Journal of Earthquake Engineering (in press)

REAL-TIME RISK ANALYSIS FOR HYBRID EARTHQUAKE EARLY WARNING SYSTEMS

IUNIO IERVOLINO
Dipartimento di Analisi e Progettazione Strutturale, Università di Napoli Federico II
Via Claudio 21, 80125 Naples, Italy

VINCENZO CONVERTITO
Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Vesuviano
C/o Laboratorio RISSC, via Coroglio 156, 80124 Naples, Italy

MASSIMILIANO GIORGIO
Dipartimento di Ingegneria Aerospaziale e Meccanica, Seconda Università di Napoli
Via Roma 29, 81031 Aversa, Italy

GAETANO MANFREDI*
Dipartimento di Analisi e Progettazione Strutturale, Università di Napoli Federico II
Via Claudio 21, 80125 Naples, Italy
gaetano.manfredi@unina.it

ALDO ZOLLO
Dipartimento di Scienze Fisiche, Università di Napoli Federico II
C/o Laboratorio RISSC, via Coroglio 156, 80124 Naples, Italy

Received (received date)
Revised (revised date)
Accepted (accepted date)

Earthquake Early Warning Systems (EEWS), based on real-time prediction of ground motion or structural response measures, may play a role in reducing vulnerability and/or exposition of buildings and lifelines. In fact, recently seismologists developed efficient methods for rapid estimation of event features by means of limited information of the P-waves. Then, when an event is occurring, probabilistic distributions of magnitude and source-to-site distance are available and the prediction of the ground motion at the site, conditioned to the seismic network measures, may be performed in analogy with the Probabilistic Seismic Hazard Analysis (PSHA). Consequently the structural performance may be obtained by the Probabilistic Seismic Demand Analysis (PSDA), and used for real-time risk management purposes. However, such prediction is performed in very uncertain conditions which have to be taken into proper account to limit false and missed alarms. In the present study, real-time risk analysis for early warning purposes is discussed. The magnitude estimation is performed via the Bayesian approach, while the earthquake localization is based on the Voronoi cells. To test the procedure it was applied, by simulation, to the EEWS under development in the Campanian region (southern Italy). The results lead to the conclusion that the PSHA, conditioned to the EEWS, correctly predicts the hazard at the site and that the false/missed alarm probabilities may be controlled by set up of an appropriate decisional rule and alarm threshold.

Keywords: early warning; real-time; risk; hazard; false alarm; missed alarm

* Corresponding Author. Tel: +39 081 7683486, Fax: +39 081 7683491.

1. Introduction

An Earthquake Early Warning System (EEWS) is made of a network of seismic instruments which can provide real-time measures, on an occurring event, to some central processing station, which elaborates the information for risk management purposes [Heaton, 1985]. In fact, EEWS may help in real-time vulnerability/exposition reduction to minimize losses, or in directing rescue operations immediately after an earthquake for emergency preparedness [Wieland, 2001]. The "early" information, provided by the EEWS in the first seconds of the earthquake, may be used, to spread the alarm to the community, to infer data or to activate different types of security measures, such as shut down of critical systems and stopping of high speed trains [Veneziano and Papadimitriou, 1998].

Earthquake Early Warning Systems may be simplistically classified, on the basis of the seismic network configuration, as *regional* or *site-specific*. Regional EEWS' typically would consist of a number of seismic stations covering a potential source zone. Such systems are designed to provide data that can be used to estimate the main parameters of the event, as time of origin, magnitude and location and to predict ground motion at some other sites in a large area (*rapid response system*). This processing may require significant time and therefore these systems are mainly devoted to near-real-time applications as shake maps, which are territorial distributions of ground shaking employed for emergency management [Wald et al., 1999].

For critical facilities with a large loss potential, a fence of seismic instruments may be placed around the equipment to protect it. This is the case of site-specific EEWS', which reduce the risk connected to the failure of nuclear power plants or lifelines by providing sufficient warning time to take measures to decrease the vulnerability or exposition (*alarm system*) [Wieland et al., 2000]. The networks devoted to site specific Earthquake Early Warning are much smaller than those of the regional type, only covering the surroundings of the system; the distance of the seismic instruments from the facility depends on the lead time needed to activate the safety procedures before the arrival of the more energetic seismic phase. Typically the alarm is issued when the ground motion at one or more sensors exceeds a given threshold. In fact, unlike the regional case, site-specific EEWS' only measure the ground shaking at the network and do not attempt to estimate the features of the event, which would require unacceptable computational time.

Due to a large development of regional networks in recent years worldwide (see SAFER project [2005] for example) the question of using EEWS' for site-specific applications is rising. In fact, nowadays *real-time seismology* allows a *hybrid* use of regional and on-site warning methods [Kanamori, 2005],[Cua, 2005]. Recent efforts on rapid estimation of the earthquake's magnitude and location make available a prediction of the ground motion at the site from a few seconds to a few tens of seconds before its arrival.

The formulation of real-time seismic risk analysis for structure-specific applications of regional EEWS is the topic of the study herein presented. The objective is to predict, in

a full probabilistic approach, a ground motion Intensity Measure (*IM*) at a site (i.e. Peak Ground Acceleration or PGA) and the performance of a structure of interest, in terms of an Engineering Demand Parameter (*EDP*), conditioned to the real-time information provided by the seismic network. A scheme of the hybrid application of a regional network for structure-specific earthquake early warning (point warning) is shown in Figure 1.

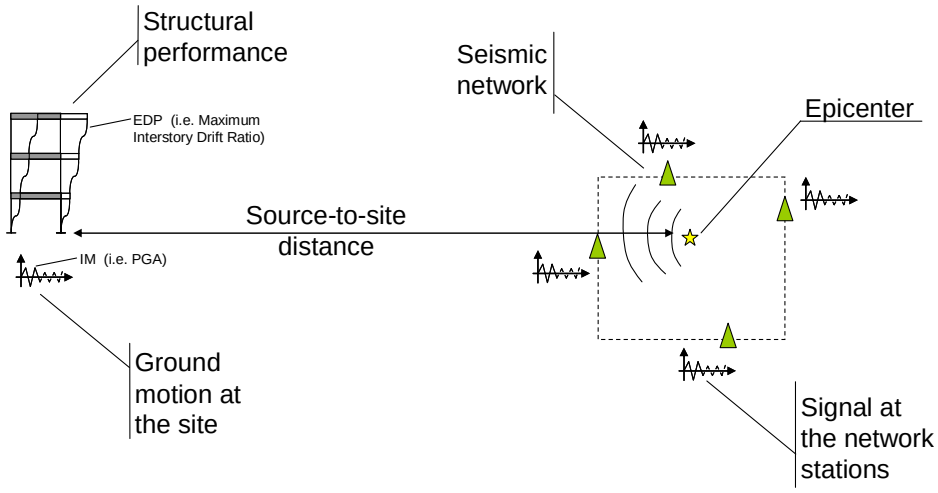

Figure 1. Regional EWWS for structure-specific applications

The risk analysis may help in real-time decision making aimed to vulnerability or exposition reduction. For example, EEWS' predictions may be used for the set-up of active or semi-active structural control, in order to achieve a safer structural response to ground motion.

In the following, in particular, it will be focused on real-time probabilistic seismic hazard analysis for risk management purposes; in this case, the seismic network estimates the earthquake's features and then the system predicts *IM* to give additional lead time. This process, however, includes significant uncertainty which may lead to false and missed alarms which both have a cost. In the case of not alarming, the loss is associated to the earthquake striking without any taken countermeasure; in the case of an alarm, the preparedness interventions have a cost (social and/or economical) which may become a loss if the actual ground motion does not require such actions. Therefore, a key issue in the use of EEWS for risk management, is the estimation of missed and false alarm probabilities associated to the adopted decisional rule. This computation, on an empirical basis, should consist of post-event analysis of EEWS predictions and would require a large strong-motion waveforms database both for the network and the site where the structure is located. Since very rarely these databases are available, especially for large earthquakes, the missed/false alarms rates of occurrence have to be estimated in a

simulation framework (e.g. Montecarlo) using appropriate characterizations of the uncertainties involved in the prediction. This approach, which requires virtually no records other than those used to calibrate the method adopted for the estimation of magnitude (M) and source-to-site distance (R), is applied herein to the developing EEWS of the Campanian region (southern Italy) to estimate the frequency of error in decisions.

2. Seismic Risk Analysis conditioned to the EEWS information

Recently seismologists have developed several methods that enable to estimate the event's magnitude on the basis of measures made on the first few seconds of P-waves signal. Similarly, as described in the following, the source-to-site distance may be predicted analyzing the triggering sequence (i.e. the order in which the seismic stations detect the earthquake). Applying these procedures to data gathered by the network, during the propagation of the event, it is possible to obtain information about M and R and perform real-time Probabilistic Seismic Hazard Analysis (PSHA) [Cornell, 1968],[McGuire, 1995]. This results in a seismic hazard analysis conditioned (in a probabilistic sense) to the real-time information provided by the EEWS. Consequently, the distribution of the structural response may be also computed by Probabilistic Seismic Demand Analysis (PSDA) [Carballo and Cornell, 2000]. The probability density function of the structural response at the site when an event is occurring contains the highest level of information available, and therefore it seems to be the appropriate tool for real-time decision making.

2.1. *Real-time PSHA and PSDA*

At a given time t from the earthquake's origin time, all the real-time information provided by the network can be synthesized in terms of Probability Density Functions (PDFs) of M and R. The PDF of M will be indicated as $f_{M|\tau_1,\tau_2,...,\tau_\nu}\left(m \mid \tau_1,\tau_2,...,\tau_\nu\right)$, because it is conditioned to $\left\{\tau_1,\tau_2,...,\tau_\nu\right\}$, the vector of measures made by the network. (Herein such measure is related to the predominant period of the first four seconds of p-waves recordings, however the procedure virtually applies to any other magnitude-correlated parameter.)

The PDF of R, which, due to the method adopted (see next sections), only depends on the triggering sequence, will be referred as $f_{R|s_1,s_2,...,s_\nu}\left(r \mid s_1,s_2,...,s_\nu\right)$; where $\left\{s_1,s_2,...,s_\nu\right\}$ is such sequence. It is worth to underline that ν is the number of instruments which have triggered and measured the parameter of interest τ at the time t. This makes both the distributions time dependent; in fact, the amount of data included in the estimation process increases with time (i.e. more stations trigger as time flows).

These PDFs may be used to compute the probabilistic distribution, or hazard curve, of a ground motion *IM* (i.e. PGA) at a site of interest by the seismic hazard integral reported in Eq.(1).

$$f_{\nu}\left(im\right) = \\ = \int_{M}\int_{R} f\left(im \mid m,r\right) f_{M|\tau_1,\tau_2,...,\tau_{\nu}}\left(m \mid \tau_1,\tau_2,...,\tau_{\nu}\right) f_{R|s_1,s_2,...,s_{\nu}}\left(r \mid s_1,s_2,...,s_{\nu}\right) dr\, dm \quad (1)$$

The PDF, $f\left(im \mid m,r\right)$, is given by an ordinary attenuation relationship. The subscript ν indicates that the computed hazard refers to a particular set of triggered stations and, consequently, also depends on *t*.

For structural applications of EEWS the prediction of the structural response in terms of an *EDP*, rather than in terms of a ground motion *IM*, may be of main concern. This requires a further integration to get the PDF of *EDP* as reported in Eq.(2),

$$f_{\nu}\left(edp\right) = \int_{IM} f\left(edp \mid im\right) f_{\nu}\left(im\right) d\left(im\right) \quad (2)$$

where the PDF, $f\left(edp \mid im\right)$, is the probabilistic relationship between *IM* and *EDP*. For Moment Resisting Frame (MRF) structures, for example, such relationship is of the type in Eq.(3), relating the Maximum Inter-storey Drift Ratio (MIDR) to $Sa(T_1)$ (first mode spectral acceleration), which are the *IM* and *EDP* respectively.

$$MIDR = a\left(Sa(T_1)\right)^{b} \varepsilon \quad (3)$$

In Eq.(3) the log of ε is a normal random variable with unity-median and variance equal to the variance of MIDR, and the coefficients *a* and *b* are obtained via non-linear incremental dynamic analysis [Vamvatsikos and Cornell, 2002]. Barroso and Winterstein [2002] have proposed a similar relationship for controlled structures.

For the sake of simplicity it will be assumed in the following that the parameter to be predicted is *IM* only. This would keep the presentation of the method clear and the results of the application more easy to interpret. Since EDP is only a probabilistic transformation of *IM*, this choice would not affect the generality of the discussion.

2.2. *Magnitude estimate*

The integral given in Eq.(1) requires the distribution of magnitude (M_s herein), $f_{M|\tau_1,\tau_2,...,\tau_{\nu}}\left(m \mid \tau_1,\tau_2,...,\tau_{\nu}\right)$, to be estimated on the basis of the data provided by the network at a given time. Herein, this PDF has been formulated combining, via the Bayes theorem, historical data and real-time information, Eq.(4), assuming that the stations always perform a measure and that the uncertainty expressed by the distribution of the magnitude, conditioned to such measure, includes eventual error of the instrument.

$$f_{M|\tau_1,\tau_2,...,\tau_{\nu}}\left(m \mid \tau_1,\tau_2,...,\tau_{\nu}\right) = \frac{f_{\tau_1,\tau_2,...,\tau_{\nu}|M}\left(\tau_1,\tau_2,...,\tau_{\nu} \mid m\right) f_M\left(m\right)}{\int_{M_{MIN}}^{M_{MAX}} f_{\tau_1,\tau_2,...,\tau_{\nu}|M}\left(\tau_1,\tau_2,...,\tau_{\nu} \mid m\right) f_M\left(m\right) dm} \quad (4)$$

In the Bayesian framework [Berger, 1985], the distribution $f_M(m)$, which incorporates the information available before the experimental data are collected (e.g. before the network performs the measures, $\{\tau_1, \tau_2, ..., \tau_v\}$), is called *a priori*. In the seismic case it is a truncated exponential, Eq.(5), derived by the well known Gutenberg-Richter recurrence relationship,

$$f_M(m): \begin{cases} \dfrac{\beta e^{-\beta m}}{e^{-\beta M_{\min}} - e^{-\beta M_{\max}}} & M_{\min} \le m \le M_{\max} \\ 0 & m \notin [M_{\min}, M_{\max}] \end{cases} \tag{5}$$

where $\{\beta, M_{\min}, M_{\max}\}$ depend on the seismic features of the region under study. The joint PDF $f_{\tau_1,\tau_2,...,\tau_v|M}(\tau_1, \tau_2, ..., \tau_v \mid m)$, which reflects all the information concerning the magnitude contained into the real-time data, is called *likelihood* function. It has been formulated assuming that, given the magnitude, the τ_i measurements are lognormal, s-independent and identically distributed random variables of parameters reported in Eq.(6)

$$\begin{cases} \mu_{\log(\tau)} = (M - 5.9)/7 \\ \sigma_{\log(\tau)} = 0.16 \end{cases} \tag{6}$$

The value of $\mu_{\log(\tau)}$ is provided by the work of Allen and Kanamori [2003] about the relationship between the magnitude of the event and the log of the predominant period, τ, of the first four seconds of the P-waves (in the velocity recording) for the TriNet network. The dispersion, $\sigma_{\log(\tau)}$, has been retrieved [Fontanella, 2005] from the data reported in the same paper under the omoskedasticity hypothesis. Then the likelihood results as reported in Eq.(7).

$$\begin{cases} f_{\tau_1,\tau_2,...,\tau_v|M}(\tau_1, \tau_2, ..., \tau_v \mid m) = \prod_{i=1}^{v} f_{\tau|M}(\tau_i \mid m) \\ f_\tau(\tau_i \mid m) = \dfrac{1}{\sqrt{2\pi}\ \sigma_{\log(\tau)}\ \tau_i} e^{-\frac{1}{2}\left(\frac{\log(\tau_i) - \mu_{\log(\tau)}}{\sigma_{\log(\tau)}}\right)^2} \end{cases} \tag{7}$$

Since the Bayes theorem enable to correct the *a priori* on the basis of the data collected in real-time, the *posterior* $f_{M|\tau_1,\tau_2,...,\tau_v}(m \mid \tau_1, \tau_2, ..., \tau_v)$, for its own nature, incorporates all the information that is effectively available.

It is worth to underline that lognormality, s-independence and omoskedasticity, hypotheses do not conflict either with results reported in Allen and Kanamori [2003] or with the methods adopted to perform the analyses presented in the same work. Nevertheless, substituting Eq.(7) and Eq.(6) in Eq.(4), is possible to show that $f_{M|\tau_1,\tau_2,...,\tau_v}(m \mid \tau_1, \tau_2, ..., \tau_v)$ depends on data only through the summation of station measurements logarithms and the number of instruments triggered. This may largely reduce the required real-time computational effort. In fact, also due to the features of the

source-to-site distance PDF, the hazard integral may be performed offline for all $\left\{\sum_{i=1}^{\nu}\log(\tau_i),\nu\right\}$ pairs.

2.3. *Real-time location and distance PDF*

The real-time localization methodology considered is that of Satriano et al. [2006] which is based on Horiuchi et al. [2005] and the Equal Differential-Time (EDT) formulation [Font et al., 2004],[Lomax, 2005]. The EDT location is given by the maximum of a stack over quasi-hyperbolic surfaces, defined by the equal differential time equations, where the difference in calculated travel-times to a pair of stations is equal to the difference in observed arrival times for the two stations. Even though for a detailed discussion of the method the reader should refer to the cited papers, a brief description of the procedure is given in the following for readability of the present study.

The algorithm defines a dense (e.g. 1km spaced) grid of points, each of those identified by a set of coordinates $\bar{x}$, in the space (V) below the network. Then the P-waves travel times to the N stations of the network, $\left\{tt_{S_1},tt_{S_2},...,tt_{S_N}\right\}_{\bar{x}}$, may be computed for all the grid points once a velocity model of the region is assigned. The travel times allow to identify the sequence of stations triggered, $\left\{S_1,S_2,...,S_N\right\}_{\bar{x}}$, if that point is the hypocenter of the event.

During an earthquake, when the first station S_1 triggers with an absolute arrival time $t\left(S_1\right)$, it is possible to define a volume that is likely to contain the hypocenter, it is the *Voronoi cell* of that station. This volume is given by the EDT surfaces on which the P-waves travel time to the first triggering station, $tt_{S_1}\left(x\right)$, is equal to the travel-time to any of the not yet triggered stations, $tt_{S_k}\left(\bar{x}\right)$, $k\neq 1$. As time progresses, the information that a new station can trigger only at a time $t\left(S_k\right)\geq t$ is gained. Thus the volume which may contain the hypocenter becomes bounded by the EDT surfaces satisfying the inequalities $tt_{S_k}-tt_{S_1}<t-t_{S_1}$, $k\in\left\{2,3,...,N\right\}$, where t is the current time and the k subscript indicates the not yet triggered stations.

When the second and later instruments trigger, a second set of equations (the standard EDT equations), which account for all pairs of triggered stations, have to be considered: $tt_{S_k}-tt_{S_m}=t_{S_k}-t_{S_m}$ (S_k, S_m are triggered and $k\neq m$). These EDT surfaces are stacked with the volume bounded by updated EDT surfaces (defined by the not yet triggered stations).

The algorithm assigns, to any of the grid points a number, which may be interpreted as the probability of the hypocenter to be coincident with that point, depending on the number of equations that are satisfied. The computation of such probability for all the points gives the PDF of the hypocentral location in *V*. Consequently, the estimate of the

epicentral distance, $f_{R|s_1,s_2,...,s_\nu}\left(r \mid s_1, s_2, ..., s_\nu\right)$, may be retrieved by a geometrical transformation assigning, to any particular distance value, a probability which is the sum of the probabilities of all points of the grid with the same epicentral distance to the site.

One of the main features of the procedure is that, as more stations trigger, the number of not yet triggered stations becomes smaller and location converge towards the hypocentral volume resulting from the standard EDT. Within 3 seconds reliable location (few kilometers uncertainty) can be obtained, also in the case of multiple, concurring events.

3. Decisional rule and the cry wolf issue

Once the EWWS provides a distribution of the ground motion intensity measure, or seismic demand for the structure of interest, a decisional condition has to be checked to launch the alarm or not. Several options are available to formulate a decisional rule, for example the alarm may be issued if the probability of the random variable exceeding a critical threshold (IM_C) outcrosses a reference value (P_C) as in Eq.(8). The P_C and IM_C values have to be set in relation to an appropriate loss function for the structure of interest and the acceptable probabilities of error in decision.

$$Alarm\ if:\ 1-\int_0^{IM_C} f_\nu\left(im\right) d\left(im\right) = P\left[IM > IM_C\right] > P_C \tag{8}$$

The performance of the early warning system may be tested verifying if it "correctly" predicts the distribution of *IM* at the site; the efficiency of the decisional rule may be evaluated in terms of false and missed alarms probabilities (the *cry wolf* issue), P_{FA} and P_{MA} respectively [Patè-Cornell, 1986]. The false alarm occurs when, on the basis of the information processed by the EEWS, the alarm is issued while the intensity measure at the site IM_T (T subscript means "true" indicating the realization of the random variable to distinguish it from the prediction) is lower than the threshold IM_C; the missed alarm corresponds to not launching the alarm if needed, Eq.(9).

$$\begin{cases} Missed\ \ Alarm: \left\{No\ \ Alarm \cap IM_T > IM_C\right\} \\ False\ \ Alarm: \quad \left\{Alarm \cap IM_T \le IM_C\right\} \end{cases} \tag{9}$$

It has been discussed how the information and the uncertainties involved are dependent on the number of stations triggered at a certain time. Therefore, in principle, the decisional rules may be checked at any time since the first station has triggered and, consequently, the false and missed alarm probabilities are, also indirectly, a function of time. From this point of view the decisional process is again time dependent, and one may decide to alarm when the trade-off between the available lead time and the losses related to a missed or a false alarm is at its optimum. As an application, probabilities of these events according to the decisional rule, have been estimated by simulation for the Campanian EEWS.

4. Simulation of the SAMS Earthquake Early Warning System

The Campanian early warning system (SAMS - Seismic Alert Management System) is based on the seismic network located in the Campano-Lucano Appennines [Weber et al., 2006]. Such network is operating in the seismically most active area for the Campanian region (100x80km^2 wide) and it is designed to acquire non-saturated data for earthquakes larger than M_W 4. In Figure 2 the 30 stations of the EEWS network (dark squares), the $M_W > 2$ events recorded from 1981 to 2002 and the faulting system of the Irpinia 1980 earthquake are given. Light color squares represent additional stations which will be used to calibrate local attenuation relationships.

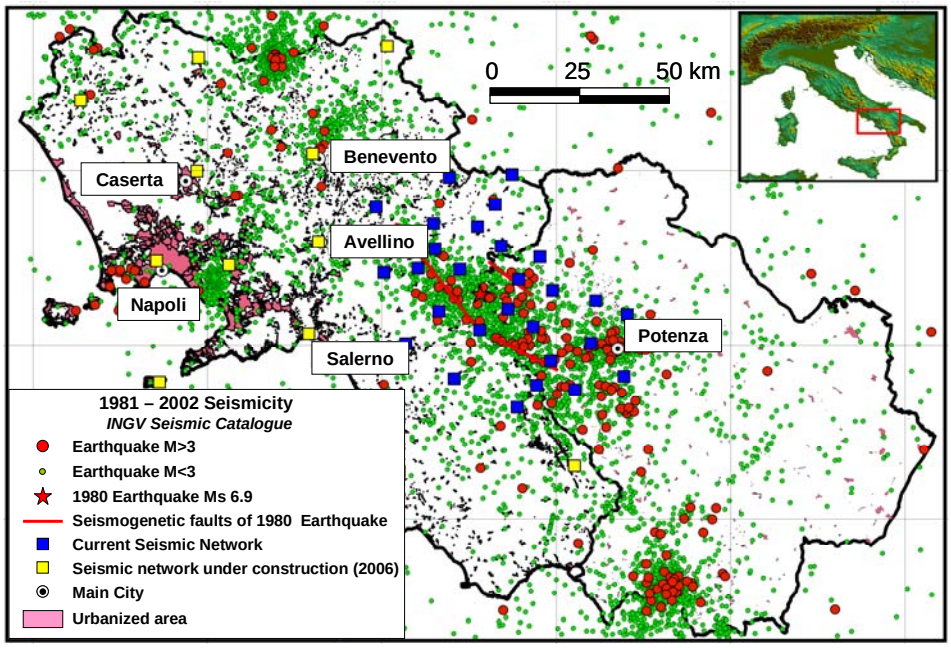

Figure 2. The sensors network and the background seismicity of the region

The real-time seismic hazard analysis procedure presented was applied to simulate the predictions of the SAMS and to assess the false/missed alarms probabilities. (In principle, to simulate the prediction of the *IM* and to compare it with the actual value at the site, for any event, a set of recordings in each station and at the site should be available. However, it is possible to simulate the behavior of the system without recorded signals, which are not available especially for rare events, but still on an empirical basis.) The procedure has been implemented in a computer code and it takes advantage of the discussed methods for the estimation of magnitude and distance. The predicted *IM* is the PGA.

Each run simulates a specific seismic event occurring in the area of interest and consists of three steps: (1) Simulation of event's features (e.g. assignment of magnitude, location and true PGA at the site); (2) Simulation of the measurements and predictions (e.g. real-time PSHA) made by the network at any instant until the trigger of all the stations; (3) Check of the decisional rule and of the false/missed alarm conditions; (4) Increase of false/missed alarms counter to compute the frequencies of occurrence. The flow chart of the simulation procedure is given in Figure 3.

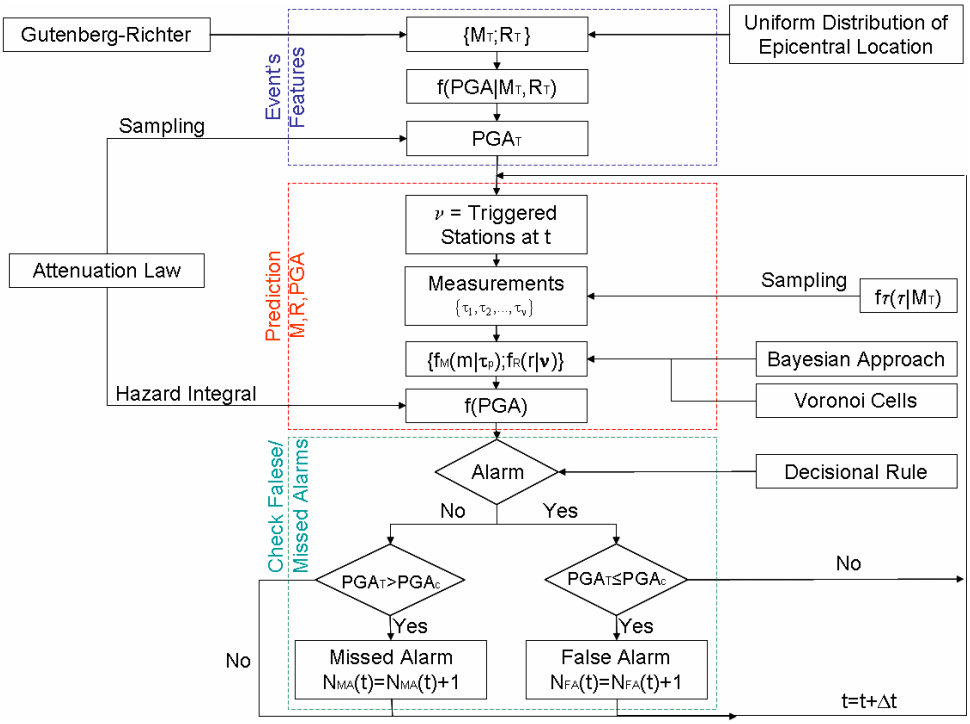

Figure 3. Simulation's flow chart

The site considered in the simulation is assumed to be in the city of Naples, which is approximately 110km far from the center of the network (30s lead time). In Figure 4 the relative position of the network and the site are given in analogy with the scheme of Figure 1.

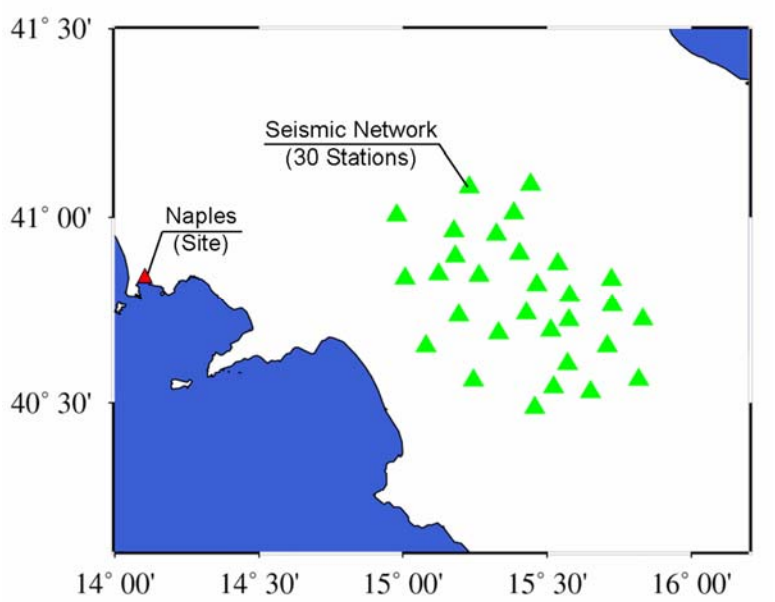

Figure 4. The EEWS-site Campanian scheme

4.1. *Event and ground motion features generation*

Each run in the Montecarlo simulation starts with the generation of the geophysical features the EEWS will try to estimate. These values will define the earthquake of that run. In other words, in order to compute the PSHA conditioned to the EEWS information, the distributions of M and R are required, then it is needed to establish the true value of those which will be called M_T and R_T (true magnitude and true source-to-site-distance respectively).

The true magnitude of the event (M_T) may be sampled according to the Gutenberg-Richter recurrence relationship for the Campanian region (in Eq.(5): $\beta = 1.69$, $M_{\min} = 4$, $M_{\max} = 7$), this will lead, over many runs, to a marginal evaluation of the EEWS performances. On the other hand, one may be interested in evaluating the EEWS' performance with respect to a specific magnitude; in this case M_T for the all runs in the simulation has to be set to a fixed value of interest. This is useful in the light of assessing the EEWS performance conditioning it to high magnitude events, which are the more threatening.

The location of the epicenter may be randomly chosen sampling its coordinates $\{x_{epi}, y_{epi}\}$ from two s-independent uniform distributions defined in the area covered by the network. Once the epicentral coordinates are set, the distance R_T to the site (e.g. Naples) is readily obtained. Again, for some purposes one may want to set the location of

the epicenter to the same point assigning the same value of R_T for all the simulations. In the following, for both M_T and R_T, this second option will be followed for sake of clarity and readability of results.

The generation (or assignment) of a "true" magnitude and "true" distance in each Montecarlo run allows to get reference values for the prediction of the EEWS; similarly, the "true" ground motion at the site (PGA_T) should also be set. The PGA experienced at the site is required to establish whether the decision, ensuing the decisional rule, produced in that run, a missed or a false alarm. The value of PGA_T, consistent with the values of M_T and R_T, is obtained by sampling the attenuation relationship which, by definition, provides the PDF of the ground motion intensity measure conditioned to $\{M_T, R_T\}$. (If many recorded signals would be available at the site for a given magnitude the empirical distribution of the PGA, as retrieved by the records, should be the same as provided by the attenuation law.) Herein the Sabetta and Pugliese [1996] attenuation is considered in its epicentral formulation. Therefore, in each run the value of PGA_T is sampled from a completely specified lognormal random variable.

Finally the event is defined for the EEWS purposes since $\{M_T, R_T, PGA_T\}$ of the run are set; the next step consists of simulating the measurements at the stations consistently with these event's features.

4.2. *Station measurements and M,R real-time distributions*

In the simulation process, at any given time *t*, the number ν and sequence of stations triggered is computed assuming a homogeneous and isotropic propagation model with P- and S-waves velocities of 5.5km/s (V_P) and 3.5km/s (V_S) respectively. Similarly the lead time, defined as the remaining seconds required by the S-waves to hit the site, is calculated.

Once the event is defined by $\{M_T, R_T, PGA_T\}$, the measurements of τ for the triggered stations are needed to perform the real-time risk analysis. For example let's consider first the case when only one station should have measured τ. It is possible to simulate the station's measurement by sampling the empirical distribution of τ conditioned to the true magnitude of the event, $f_{\tau|M}(\tau \mid M_T)$. Real τ values measured from recorded signals would be distributed as $f_{\tau|M}(\tau \mid M_T)$ by definition, and therefore such sampling is appropriate in a simulation approach.

To generate τ for more than one station, consistently with section 2.2, the measurements are considered s-independent conditionally to the event's magnitude (M_T). Therefore, if ν stations are triggered, all the ν components of the $\{\tau_1, \tau_2, ..., \tau_\nu\}$ vector are obtained by sampling ν times the $f_{\tau|M}(\tau \mid M_T)$ PDF. Specific Campanian $f_{\tau|M}(\tau \mid m)$ are not yet available and data by Allen and Kanamori [2003] have to be used, they are based on τ measurements on four seconds of recording, then herein the

working hypothesis is that any station's measurement is considered in the process if four seconds have passed from its trigger.

Once the measurements vector $\{\tau_1, \tau_2, ..., \tau_\nu\}$ is defined, the discussed Bayesian method may be applied to compute the magnitude's distribution. In Figure 5 the resulting magnitude PDFs, for a single simulation of an M 6 event, are given.

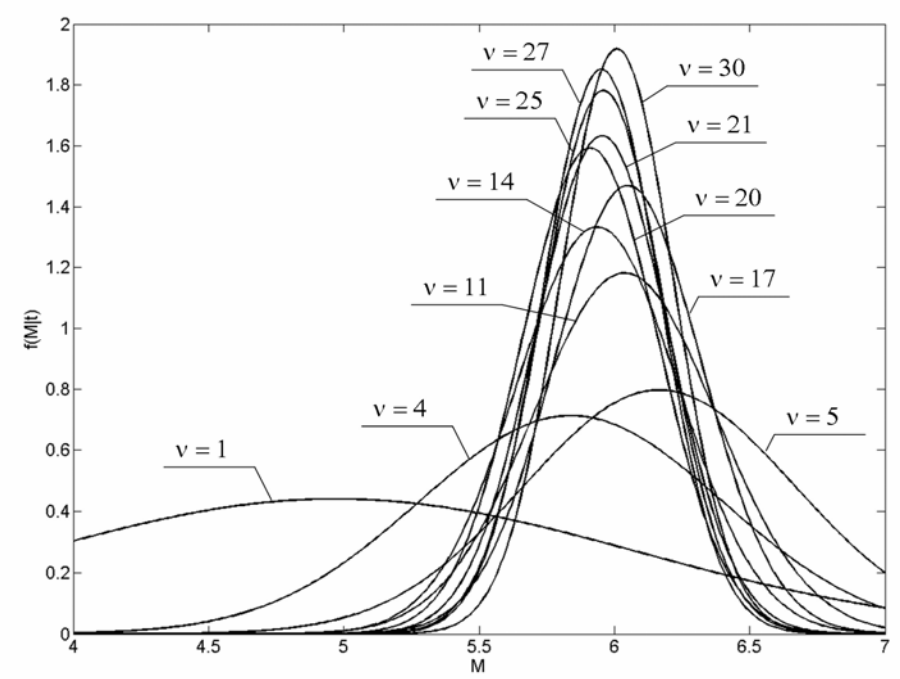

Figure 5. Magnitude distribution as the number of triggered stations increases ($M_T = 6$, $R_T = 91km$)

When only few stations are triggered, the distributions underestimate the magnitude. In fact, when data are few the dominating information is that *a priori* of Eq.(5), which naturally tends to give larger occurrence probability to low magnitude events. More precisely, the Bayesian approach will tend to produce overestimates of magnitude when it is below the a priori mean and it will tend to underestimate it when it is larger than the mean. This effect is directly proportional to the difference between the expected value of the a priori and M_T and inversely proportional to the size of measurements' vector. As more measurements became available, the prediction centers on the real value with a relatively small uncertainty. An estimator with these features is said biased by classic statisticians and other methods can be considered to get an unbiased estimators (i.e. maximum likelihood). Nevertheless the Bayesian approach was preferred because, even slightly biased, it gives, in the mean, smaller estimation errors due to the use of the a priori information.

The distribution of the source-to-site distance, $f_{R|S_1,S_2,...,S_\nu}\left(r \mid S_1, S_2, ..., S_\nu\right)$, is only dependent on the sequence of stations triggered at *t*, which is known by computing the P-waves travel times ($t_{e,j} = R_j / V_P$ where R_j is the distance of the *j-th* station from the epicenter) for all the stations in the network. It has been shown that, due to the features of the SAMS, the localization method, after only three seconds from the first trigger, determines the epicenter's with only 1km of uncertainty [Satriano et al., 2006], which is negligible in respect to the other uncertainties involved in the process. Therefore, in principle, since the magnitude estimation starts after four seconds from the trigger of the first station, at that time it may be assumed that the source-to-site distance is known.

4.3. *Results and discussion*

The hazard integral of Eq.(1), with the estimated distributions of M and R and the attenuation relationship, allows the exceeding probability of PGA at the site to be computed as the event evolves and the stations trigger. The hazard curves corresponding to the event simulated in Figure 5 are given in Figure 6; the latter shows that the hazard increases as time flows (more τ measurements available), consistently with the magnitude estimation results given in the former.

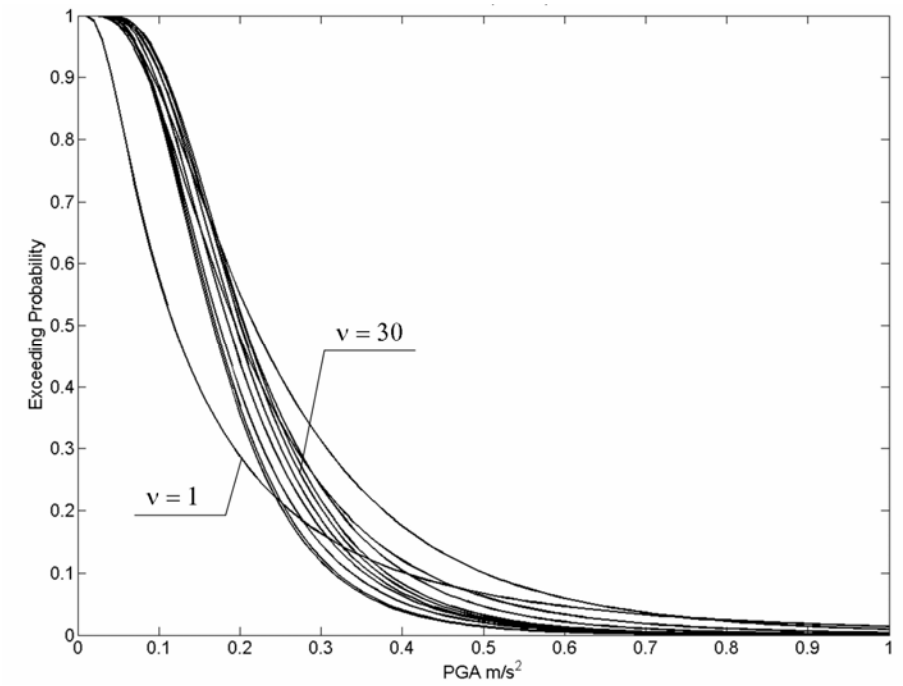

Figure 6. EWWS conditioned seismic hazard as the number of stations increases ($M_T = 6$, $R_T = 91km$)

To discuss the hazard computed with the EEWS information, it may be compared to the "maximum knowledge status" which is the hazard calculated adopting the true value of magnitude and distance (this corresponds to consider no uncertainty in the M and R estimations). Such comparison is shown in Figure 7: the thick curve represents the complementary Cumulative Density Function (CDF) for the PGA when M_T (7) and R_T (110km) are deterministically known; the thin curves are the results of 200 simulated events (see subsections 4.1 and 4.2) with the same magnitude and distance features of the maximum knowledge hazard. In the figure only the hazard curves corresponding to the case when all the stations triggered, $v = 30$, are reported.

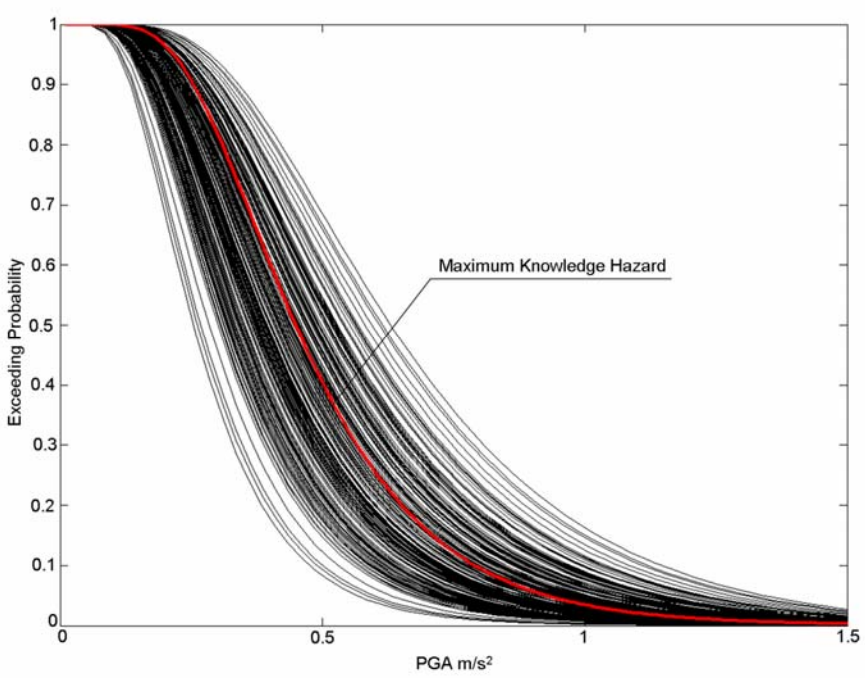

Figure 7. EWWS conditioned seismic hazard in 200 simulations compared to the maximum knowledge condition ($M_T = 7$, $R_T = 110km$)

The EEWS' hazard correctly approximates the maximum knowledge condition even with significant variability of the curves. To reduce the latter, two strategies are possible: (a) developing more efficient procedures for rapid-estimation of magnitude or (b) increasing the size of the measurement sample, for instance, having a larger set of seismic stations in the source area. Option (b) may be costly or even unfeasible, but the estimation procedure would take significant advantage of it. As an example, the simulation of the SAMS has been performed fictitiously considering a network of a double size (60 stations) in the same area. In Figure 8 200 simulations, for events with

the same features of those in Figure 7, are given. The reduction in variability of the curves is significant.

The thick curve in the two figures also point out that, in the hazard integral, the dominating uncertainty is that of the attenuation relationship; then to ultimately improve the prediction of the PGA, a local (i.e. less disperse) attenuation model is necessary.

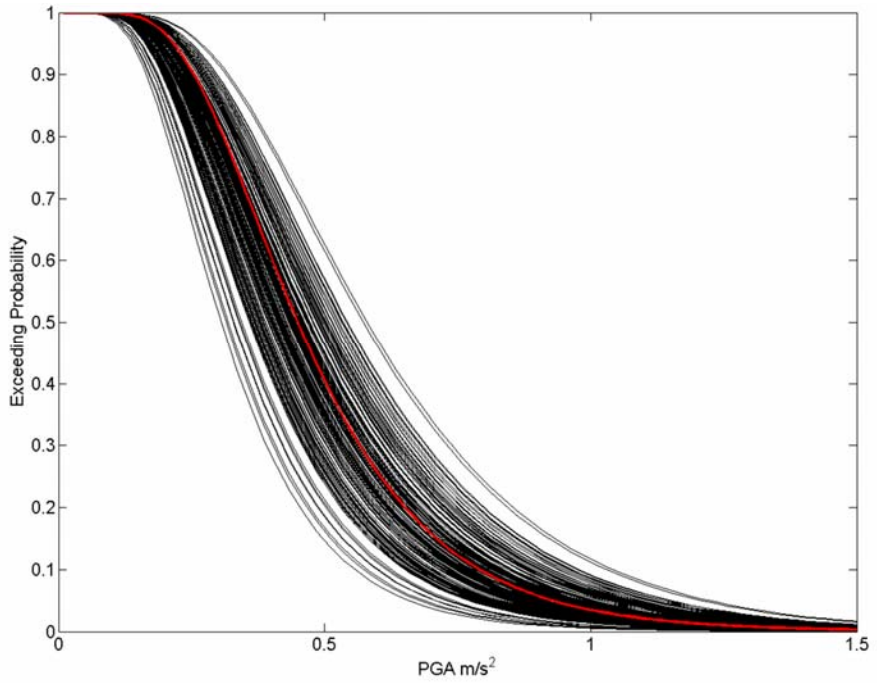

Figure 8. EWWS conditioned seismic hazard in 200 simulations with a seismic network of a double size ($M_T = 7$, $R_T = 110km$)

The simulation allows the computation of the frequency of missed and false alarms. For example, these probabilities are estimated by Eq.(10), where the numerator is the number of occurrence of MA or FA and N_{TOT} is the number of simulated events. Herein MA and FA only refer to wrong ground motion predictions neglecting issues due to system malfunctioning.

$$\begin{cases} P_{MA} \cong \dfrac{N\left[P\left[PGA > PGA_C\right] \le P_C \cap PGA_T > PGA_C\right]}{N_{TOT}} \\ P_{FA} \cong \dfrac{N\left[P\left[PGA > PGA_C\right] > P_C \cap PGA_T \le PGA_C\right]}{N_{TOT}} \end{cases} \tag{10}$$

In Figure 9 such estimations are given for M 7 events with an epicentral distance of 110km (10^4 simulations); Δ and + symbols distinguish missed and false alarm plots

respectively. The curves are a function of time because the real-time PSHA is performed at each second from the first measurement, therefore, the decisional rule may be checked at any instant; consequently the false and missed alarm occurrences reflect the case if the decision of alarming or not would be taken at that instant.

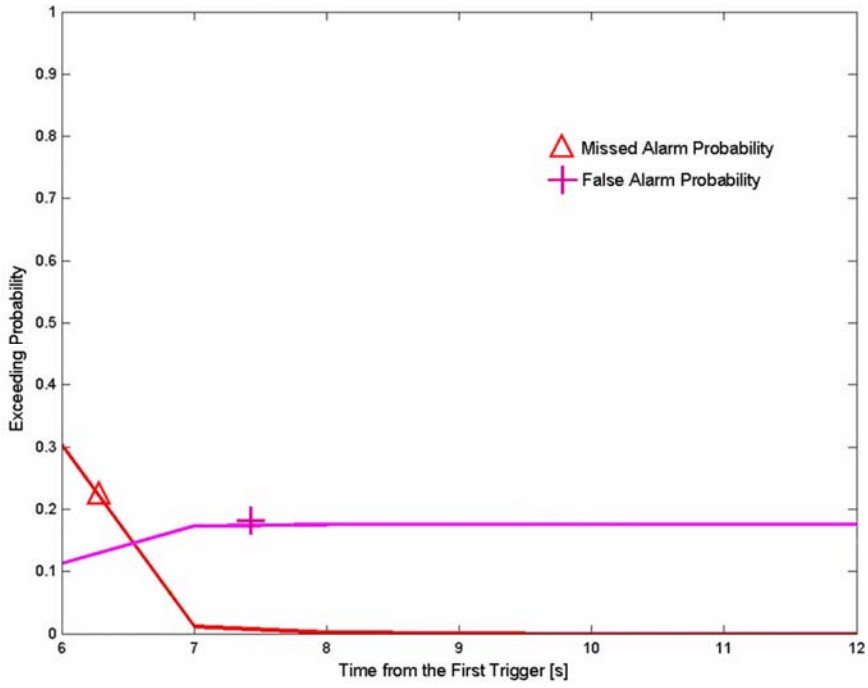

Figure 9. False and missed alarm probabilities for 10^4 events ($M_T = 7$, $R_T = 110km$, $PGA_C = 0.3m/s^2$, $P_C = 0.2$)

To better understand the results of Figure 9 it is useful to discuss the given curves. The PGA_C, in decisional rule of Eq.(8), is arbitrarily set to 0.3m/s^2 and the critical probability of exceedance (P_C) is 0.2; the true values of magnitude and distance are $M_T = 7$ and $R_T = 110km$ respectively. The chosen maximum knowledge status (e.g. the attenuation relationship conditioned to M_T and R_T) gives $P\left[PGA > PGA_C\right] = 0.81$, then if P_C is equal to 0.2, the right decision would be to alarm in every run. As a consequence the probability of missed alarm is zero by design, because the alarm should always be issued, and the probability of a false alarm is $P\left[PGA \le PGA_C\right]$ or 1-0.81 = 0.19. These probabilities are intrinsic to the decisional rule and the thresholds set. However, as discussed, the EEWS cannot perfectly estimate the hazard curve with M_T and R_T known. In fact, due to the variability of the estimations (see Figure 7) the value of $P\left[PGA > PGA_C\right]$ is sometimes underestimated and sometimes overestimated. For

example the underestimation of $P[PGA > PGA_C]$ may lead to not issue the alarm even if required and therefore the missed alarm probability is not zero. In particular, when there are few triggered stations this underestimation effect is strong, and the missed alarm probability is relatively high because when the alarm is not launched (incorrectly) it will most likely result is a missed alarm (in fact the true $P[PGA > PGA_C]$ is 0.81). As time increases the estimation improves; at 10s since the first trigger 23 station out of 30 have measured τ, and $P[PGA > PGA_C]$ approaches its correct value (0.81), then the missed and false alarm probabilities also tend to their correct values, 0 and 0.19 respectively. This means that, when all stations are triggered, the systems would work according to what designed. This would suggest to alarm when P_{MA} and P_{FA} are at their a priori values, however this happens only since a certain time, when a sufficient number of stations is triggered, this means that alarming in the early seconds of the event gives additional lead time but implies to accept some error probability which may be intolerable.

The shape of the curves depends on what both the chosen values of PGA_C and P_C may be, consequently, they may vary different from those discussed in this example if other values of the thresholds are concerned. Nevertheless, given the missed and false alarms reference values, calculated by means of the hazard conditioned to M_T and R_T, the system may be calibrated setting PGA_C and P_C by an appropriate loss function. The choice of PGA_C mainly depends on the seismic response of the protected infrastructure and on the related damage to structural and non structural elements, while the choice of P_C is related to the consequences of a false/missed protection action and on the minimum lead time necessary to develop this protection action in a safe manner.

5. Conclusions

The study presented in this paper investigated use of Earthquake Early Warning Systems for real-time decision making and seismic risk management. The information provided by the seismic network on magnitude and source-to-site distance, on a developing event, may be treated as in the classic hazard analysis which is the basis to obtain a prediction of the required structural or non structural performance. The approach has been tested simulating the Campanian early warning system implementing recent advances of real-time seismology in a probabilistic framework. Results indicate that the PSHA, conditioned to the EEWS measures, correctly approximates the hazard computed if the magnitude and distance would be deterministically known, which is the maximum level of knowledge. The residual variability may be reduced adopting more efficient estimation methods (i.e. regionally calibrated relationships or different inferential strategies) and/or, if possible, designing a more dense seismic network.

While real-time seismic risk analysis seems the way to use all the information provided by the Earthquake Early Warning System, on the other hand the prediction involves significant uncertainty which cannot be neglected because it may result in a low efficiency of hybrid EEWS in respect to site-specific systems. Decisional rule and alarm

thresholds have intrinsic missed and false alarm probabilities which should be set according to appropriate loss functions related to the system to protect. Simulations, by means of time dependent curves, show how the rates of error in decision evolve, with time approaching to their design values as the number of triggered stations increases. Such curves may be used for risk management optimizing the trade off between the probability of wrong decisions and the available lead time for risk reduction actions.

Acknowledgements

This research was carried out in the framework of AMRA scarl [http://www.amra.unina.it/]. The authors would also acknowledge Professor Daniele Veneziano (Massachusetts Institute of Technology, USA) and Dr. Gianpaolo Pulcini (Consiglio Nazionale delle Ricerche, Italy) for their helpful discussions on the topic. Finally authors thank Miss Racquel K. Hagen (Stanford University) for proofreading the paper.

References

Allen, R.M., Kanamori, H., The Potential for Earthquake Early Warning in Southern California. *Science* 2003; **300**: 786-789.

Barroso, L.R., Winterstein, S., Probabilistic Seismic demand analysis of controlled steel moment-resisting frame structures, *Earthquake Engineering and Structural Dynamics* 2002; **31**: 2049-2066.

Berger, J.O., *Statistical Decision theory and Bayesian Analysis*, Springer-Verlag New York, 1985.

Carballo, J.E., Cornell, C.A., *Probabilistic Seismic Demand Analysis: Spectrum Matching and Design*. Department of Civil and Environmental Engineering, Stanford University. Report No. RMS-41, 2000.

Cornell, C.A., Engineering seismic risk analysis. *Bull. Seismol. Soc. Am.* 1968; **58**: 1583–1606.

Cua, G., *Creating the Virtual Seismologist: developments in ground motion characterization and seismic early warning*, Ph.D. Thesis, California Institute of Technology. Advisor: T. Heaton, 2005.

Font, Y., Kao, H., Lallemand, S., Liu, C.-S., Chiao, L.-Y., Hypocentral determination offshore Eastern Taiwan using the Maximum Intersection method. *Geophys. J. Int.* 2004: **158**: 655-675.

Fontanella, N., *Gestione del Rischio Sismico nella Regione Campania: Formulazione e Calibrazione del Simulatore del Sistema di Early Warning Sismico per il Progetto SAMS*, B.A. Thesis, University of Naples Federico II. Advisors: M. Giorgio, I. Iervolino, V. Convertito, 2005 (in Italian).

Heaton, T.H., A model for a seismic computerized alert network. *Science* 1985; **228**: 987-990.

Horiuchi, S., Negishi, H., Abe, K., Kamimura, A., Fujinawa, Y., An Automatic Processing System for Broadcasting Earthquake Alarms. *Bull. Seism. Soc. Am.* 2005; **95**: 708-718.

Kanamori, H., Real-Time Seismology and Earthquake Damage Mitigation. *Annual Review of Earth and Planetary Science* 2005; **33**: 195-214.

Lomax, A., A Reanalysis of the Hypocentral Location and Related Observations for the Great 1906 California Earthquake. *Bull. Seism. Soc. Am*. 2005; **95**: 861-877.

McGuire, R.K., Probabilistic seismic hazard analysis and design earthquakes: Closing the loop, *Bull. Seismol. Soc. Am.* 1995; **85**: 1275-1284.

Patè-Cornell, M.E., Warning systems in risk management. *Risk Management* 1986; **6**: 223-234.

Sabetta, F. Pugliese, A., Estimation of response spectra and simulation of nonstationarity earthquake ground motion. *Bull. Seismol. Soc. Am.* 1996; **86**: 337 - 352.

Satriano, C., Lomax, A., Zollo, A., Optimal, real-time earthquake location for early warning. *Seismic Early Warning*, P. Gasparini, G. Manfredi and J. Szchau (eds.), Springer-Verlag, 2006.

Seismic eArly warning For EuRope – Safer. *Sixth Framework Programme Call: Fp6-2005-Global-4 Sustainable Development, Global Change and Ecosystem Priority 6.3.IV.2.1: Reduction of seismic risks*, 2005.

Vamvatsikos, D., Cornell, C.A., Incremental Dynamic Analysis, *Earthquake Engineering and Structural Dynamics* 2002; **31**:491-514.

Veneziano, D., Papadimitriou, A.G., Optimization of the Seismic Early Warning System for the Tohoku Shinkansen. *11th European Conference on Earthquake Engineering.* Paris, France, 1998.

Wald, D.J., Quitoriano, V., Heaton, T.H., Kanamori, H., Scrivner, C.W., orden, B.C., TriNet "ShakeMaps": Rapid Generation of Peak Ground Motion and Intensity Maps for Earthquake in Southern California. *Earthquake Spectra* 1999; **15**: 537-555.

Weber, E., Iannaccone, G., Zollo, A., Bobbio, A., Cantore, L., Corciulo, M., Convertito, V., DiCrosta, M., Elia, L., Emolo, A., Martino, C., Romeo, A., Satriano, C., Development and testing of an advanced monitoring infrastructure (ISNet) for seismic early warning applications in the Campania Region, Southern Italy. *Seismic Early Warning*. P. Gasparini, G. Manfredi and J. Szchau (eds.), Springer-Verlag, 2006.

Wieland, M., Earthquake Alarm, Rapid Response, and Early Warning Systems: Low Cost Systems for Seismic Risk Reduction. *International Workshop on Disaster Reduction*, Reston, VA, 2001.

Wieland, M., Griesser, M., Kuendig, C., Seismic Early Warning System for a Nuclear Power Plant. *Proc. of 12° World Conference on Earthquake Engineering*. Auckland, New Zealand, 2000.