Uncertainty in early warning predictions of engineering ground motion parameters: What really matters?

Iunio Iervolino,[1] Massimiliano Giorgio,[2] Carmine Galasso,[1] and Gaetano Manfredi[1]

Received 12 November 2008; revised 7 January 2009; accepted 9 January 2009; published 26 February 2009.

[1] From the engineering perspective, the effectiveness of earthquake early warning systems (EEWS) depends only on the possibility of immediately detecting the earthquake and estimating the expected loss at a location of interest, in order to undertake actions to manage/mitigate the risk before the strike. The simplest proxy for the earthquake's destructive potential is the peak ground acceleration (PGA), which is predicted through probabilistic seismic hazard analysis in the framework of EEW. In this paper, the effects of different sources of uncertainty on the prediction of PGA are assessed with reference to the ISNet (Irpinia Seismic Network) EEWS. First the analyses show how the uncertainty of the ground motion prediction equation (GMPE) dominates those of magnitude and distance, almost independently of the information available for the event. Secondly, based on these findings, information-dependent lead-time maps are provided for the Campania (southern Italy) region. Finally, different real-time magnitude estimators are compared in terms of errors in the prediction of PGA, as a more efficient estimator may give additional lead-time for risk reduction. **Citation:** Iervolino, I., M. Giorgio, C. Galasso, and G. Manfredi (2009), Uncertainty in early warning predictions of engineering ground motion parameters: What really matters?, *Geophys. Res. Lett.*, *36*, L00B06, doi:10.1029/2008GL036644.

1. Introduction to RTPSHA

[2] Seismologists have recently developed several methods to estimate the magnitude (M) of an event, given limited information of the P-waves, for real-time applications. Similarly, the source-to-site distance (R) may be determined by analyzing the time and order of the seismic stations detecting the developing earthquake. Therefore, for earthquake early warning purposes, it is possible to assume that at a given instant estimates of M and R are available, and the peak ground motion at the site can be predicted via probabilistic seismic hazard analysis (PSHA) conditional to the real-time information given by the seismic sensors network [*Iervolino et al.*, 2006].

[3] In fact, assuming that at a given time t from the earthquake's origin, the seismic network can provide a vector of measures informative for the magnitude, $\{\tau_1, \tau_2, \ldots, \tau_n\}$, for example the predominant periods of the first four seconds of the P-waves [*Allen and Kanamori*, 2003], then the posterior probability density function (PDF) of M conditional to the measures, $f(m|\tau_1, \tau_2, \ldots, \tau_n)$, may be obtained via the Bayes theorem, equation (1).

$$f(m|\tau_1,\tau_2,\ldots,\tau_n)=\frac{f(\tau_1,\tau_2,\ldots,\tau_n|m)f(m)}{\int_{M_{\min}}^{M_{\max}} f(\tau_1,\tau_2,\ldots,\tau_n|m)f(m)dm} \quad (1)$$

[4] In the Bayesian framework $f(m)$, the *a priori* distribution, is used to incorporate the information available before the seismic network performs the measurements. Then, in this application, a natural candidate for $f(m)$ is the truncated exponential of equation (2), derived by the Gutenberg-Richter relationship. In equation (2), $\{\beta, M_{\min}, M_{\max}\}$ are dependent on the seismic features of the region under study and will be assumed in the following as {1.69, 4, 7}.

$$f(m)=\begin{cases}\frac{\beta e^{-\beta m}}{e^{-\beta M_{\min}}-e^{-\beta M_{\max}}} & M_{\min}\le m\le M_{\max}\\ 0 & m\notin[M_{\min},M_{\max}]\end{cases} \quad (2)$$

[5] The joint PDF $f(\tau_1, \tau_2, \ldots, \tau_n|m)$ is the *likelihood* function. It is used to incorporate into the analysis all information on M contained into the real-time data. Under the hypothesis that the τ measurements are stochastically independent and identically distributed lognormal random variables it may be formulated as in equation (3).

$$f(\tau_1,\tau_2,\ldots,\tau_n|m)=\left(\frac{1}{\sqrt{2\pi}\sigma_{\ln(\tau)}}\right)^n \cdot\left(\prod_{i=1}^{n}\frac{1}{\tau_i}\right)e^{\left(-\frac{\sum_{i=1}^{n}(\ln(\tau_i))^2}{2\sigma^2_{\ln(\tau)}}\right)}e^{\left(\frac{2\mu_{\ln(\tau)}\left(\sum_{i=1}^{n}\ln(\tau_i)\right)-n\mu^2_{\ln(\tau)}}{2\sigma^2_{\ln(\tau)}}\right)} \quad (3)$$

[6] The parameters in equation (3) are the mean, $\mu_{\ln(\tau)}$, and the standard deviation, $\sigma_{\ln(\tau)}$, of τ from the study of *Allen and Kanamori* [2003], equation (4).

$$\begin{cases}\mu_{\ln(\tau)}=(M-5.9)/(7\log(e))\\ \sigma_{\ln(\tau)}=0.16/\log(e)\end{cases} \quad (4)$$

[7] Combining equation (3) and equation (2) into equation (1), the posterior PDF of the magnitude results that

[1]Dipartimento di Ingegneria Strutturale, Università degli Studi di Napoli Federico II, Naples, Italy.
[2]Dipartimento di Ingegneria Aerospaziale e Meccanica, Seconda Università degli Studi di Napoli, Aversa, Italy.

0094-8276/09/2008GL036644$05.00

of equation (5), which depends on the measures only via the summation of the logs, $\sum_{i=1}^{n} \ln(\tau_i)$.

$$f(m|\tau_1,\tau_2,\ldots,\tau_n) = \frac{e^{\left(2\mu_{\ln(\tau)}\left(\sum_{i=1}^{n}\ln(\tau_i)\right)-n\mu^2_{\ln(\tau)}\right)\Big/2\sigma^2_{\ln(\tau)}} e^{-\beta m}}{\int_{M_{\min}}^{M_{\max}} e^{\left(2\mu_{\ln(\tau)}\left(\sum_{i=1}^{n}\ln(\tau_i)\right)-n\mu^2_{\ln(\tau)}\right)\Big/2\sigma^2_{\ln(\tau)}} e^{-\beta m} dm} \tag{5}$$

[8] A simpler approach to the estimation of the magnitude is proposed by *Convertito et al.* [2008]. It consists of using for M, in place of its full PDF, a classical point estimate that, for example, can be obtained via the constrained maximum likelihood estimator of equation (6).

$$\bar{M} = \begin{cases} \hat{M} = 5.9 + \dfrac{7}{n}\sum_{i=1}^{n}\log(\tau_i); & if \quad M_{\min} \le \hat{M} \le M_{\max} \\ M_{\min}; & if \quad \hat{M} < M_{\min} \\ M_{\max}; & if \quad \hat{M} > M_{\max} \end{cases} \tag{6}$$

[9] Regarding the source-to-site distance, because of rapid earthquake localization procedures, a probabilistic estimate of the epicenter may also be available. For example, during an earthquake the RTLoc algorithm [*Satriano et al.*, 2008] allows the assigning to each point of a grid, defined in the region where the network operates, a probability that the hypocenter is coincident with that point based on the sequence according to which the stations trigger, $\{s_1, s_2, \ldots, s_n\}$. Consequently, also the PDF of R, $f(r|s_1, s_2, \ldots, s_n)$, may be retrieved numerically in real-time. Thus, it is possible to compute the probabilistic distribution (or hazard curve) of a ground motion intensity measure (IM), for example the PGA, at a site of interest as in equation (7), which requires $f(im|m, r)$ available for the chosen IM. This is typically a lognormal distribution the parameters of which are given by a ground motion prediction equation.

[10] The subscript n in the left-hand side of equation (7) means that the computed hazard curve is still conditional on the specific set of triggered stations/measures and, therefore, evolves with time. This procedure will be referred to as real-time probabilistic seismic hazard analysis (RTPSHA).

$$f_n(im) = \int_M \int_R f(im|m,r) f(m|\tau_1,\tau_2,\ldots,\tau_n) f(r|s_1,s_2,\ldots,s_n) dr dm \tag{7}$$

[11] One of the main purposes of this letter is to assess and compare the effects of different possible approaches to magnitude estimation on RTPSHA. In addition, the other sources of uncertainty (i.e., localization and GMPE) are also evaluated. Finally, based on the results of these analysis, information-dependent lead-time maps are given for the Campania region (southern Italy).

2. Simulation of the ISNet EEWS

[12] ISNet is a local network of 29 strong motion seismic stations deployed in a 100 × 70 km^2 area covering the epicentral location of the main earthquakes that occurred in the southern Appennines in Italy [*Weber et al.*, 2007]. The EEWS under development in Campania using ISNet has been designed to be a hybrid application of regional and on-site early warning approaches [*Kanamori*, 2005], where a seismic network operates to protect several critical structures/infrastructures at the same time.

[13] To analyze the performances of the ISNet infrastructure in respect to the RTPSHA a MATHWORKS-MATLAB/SIMULINK® simulator was developed. It may be represented by the three following blocks.

2.1. Acquisition

[14] The simulation starts with the assignment of the basic features of the earthquake (i.e., the *true* values of magnitude and location). Then, the τ measurements for the triggered stations are needed to perform the RTPSHA. The stations' measurements may be simulated sampling from the $f(\tau|m)$ distribution (see section 1). In the simulation process, the number and sequence of stations triggered is also computed assuming an appropriate velocity model (to follow) for the region. A station is considered to be in the simulation process (i.e., its τ measurement is simulated) if 4 s have passed from its trigger.

2.2. Computation

[15] This block furnishes evolving real-time estimates of the hypocenter (i.e., the distance of the source from the site for which the prediction of peak ground motion is sought) and the magnitude of the earthquake as described in section 1. At each 1 s time step, the estimates of R and M are updated on the basis of new information collected by the network, when available.

2.3. Decision

[16] Here the hazard integral of equation (7) is computed, i.e., the prediction of the IM at the target site is obtained and the decision whether to issue an alarm or not is taken. This implies a decisional rule: a simple one, if the predicted IM is the PGA, is to issue the alarm if the probability that a critical peak ground motion value (PGA_c) will be exceeded, when the earthquakes strikes at the site, is larger than a probability threshold (P_c), as in equation (8).

$$P[PGA > PGA_c] > P_c \tag{8}$$

[17] Probabilistic procedures that can be used to set PGA_c and P_c based on the minimization of the expected loss for structures, are provided by *Iervolino et al.* [2007].

3. Results and Discussion

[18] As an example, results of the simulation for an M 6 event located within the ISNet network are given. Two target sites are considered, the city of Naples, capital of the Campania region, and S. Angelo dei Lombardi, a small town severely affected by the 1980 Irpinia earthquake (Figure 1a). The probabilistic PDFs of the magnitude (obtained via the Bayesian approach) are given in Figure 1b. Figure 1c and 1d represent the distributions of R for the two sites. Note that the true distances of Naples and S. Angelo dei Lombardi from the epicenter of the simulated event, 124 km and 60 km respectively, are well captured by the PDFs. In Figure 1e

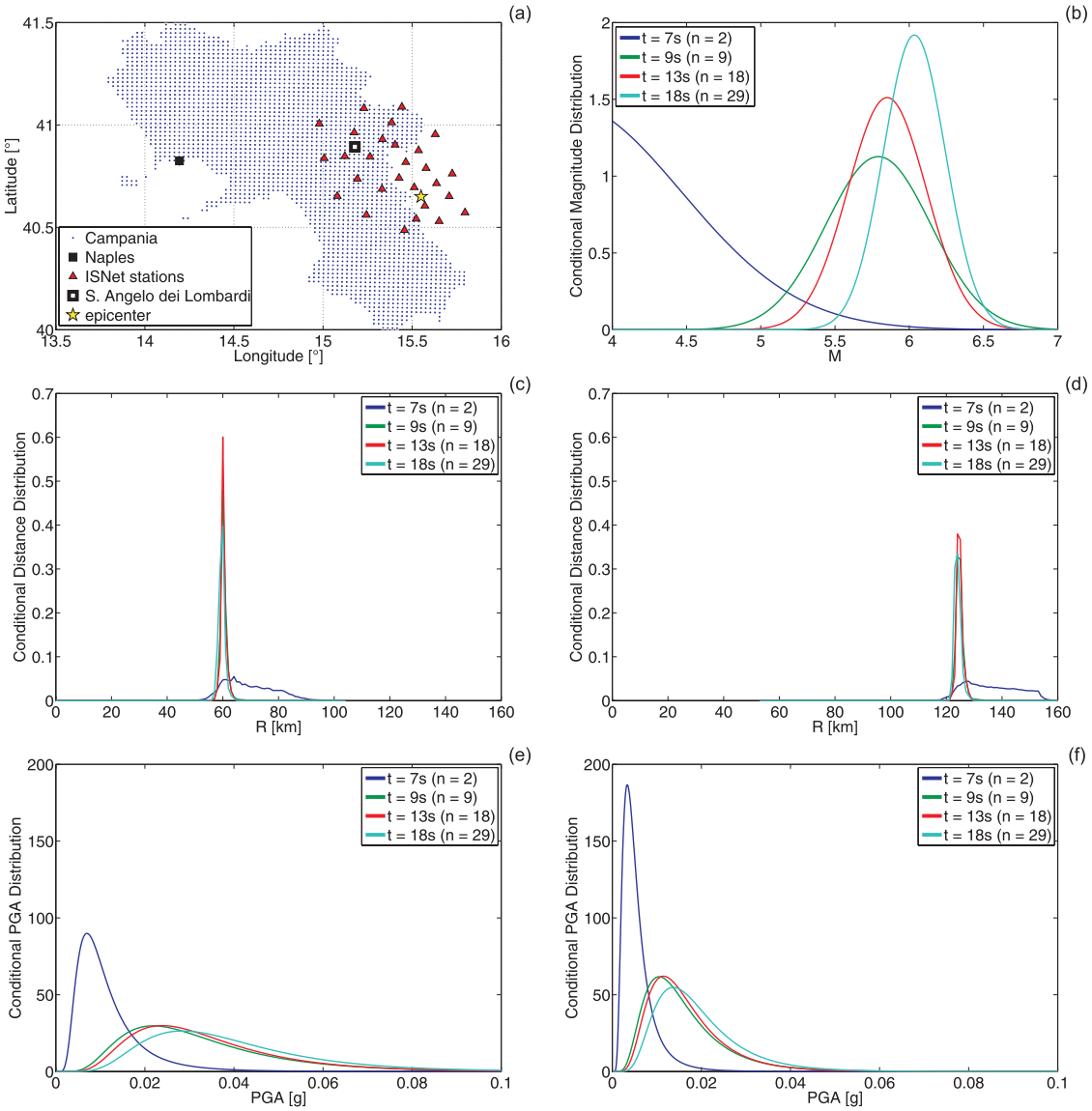

Figure 1. Results for the M 6 simulated event in (left) S. Angelo dei Lombardi and (right) Naples. (a) Scheme of the network and sites; (b) magnitude distributions; (c and d) distance distributions; and (e and f) PGA distributions.

and 1f, the real-time hazards are given for the two target sites (the used GMPE is that from *Sabetta and Pugliese* [1996], assumed, in this study, to be applicable moderately beyond the M and R limits prescribed for the model). The curves in Figure 1 are given for selected instants from the beginning of the event and, therefore, different numbers of stations providing τ.

3.1. Uncertainty Analysis

[19] Referring to the M 6 event and the decisional rule discussed, in Figures 2a and 2b the exceedance probabilities are given for different possible values of PGA_c. The plots in the figure refer to 100 averaged simulations (i.e., the τ measurements were simulated 100 times for the same earthquake). It appears that the probability of exceedance does not change after 10 s–13 s. In other words, after on average 11–18 stations have measured τ, the estimation of the critical PGA do not benefit much from further information.

[20] In Figures 2c and 2d, the coefficient of variation (CoV, the ratio of the standard deviation to the mean) of the PGA is given, as the number of stations providing τ increases. In particular, the CoV is computed, using equation (7), in the following cases: (i) considering both PDFs of M and R; (ii) considering only the modal value of the distance (R*) from Figures 1c and 1d, in place of its full PDF; and (iii) considering only the mode of R and the maximum likelihood value of M. Note that in case (iii) the real-time hazard is simply given by $f(im|\overline{M}, R^*)$, i.e., the distribution of the PGA from the GMPE computed for the specific $\{\overline{M}, R^*\}$ pair.

[21] It appears from the results that the uncertainty of the distance is negligible with respect to the prediction of PGA as the CoV is almost the same with or without uncertainty on distance as green and blue curves are overlapping (it is to note that, for events outside the network or for networks not as dense as the ISNet, cases not investigated herein, the uncertainty on distance can be larger). Also the contribution

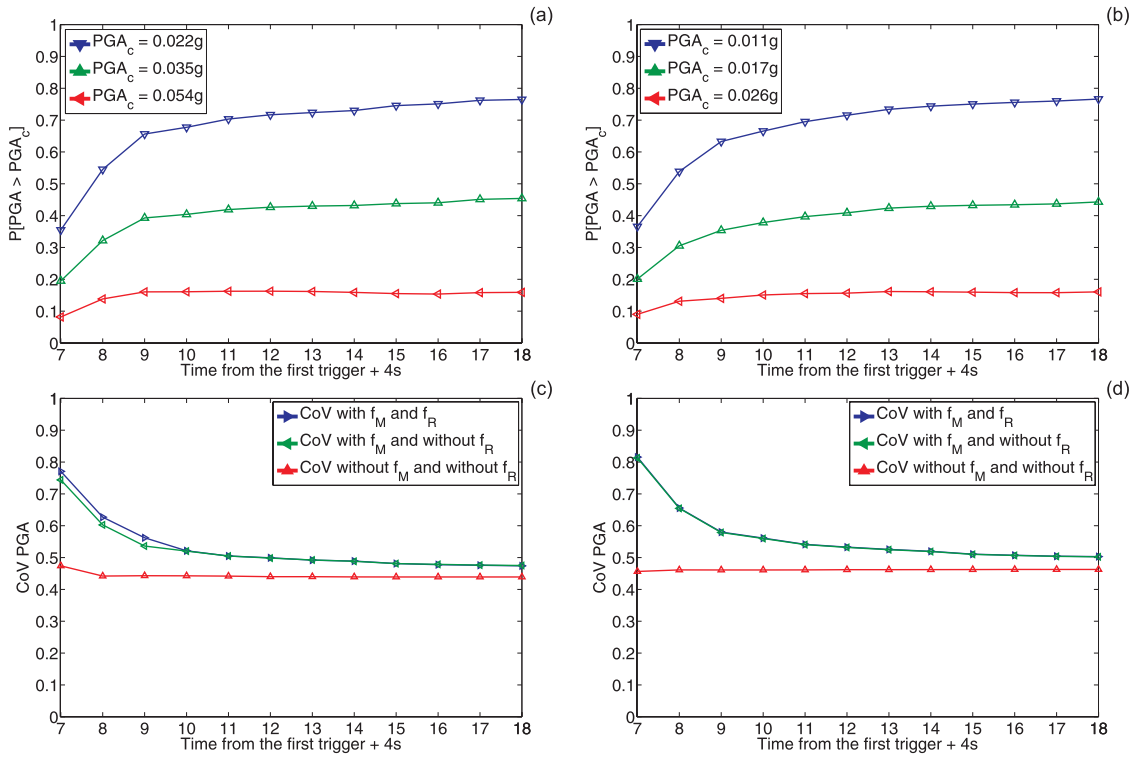

Figure 2. Evolution of hazard estimates for three different critical PGA values and COVs in the case of different considered uncertainties. (left) S. Angelo dei Lombardi and (right) Naples.

of uncertainty of magnitude to CoV, although larger than distance, is small if compared to that of the GMPE. This is more evident when, during an event, several τ measures have been collected by the seismic network.

3.2. Information-Dependent Lead-Time

[22] It was shown that, in the adopted RTPSHA approach, the estimation of critical ground motion becomes stable only after a number of stations have measured the early signal of the event. Therefore, there is a trade-off between the lead-time and the level of information based on which the alarm issuance is decided. Consequently, different lead-times may be considered, each of those correspond to a different number (k) of stations providing τ, for example 4, 18 and 29. In these cases, the lead-times associated to each point (j) in Campania, T_k^j, have been computed as in equation (9) using a 1D model in which P and S velocities, V_P and V_S respectively, are a function of hypocentral depth (h) and have a constant $V_P/V_S = 1.68$ ratio.

$$T_k^j = T_S^j(h) - T_P^k(h) - \Delta t \tag{9}$$

[23] In equation (9), $T_S^j(h)$ is the S-wave arrival time at the specific site (j); $T_P^k(h)$ represent the time to trigger the first k stations; Δt is the required processing time (assumed that it is equal to 5 seconds including the required P-wave data recording time at each station and some computing time). $T_S^j(h)$ and $T_P^k(h)$ were approximated using travel-time curves as a function of distance from the seismic source and h (A. Zollo, personal communication, 2008).

[24] The lead-times were computed for the region considering as possible epicenters those randomly occurring in the area which includes the ISNet sensors, while the hypocentral depth may be up to 12 km. Results in form of maps are given in Figure 3, where mean lead-times are given for each node of a regular grid having about 2 km spacing and covering the territory of the Campania region. The map corresponding to 18 stations was analyzed in respect to a list of real-time risk reduction actions and the time they require [*Goltz*, 2002]. Because it was shown that the information on the impending ground motion stabilizes when about 18 stations have provided τ, this map may be a tool for the design of engineering applications of EEWS.

3.3. Comparison of Magnitude Estimators

[25] In this section the performances of the magnitude estimators presented in section 1 are compared. The better estimator is assumed to be the one giving hazard estimates closer to *true* one (i.e., the hazard calculated using the real values of magnitude and distance for the event). Since an estimator never performs systematically better than another one, except in trivial cases, comparisons are typically made in terms of average performance, adopting as reference index the mean squared error (MSE). Because this index is calculated for fixed values of the estimated parameter, the MSE results as a function of the parameter. Unfortunately, for different estimators the functions may interweave, not allowing the identification of the better estimator. In these cases, when an a priori distribution is available for the unknown parameter, a resolving approach consists in computing the Bayes risk (BR). It is obtained by averaging the MSE over the a priori. The better estimator is the one having the smaller BR.

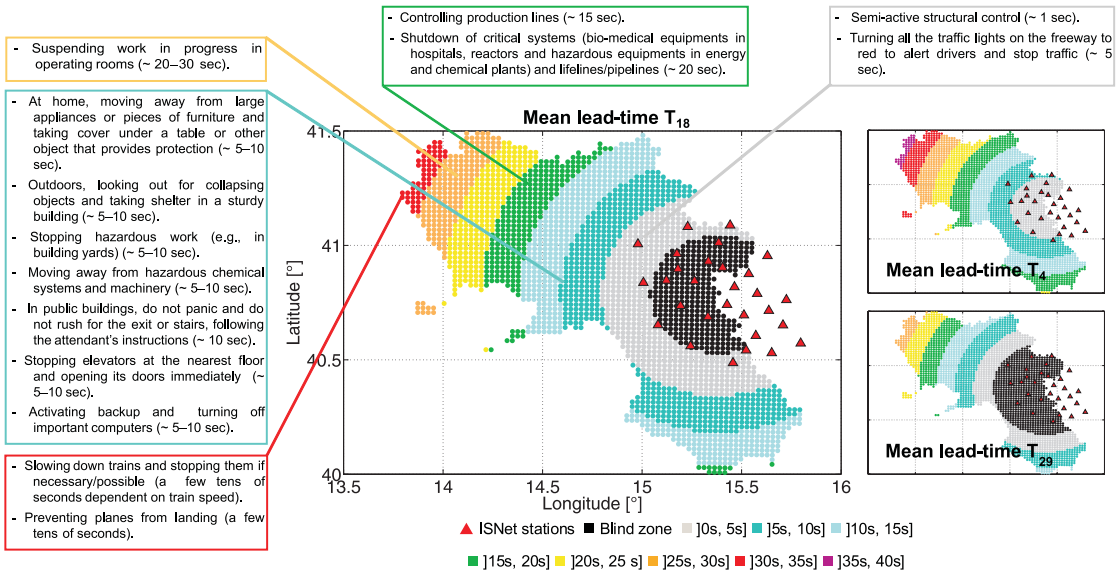

Figure 3. Information-dependent average lead-time maps for the Campania region in the case of an event within the area covered by the ISNet network and possible risk reduction actions. The three maps refer to different numbers of τ measures available.

[26] Herein, the comparison of magnitude estimation approaches is carried out by an exemplificative case study, in which MSE and BR were calculated via a Montecarlo simulation. It should be noted preliminarily that in the RTPSHA framework there is a function to evaluate (i.e., the PGA distribution). Thus, for each value of the magnitude the MSE is still a function of the PGA. The site for which the hazard is computed is Sant'Angelo dei Lombardi and the epicentral location is that of Figure 1. In all the simulations the distance was considered known, because of its negligible uncertainty.

[27] Results of the analysis are given in Figures 4a–4c, which report the MSE in six cases. The curves give for each PGA in abscissa the MSE computed as in equation (10) where

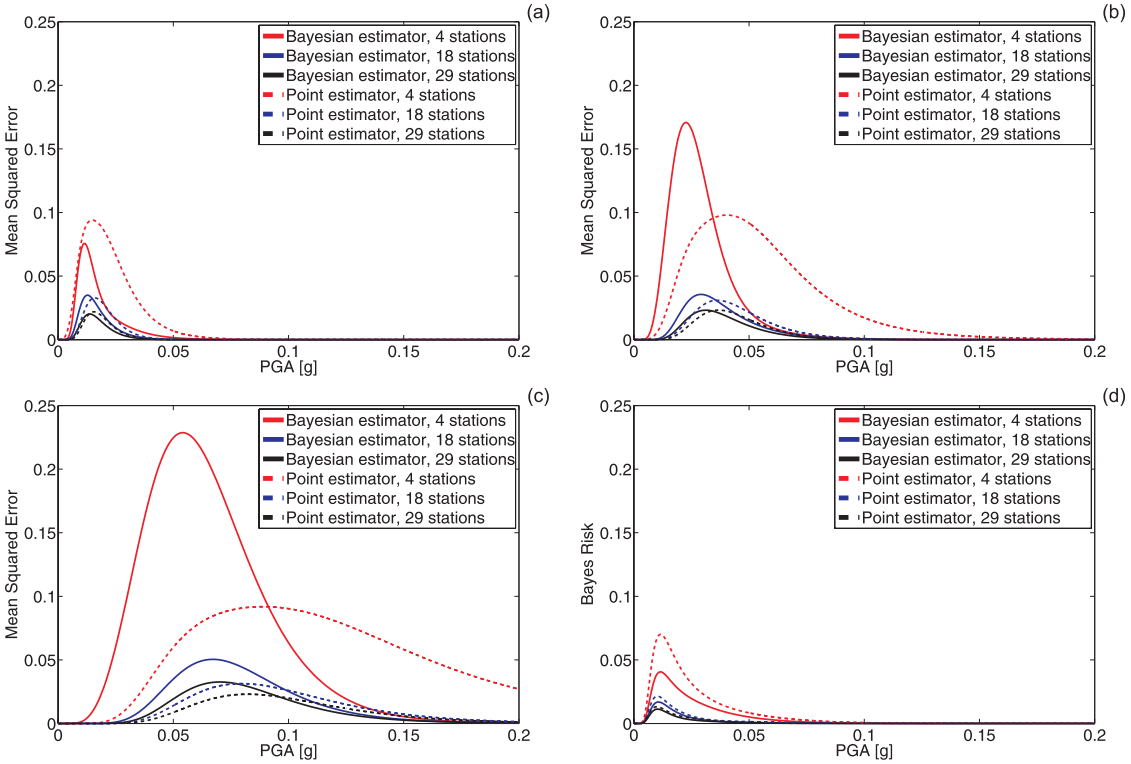

Figure 4. Effects on real-time hazard of different magnitude estimators in the case of different numbers of τ measures available. MSE for (a) M 5, (b) M 6, and (c) M 7 events; and (d) Bayes risk.

NS is the number of samples generated in the simulation, $P_E[PGA]_T$ is the true hazard (expressed in terms of probability of exceedance of PGA), and $P_E[PGA]_i$ is its estimate obtained with the i-th sample.

$$MSE = \frac{1}{NS}\sum_{i=1}^{NS}\left(P_E[PGA]_i - P_E[PGA]_T\right)^2 \quad (10)$$

[28] Dashed lines in the figure refer to the classical point estimator. Three different dashed lines are represented, to show how performances depend on the number of stations which have triggered at the moment of the analysis; solid lines are their analogous computed with the Bayesian approach. The curves were computed simulating NS = 500 samples/events of fixed magnitude, M 5, M 6, and M 7 respectively.

[29] The MSE, in all cases, does not establish which estimator performs better, as the result of the comparison depends on both M and PGA. Thus, with an a priori being available for the magnitude (i.e., the Gutenberg-Richter), the BR was computed, again adopting equation (10). In this case 500 samples/events, whose magnitude (one for each event) was randomly sampled from the a priori, were generated. Therefore, the true hazard value also varies, with M, from sample to sample. Note that because the a priori is the historical recurrence relationship of M, BR has the objective meaning of long term estimation error of RTPSHA.

[30] Results of the simulation (Figure 4d) demonstrate that the Bayesian estimator, performs better than the alternative one. In fact, the BR for the former is smaller for every value of the PGA with respect to the latter, with differences which, as expected, decrease when the number of triggered stations increases, as larger amounts of measures render the a priori information less precious.

4. Conclusions

[31] The uncertainty associated to engineering ground motion predictions in the framework of earthquake early warning was investigated. The study was carried out referring to the ISNet seismic network, installed in southern Appennines (Italy). Results of the analyses lead to conclude that: (1) the dominating uncertainty in the real-time hazard analysis is that of the ground motion prediction equation; (2) the estimation of peak ground motion stabilizes only when a certain level of information is reached and, therefore, information-dependent lead-time should be considered for design of engineering applications of EEW; (3) Bayesian magnitude estimation may provide larger efficiency in respect to simpler point estimates if the peak ground motion is the parameter of interest.

[32] **Acknowledgments.** This research was funded by AMRA scarl (http://www.amracenter.com/), through the SAFER project (http://www.saferproject.net/) of the EU's Sixth Framework Programme, contract 036935. Authors want also to acknowledge Claudio Satriano for the RTLoc data regarding the simulated event.

References

Allen, R. M., and H. Kanamori (2003), The potential for earthquake early warning in southern California, *Science*, *300*, 786–789.

Convertito, V., I. Iervolino, M. Giorgio, G. Manfredi, and A. Zollo (2008), Prediction of response spectra via real-time earthquake measurements, *Soil Dyn. Earthquake Eng.*, *28*, 492–505.

Goltz, J. D. (2002), Introducing earthquake early warning in California: A summary of social science and public policy issues, report to OES and the operational areas, Governor's Off. for Emergency Serv., Pasadena, Calif. (Available at http://www.cisn.org/docs/Goltz.TaskI-IV. Report.doc)

Iervolino, I., V. Convertito, M. Giorgio, G. Manfredi, and A. Zollo (2006), Real-time risk analysis for hybrid earthquake early warning systems, *J. Earthquake Eng.*, *10*, 867–885.

Iervolino, I., M. Giorgio, and G. Manfredi (2007), Expected loss-based alarm threshold set for earthquake early warning systems, *Earthquake Eng. Struct. Dyn.*, *36*, 1151–1168.

Kanamori, H. (2005), Real-time seismology and earthquake damage mitigation, *Annu. Rev. Earth Planet. Sci.*, *33*, 195–214.

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

Satriano, C., A. Lomax, and A. Zollo (2008), Real-time evolutionary earthquake location for seismic early warning, *Bull. Seismol. Soc. Am.*, *98*, 1482–1494.

Weber, E., et al. (2007), An advanced seismic network in the southern Apennines (Italy) for seismicity investigations and experimentation with earthquake early warning, *Seismol. Res. Lett.*, *78*, 622–634.

C. Galasso, I. Iervolino, and G. Manfredi, Dipartimento di Ingegneria Strutturale, Università degli Studi di Napoli Federico II, via Claudio 21, I-80125 Naples, Italy. (iunio.iervolino@unina.it)

M. Giorgio, Dipartimento di Ingegneria Aerospaziale e Meccanica, Seconda Università degli Studi di Napoli, via Roma 29, I-81031 Aversa, Italy.