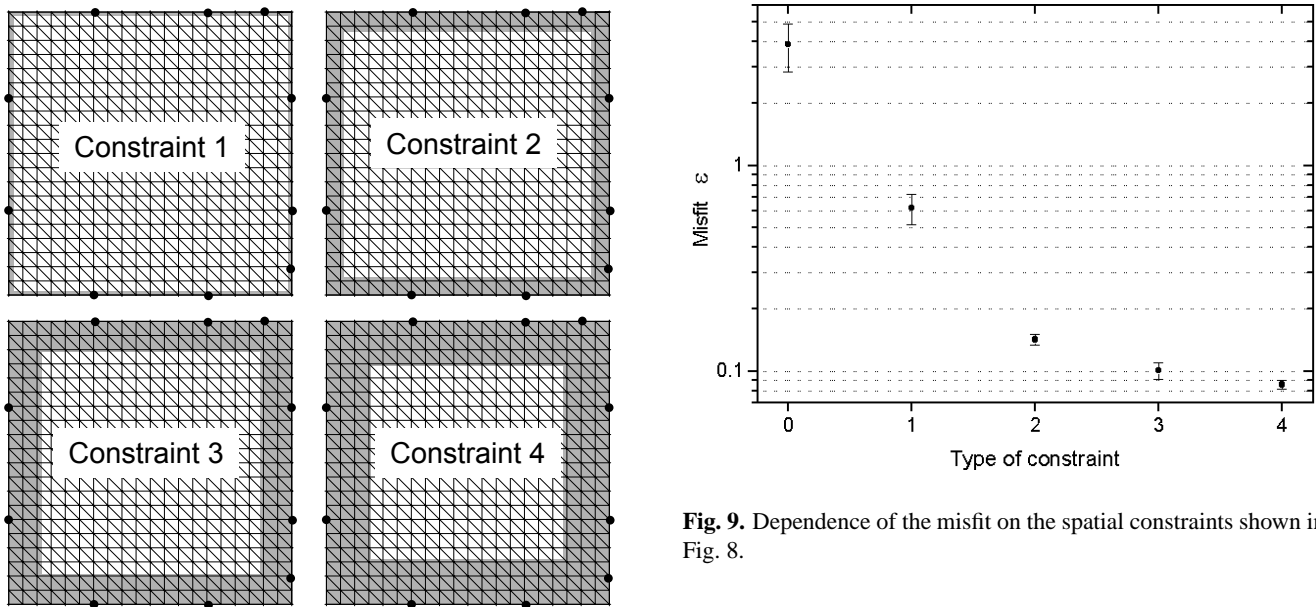


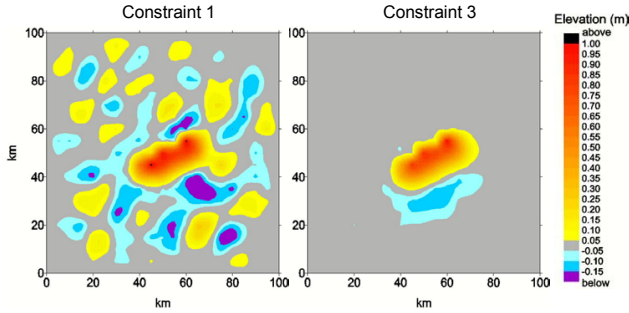
Fig. 8. Sketch of the spatial constraints used to improve the inversions with 10 waveforms: the initial field is taken equal to zero in the shaded zones.

fields inverted using a growing number of records. The inversion performed with 10 stations leads to an initial elevation field completely different from the theoretical one (see Fig. 1 for comparison): it is characterised by large and uncorrelated wave amplitudes that are indicative of an unstable inversion. The initial field inverted with 12 stations, corresponding to a misfit of 0.36, is visibly closer to the correct shape: the main positive-negative wave system, located in the middle of the basin, is reproduced in its main features, whereas several patches of noise are still present. Finally, when the inversion is performed with 16 waveforms, the inverted field is very close to the theoretical one: both the shape and the amplitude are well reproduced and the elevation field is noiseless.

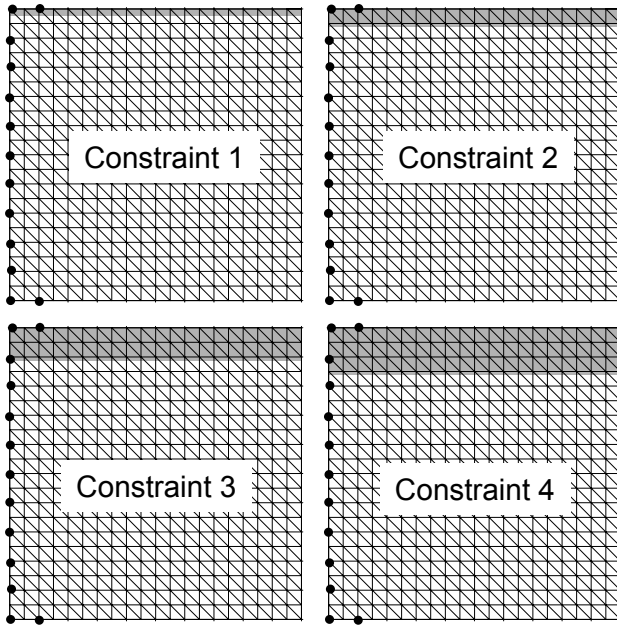
The number of records is only one of the parameters that

Fig. 9. Dependence of the misfit on the spatial constraints shown in Fig. 8.

controls the goodness of the inversion procedure: actually, the spatial distribution of the stations also plays an important role. To investigate this feature, we perform some inversion experiments using a fixed number of waveforms, whose stations are distributed in a different way. As shown in Fig. 5, we consider four types of station distributions, numbered from 1 to 4 and characterised by a growing azimuthal coverage of the basin. It is evident, from Fig. 6, that the larger the azimuthal coverage of the stations, the smaller the misfit and the more robust the inversion becomes. Anyway, the misfit gives only a global estimation of the goodness of the inversion: in this case it is of particular interest to analyze the shape of the inverted initial fields that are shown in Fig. 7. We can see that due to the upper/left-bottom/right directivity of the source, the lack of station coverage of distributions 1 and 2 generates large spurious wave amplitudes mainly in the bottom right corner of the basin. On the contrary, the upper/left part of the basin is more constrained as far as the source directivity is concerned, so that the noise threshold is



**Fig. 10.** Representative initial fields inverted by using 10 waveforms and applying the constraints shown in Fig. 8.

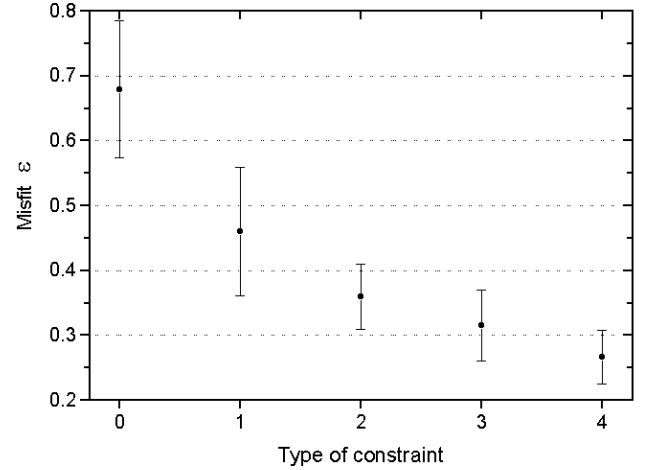


**Fig. 11.** Sketch of the spatial constraints used to improve the inversions with the station distribution of type 1, characterised by the lowest azimuthal coverage: the initial field is taken to be equal to zero in the shaded zones.

very low there. This feature does not appear in the case of distribution 4, characterised by an uniform azimuthal coverage.

## 5 A priori information: spatial constraints

In all the numerical experiments carried out until now, we have made the explicit assumption that the inversion is unconstrained, so that the degrees of freedom of the problem are equal to  $N$ , i.e. to the number of the grid nodes where we look for the solution. Anyway, for the kind of problem we are dealing with, i.e. the inversion of the displacement field induced by a tsunamigenic seismic source, this is an unlikely work condition. From the inversion of seismological data, in most cases we have at least some knowledge about the source location that enables us to look for the solution only in a re-



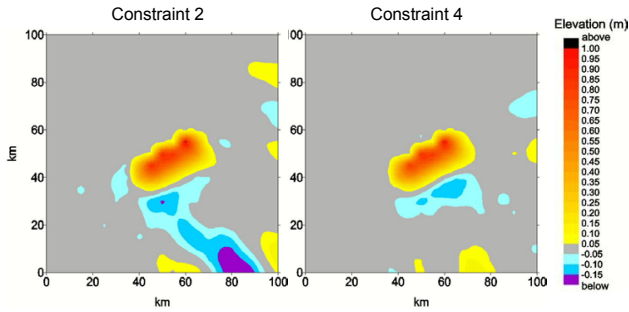
**Fig. 12.** Dependence of the misfit on the spatial constraints shown in Fig. 11.

stricted region of the domain: this represents a type of the so-called *a priori* information. In our inverse modeling this type of *a priori* information can be easily implemented as a set of spatial constraints: more precisely, the elevation on the nodal points belonging to regions where we have not looked for the solution are set equal to zero. As we will show in the following, the use of the *a priori* information reduces the dimension of the model space parameters, leading, in general, to better solutions. We also present an example to show that this kind of constraint should be used with some prudence.

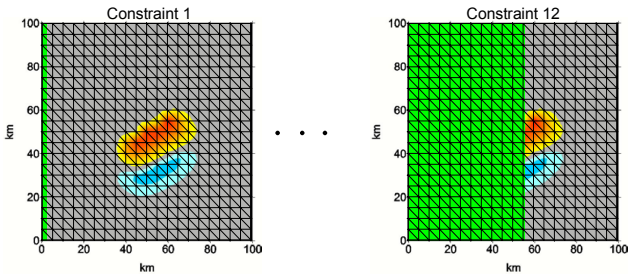
### 5.1 Constraints improving the solution

To show the improvement of the inversion due to the application of the spatial constraints, we consider two cases, already discussed in Sect. 4, for which the unconstrained inversion leads to bad results. The first case is the inversion performed using 10 stations: assuming that the initial field is confined in the middle of the basin, we apply constraints that are progressively stronger, setting to zero the water elevation on the nodes belonging to the grey shaded regions, shown in Fig. 8. The application of the constraint of type 1 reduces the model space dimension from  $N = 441$  to  $N = 361$  ( $\sim 18\%$  of reduction) and leads to a misfit reduction by a factor 6 (see Fig. 9): we may appreciate, in Fig. 10, that now the corresponding inverted initial field shows the main positive wave patch, even if the field is still largely noisy. As shown in Fig. 9 and in Fig. 10, running the inversion with even stronger constraints, rapidly leads to very good results, qualitatively equivalent to those obtained using a much larger number of waveforms (see Figs. 3 and 4 for comparison).

As a second example, we consider the inversion performed using the stations distribution of type 1, characterised by a low azimuthal coverage of the domain and we apply constraints that are somewhat lighter with respect to those used in the previous example (see Fig. 11). Even in this case the information conveyed by the spatial constraints makes the



**Fig. 13.** Representative initial fields inverted by using the stations distribution of type 1 and applying the constraints shown in Fig. 11.

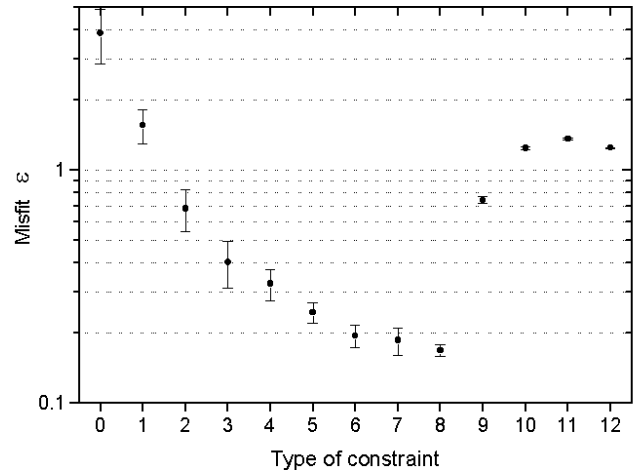


**Fig. 14.** Sketch of progressively stronger constraints: the initial field is taken to be equal to zero in the shaded zones. The theoretical initial elevation field is also shown.

inversion more robust, also leading to a consistent misfit reduction (see Fig. 12). As shown in Fig. 13, even with a bad distribution of stations, we are able to correctly recover the initial elevation field (see Fig. 7 for comparison).

### 5.2 An example of bad constraints

We have seen, until now, that the application of the spatial constraints had always led to improved solutions. Here, we show that the use of a constraint that is too strong will lead to a completely wrong solution. Again, we consider the inversion performed with 10 stations, and we progressively apply the constraints shown in Fig. 14. We found that the misfit steadily decreases (see Fig. 15), as long as an increasing number of constraints are added, until reaching a minimum in correspondence of the constraint number 8; then it abruptly increases as soon as a stronger constraint is applied. The reason for this worsening in the inversion is that with constraint 9 and the following, we make a bad choice of the model space parameters. More precisely, with constraint 9 we start to force to zero a region of the domain where the initial field is consistently different from that value. The worsening of the inversion can be better understood looking at Fig. 16, where three representative inverted initial fields, relative to constraints 3, 8 and 9, are shown. The initial fields are reproduced more and more precisely as stronger constraints are used (constraints 1 to 8); then the inverted field abruptly worsens (constraints 9 to 12). Actually, the inversion model is forced, by taking into account the information



**Fig. 15.** Dependence of the misfit on the spatial constraints shown in Fig. 14.

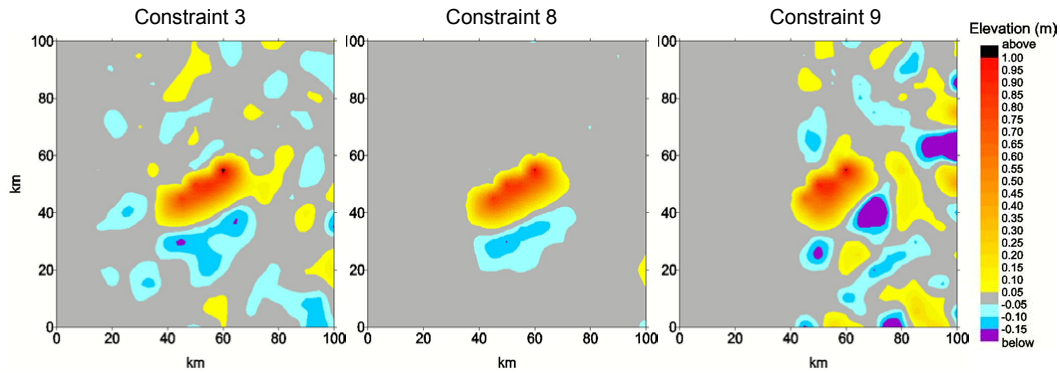
conveyed by the various waveforms and by the applied constraints, to build-up fictitious initial wave systems.

## 6 Conclusions

We applied an inversion method to the problem of retrieving the initial water elevation field that generates a tsunami. When the inversions are unconstrained, we found that to attain good results, the dimension of the data space has to be much larger than that of the model space parameter: to correctly invert the initial field, it was necessary to perform an inversion using at least 12–13 waveforms. We also showed that a large number of records is not sufficient to ensure a good inversion if the corresponding stations do not have a good azimuthal coverage with respect to source directivity. This result should be kept in mind in designing a tide-gage network to study a tsunami source. Since tsunami sources frequently feature a dipolar shape, a lack of tide-gages in the direction perpendicular to the dipole will lead to low quality inversions.

A way to improve the inversions, with a lesser number of waveforms and/or with a tide-gage network that does not have an optimal azimuthal coverage, is to use the available *a priori* information on the source, generally coming from the inversion of seismological data. In this paper we showed how to implement very common information about a tsunami-genic seismic source, i.e. the earthquake source region, as a set of spatial constraints. The results are very satisfactory, since even a rough localisation of the source enables us to correctly invert the initial elevation field.

To conclude, we would like to point out some computational characteristics of the inversion method used in this paper. The main drawback is that the proposed method holds only for linear theory. Moreover, large computer memory is needed to deal with real cases, for which the number of nodes of the finite-element mesh is large. Anyway, given



**Fig. 16.** Representative initial fields inverted by using 10 waveforms and applying the constraints shown in Fig. 14.

a domain and the position of the stations, a large computational effort is required only once, to generate the appropriate Green's functions. Once the Green's functions are computed, the inversions run very fast. To fix some numbers, for the case presented in this paper, 45 min of CPU time (on a 800 MHz clock time based processor) are needed to generate the Green's functions, but only few seconds are needed to run a full inversion.

*Acknowledgements.* This work was carried out on funds from the Gruppo Nazionale di Difesa dai Terremoti (GNDT) of the Istituto Nazionale di Geofisica e Vulcanologia (INGV) and from the Ministero dell'Università e della Ricerca Scientifica e Tecnologica (MURST).

## References

- Arnold V. I.: Ordinary differential equations, MIT Press, 1978.
- Johnson, J. M.: Heterogeneous coupling along Alaska-Aleutian as inferred from tsunamis, seismic and geodetic inversion, *Advances in Geophysics*, 39, 1–116, 1999.
- Johnson, J. M., Satake, K., Holdahl, S. R., and Sauber, J.: The 1964 Prince William Sound earthquake: joint inversion of tsunami and geodetic data, *J. Geophys. Res.*, 101, 523–532, 1996.
- Okada, Y.: Internal deformation due to shear and tensile faults in a half-space, *Bull. Seismol. Soc. Am.*, 82, 1018–1040, 1992.
- Piatanesi, A. and Tinti, S.: The slip distribution of the 1992 Nicaragua earthquake from tsunami run-up data, *Geophys. Res. Lett.*, 23, 37–40, 1996.
- Pires, C. and Miranda, P. M. A.: Tsunami waveform inversion by adjoint methods, *J. Geophys. Res.*, 106, C9, 19 773–19 796, 2001.
- Satake, K.: Inversion of tsunami waveforms for the estimation of a fault heterogeneity: method and numerical experiments, *J. Phys. Earth*, 35, 241–254, 1987.
- Satake, K.: Inversion of tsunami waveforms for the estimation of hereogeneous fault motion of large submarine earthquakes: the 1968 Tokachi-oki and the 1983 Japan sea earthquake, *J. Geophys. Res.*, 94, 5627–5636, 1989.
- Tinti, S. and Piatanesi, A.: Wave propagator in finite-element modeling of tsunamis, *Marine Geodesy*, 18, 273–298, 1995.
- Tinti, S., Piatanesi, A., and Bortolucci, E.: The finite-element wave propagator approach and the tsunami inversion problem, *J. Phys. Chem. Earth*, 12, 27–32, 1996.
- Titov, V. V., Mofjeld, H. O., Gonzales, F. I., and Newman, J. C.: Offshore forecasting of Alaska-Aleutian Subduction Zone tsunamis in Hawaii, NOAA Tech. Memo. ERL PMEL-114, 22, 1999.