Flash sourcing, or rapid detection and characterization of earthquake effects through website traffic analysis

Rémy Bossu[*], Sébastien Gilles, Gilles Mazet-Roux, Fréderic Roussel, Laurent Frobert, Linus Kamb

European Mediterranean Seismological Centre (EMSC), earthquake information website (www.emsc-csem.org)

ABSTRACT

This study presents the latest developments of an approach called 'flash sourcing', which provides information on the effects of an earthquake within minutes of its occurrence. Information is derived from an analysis of the website traffic surges of the European–Mediterranean Seismological Centre website after felt earthquakes. These surges are caused by eyewitnesses to a felt earthquake, who are the first who are informed of, and hence the first concerned by, an earthquake occurrence. Flash sourcing maps the felt area, and at least in some circumstances, the regions affected by severe damage or network disruption. We illustrate how the flash-sourced information improves and speeds up the delivery of public earthquake information, and beyond seismology, we consider what it can teach us about public responses when experiencing an earthquake. Future developments should improve the description of the earthquake effects and potentially contribute to the improvement of the efficiency of earthquake responses by filling the information gap after the occurrence of an earthquake.

Introduction

Rapid characterization of earthquake effects is essential for timely and appropriate responses to aid victims. In the immediate aftermath of damaging earthquakes, any field observations of their effects can fill the information gap and contribute to more efficient rescue operations. A small magnitude shallow earthquake below an urban area can be widely felt and become a major event from a public and media perspective. A rapid assessment of their impacts helps seismologists dealing with public information to timely respond to questions about these concerns.

This study presents the latest developments of a method called 'flash sourcing', which addresses these issues [Bossu et al. 2007, 2008, 2011]. Flash sourcing relies on eyewitnesses, the first people who are informed of, and hence the first who are concerned by, an earthquake occurrence. More precisely, the use by these eyewitnesses of the European–Mediterranean Seismological Centre (EMSC) earthquake information website (www.emsc-csem.org) is analyzed in real time to map the area where an earthquake is felt, and to identify, at least under certain circumstances, the zones of widespread damage. This approach is based on the natural and immediate responses of eyewitnesses, who rush to the Internet to investigate the cause of shaking that they have just felt, converging on the EMSC website and thus increasing the website traffic [Bossu et al. 2007]. The area where an earthquake has been felt is mapped simply by locating the Internet Protocol (IP) addresses of the visitors to the EMSC website during these website traffic surges [Bossu et al. 2008]. In addition, the presence of eyewitnesses browsing our website within minutes of an earthquake occurrence is evidence excluding the possibility of widespread damage in the localities they originate from; in the case of severe damage, the networks would probably be down [Bossu et al. 2008]. The validity of the information derived from this website traffic analysis is confirmed by comparisons with macroseismic maps of the European Macroseismic Scale EMS98 [Grünthal 1998] obtained from online questionnaires [Bossu et al. 2011].

The name of this approach, flash-sourcing, is a combination of 'flash crowd' and 'crowd-sourcing', which reflects the rapid collection of data from the public. For computer scientists, a flash crowd indicates a traffic surge on a website. Crowd-sourcing refers to work being done by a 'crowd' of people, and it is also used to characterize Internet and mobile applications that collect information from the public, such as online macroseismic questionnaires. Like crowd-sourcing techniques, flash sourcing is a crowd-to-agency system, although unlike crowd-sourcing techniques, flash sourcing is not based on declarative information (e.g. answers to a questionnaire), but on implicit data. In our case, this is the real-time analysis of the website traffic observed at the EMSC website.

In the first part of the report, we present the main improvements of the method, the improved detection of
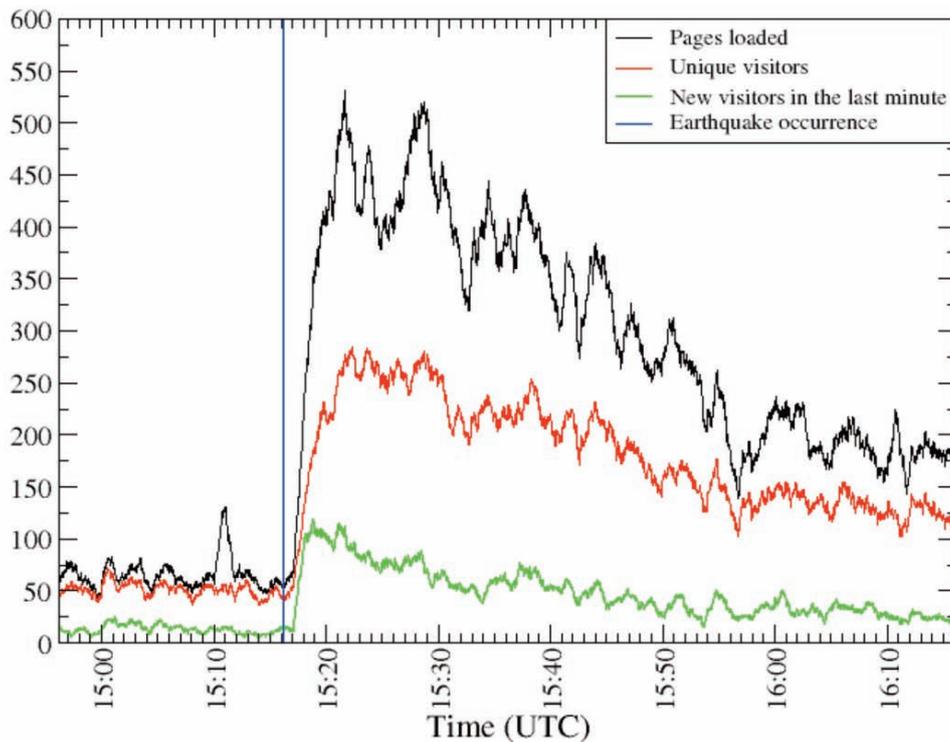
Figure 1. The different traffic metrics for the M 4.5 Vrancea (Romania) earthquake of June 6, 2010. The number of loaded pages (black line) characterizes the activity that results from all of the website visitors. The number of unique website visitors (red line) is the number of different IP addresses that loaded at least one page in the previous minute. The number of new website visitors in the previous minute (green line) is a new metric defined to reliably detect the convergence of eyewitnesses following a felt earthquake. HTTP requests from robots, referrers and pre-identified seismological institutes are systematically removed as the first step of these website traffic measurements. The vertical blue line represents the origin time of the earthquake. The first location and magnitude for this earthquake were reported on the EMSC Web site 4 min after its occurrence, while the website traffic surge was detected in 112 s.

website traffic surges, and a way to instantly map areas affected by severe damage or network disruption. The second part describes how the information derived can improve and speed up the delivery of public earthquake information, and beyond seismology, what it can teach us about public behavior when an earthquake is experienced. Finally, the discussion focuses on future developments and how flash sourcing can ultimately improve earthquake response.

Detection of website traffic surges resulting from felt earthquakes

A website traffic surge following a felt earthquake is a result of the rapid influx of site visits by eyewitnesses who were not present on the website before the event occurrence. Rather than relying on the classical metrics of website traffic, such as the rate of loaded pages or number of visitors, which can be estimated through the number of IP addresses, we compute the number of new visitors who have arrived in the last minute. This is defined as the number of different IP addresses that have loaded at least one page within the last minute, and that have had no activity (i.e. no loaded pages) on the website in the previous 30 min (Figure 1). These arbitrarily defined time values are aimed at detecting fast increases in website traffic. The metric has little sensitivity for the level of baseline traffic or for more gradual increases

observed every morning or as a result of the propagation of the news of an earthquake.

As the first step, the raw website traffic is systematically filtered to exclude identified indexing robots, IP addresses from seismological institutes [Bossu et al. 2008], and visitors coming from HTTP referrers; i.e. visitors linking from an external website. The exclusion of this last component of the website traffic avoids false triggers caused by links to the EMSC website that can be published on popular forums or websites [Bossu et al. 2011]. There is currently no discrimination between mobile and land-line Internet access (see *Future developments* section for more details), and all visitors to the website are treated equally.

Real time monitoring of the number of new visitors in the last minute allows the automatic detection of felt earthquakes generating a significant traffic increase. This monitoring is performed every second in real time. The latency due to the different steps of the computation is about one second, including for removal of referrers and robots, and determination of the number of new visitors in the last minute. EMSC-specific trigger criteria have been defined following an analysis of our website traffic characteristics, to reduce false triggers rather than to optimize the number or rapidity of the automatic detections. The system triggers when the number of new visitors in the last minute exceeds the observed average over the previous 30
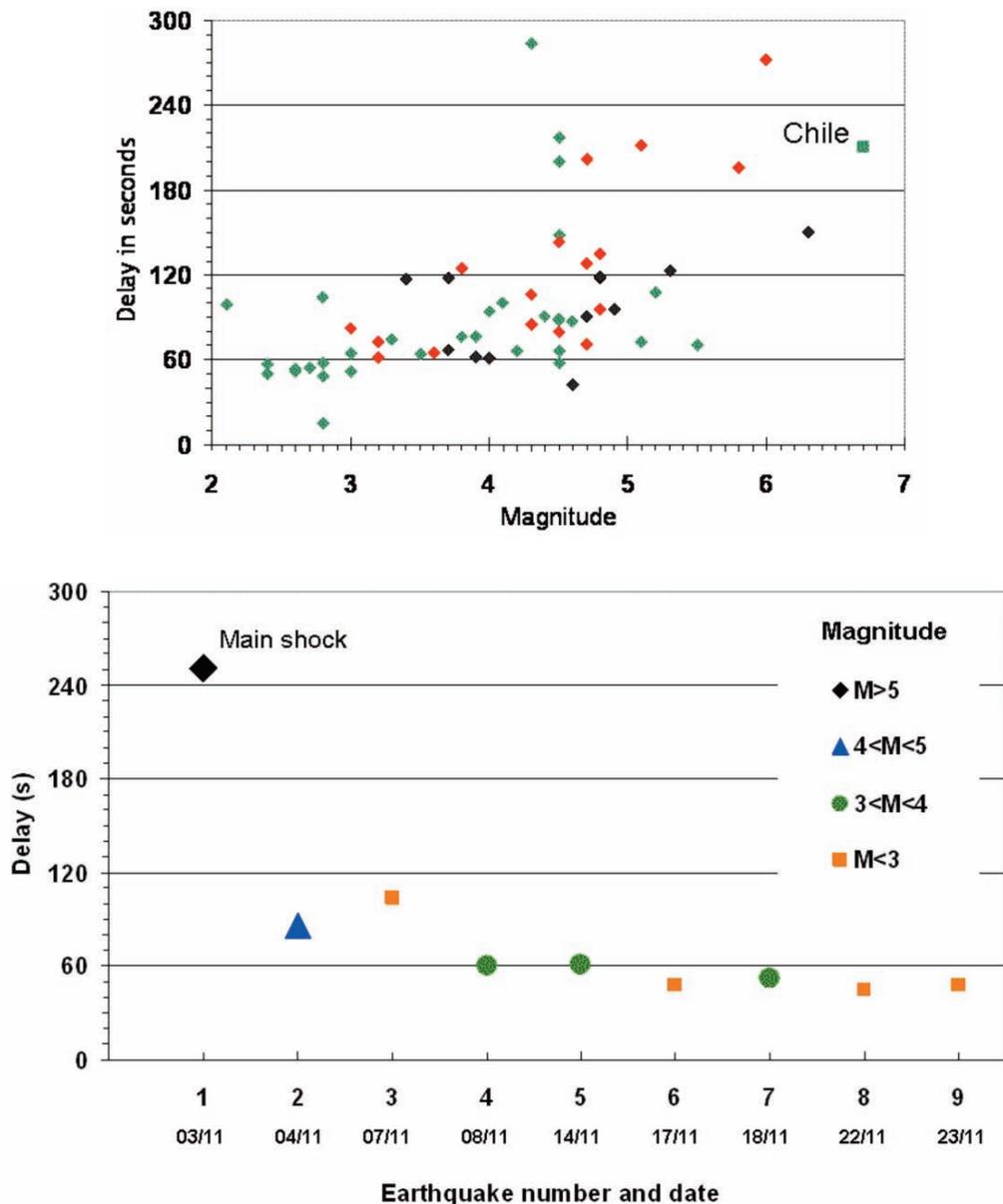
Figure 2 (top). Repartition of the time delays of the 63 automatically detected felt earthquakes as a function of their magnitude for the period January 2009 to May 2011, with black symbols for 2009 events, green for 2010, and red for 2011. The delay is measured from the earthquake occurrence. The Chilean earthquake, an aftershock of the M 8.8 earthquake on February 27, 2010, is the only earthquake detected outside the European–Mediterranean, region where the majority of the EMSC website visitors come from. **Figure 3** (bottom). Evolution of the detection delay over the 20 days following the M 5.4 earthquake in Kraljevo (Serbia). Nine earthquakes were automatically detected during this period. The detection delay is not related to the earthquake magnitude but decreases with time. The audience of the EMSC website dramatically increased in the epicenter region in the hours following the main shock, easing the detection of the following aftershocks regardless of their magnitudes.

min by more than 40. These felt events are automatically detected with delays ranging from 15 s to 284 s of their occurrence, with a median value of 88 s (Figure 2). Only one of the 63 detected events in the period from January 2009 to May 2011 was outside the European–Mediterranean region; namely, a M 6.7 earthquake in Chile. This illustrates that such a system can only detect earthquakes felt by the normal users of the website [Bossu et al. 2008], and that the EMSC users originate mainly from the European–Mediterranean region.

The magnitudes of the automatically detected earthquakes range from M 2.1 to M 6.7. Some smaller magnitude earthquakes can potentially generate a website response. However, they do not meet the current trigger criteria defined to optimize the reliability of the EMSC felt-earthquake detection. Smaller shocks appear to be detected faster than larger events (Figure 2). In practice, the key factor is not the magnitude, but the order of the shock in a sequence of earthquakes, with aftershocks detected faster than a mainshock. For example, on November 3, 2010, a M 5.4 earthquake occurred at Kraljevo (Serbia), and it was detected in 250 s. The next day, a M 4.4 aftershock was detected in 80 s, and the majority of the subsequent shocks were detected within less than 1 min, irrespective of the magnitudes (Figure 3). The explanation is probably the following: the first shock enlarges the visibility of the EMSC website in the
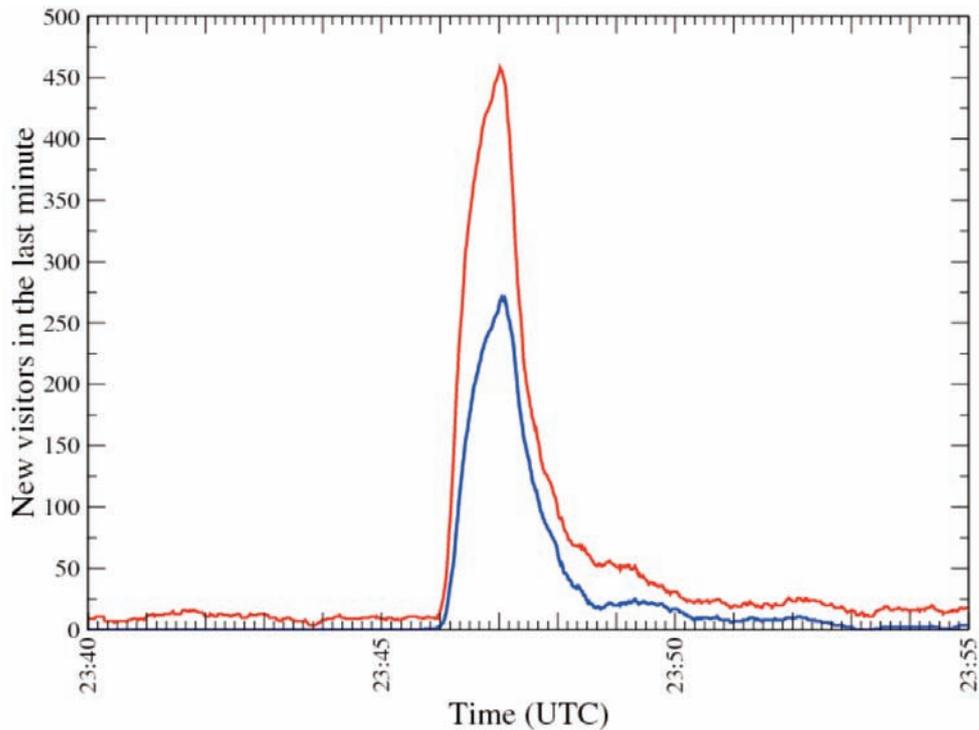
Figure 4. The mention of the EMSC website during a television program in Spain on March 21, 2010, led to a website traffic surge that was not related to an earthquake. The blue curve represents the website visitors originating from the Spanish Google search engine. They represent about 50% of the website visitors, which is a significantly higher fraction that what was observed during a website surge as a result of an earthquake, which is typically less than 30%.

epicenter area, making the subsequent triggers much more rapid due to the more rapid convergence when new shocks occur of eyewitnesses who might have already bookmarked the URL of the site [Bossu et al. 2011]. This would also explain the detection of the M 6.7 Chilean earthquake on March 16, 2010, which was an aftershock of the M 8.8 Chilean earthquake of February 27, 2010.

Television programs that mention the EMSC website can create a dramatic surge in the website traffic and result in a false trigger; i.e. a trigger that is not associated with a felt earthquake. This has been the cause of the 2 false triggers in the last 29 months. With 63 correct triggers in this period, this gives a false trigger rate of 3%. The first case of television-generated website traffic occurred in May, 2009, following a program on a Romanian television channel [Bossu et al. 2011]. In another case, a night-time scientific program on a Spanish television channel, on March 21, 2010, caused an immediate and massive increase in the website traffic (Figure 4) when the EMSC URL was displayed (E. Carreno, private communication). In both cases, the traffic originated from the Google search engine is significantly larger than during true triggers: 70% for the television program in Romania, 50% for the television program in Spain compared to less than 30% for surges associated to an earthquake. We believe that television programs drive on our site primarily persons who have never visited it before, increasing the use of search engines. If this is confirmed, this ratio could help in the automatic discrimination of television-generated triggers.

Theoretically, a group of well-organized Internet users can also generate false triggers by launching concurrent website visits. This has not been observed yet. Similarly, we have also never observed coordinated false answers to our online macroseismic questionnaires. This suggests that if such events cannot be excluded, at least their occurrence should remain rare.

Mapping the felt area

The EMSC uses the Digital Element IP Intelligence software to identify the geographic location of each website visitor through the IP address. Once a website surge is detected, the area where the earthquake was felt is mapped by the geographic origins of the statistically significant increased website traffic [Bossu et al. 2008, 2011]. This map is obtained within a couple of minutes of the website surge detection, and it is updated after the origin time and location of the earthquake have been determined by the seismic networks. Maps can also be produced for earthquakes that have not generated a detected website traffic surge, and therefore have not triggered the system.

In practice, the number of observed visitors at each IP location since the origin time of an earthquake is compared with the average website traffic from the same place, during the same days of the week, and at the same hour of the day, during the previous 4 weeks. Increases are considered significant when they reach a 99% confidence level. Locations showing significantly increased website traffic delineate the felt area (Figure 5a).

Figure 5. (**a**) Geographical origins of the website visitors to the EMSC website within 5 min of the occurrence of the M 5.4 earthquake (star) in Serbia on November 3, 2010. Red circles represent geographical origins of statistically significant increased traffic observed during this 5-min time window. Yellow circles represent the geographical origins of website traffic with no significant variations. Black triangles represent the observed geographic origins of website visitors in past 12 months. This characterizes the regional audience of the EMSC website as well as the maximum spatial resolution of the IP locations. Circle size is a function of the difference between the expected and the observed numbers of unique IPs, with the smallest circle equal to 0 and the largest greater than or equal to 100. The star represents the epicenter location. IP locations are determined by Digital Element. There is an absence of visitors originating from the epicenter area, where damage was observed (epicenter intensity in Kraljevo: EMS98 VII from online questionnaires collected by the EMSC). (**b**) Macroseismic maps based on online questionnaires collected by the USGS for the same earthquake. The area where the earthquake was reportedly felt is comparable with the map derived from flash-sourcing analysis in (a). The datasets used to derive these two maps were independently collected by two distinct agencies.
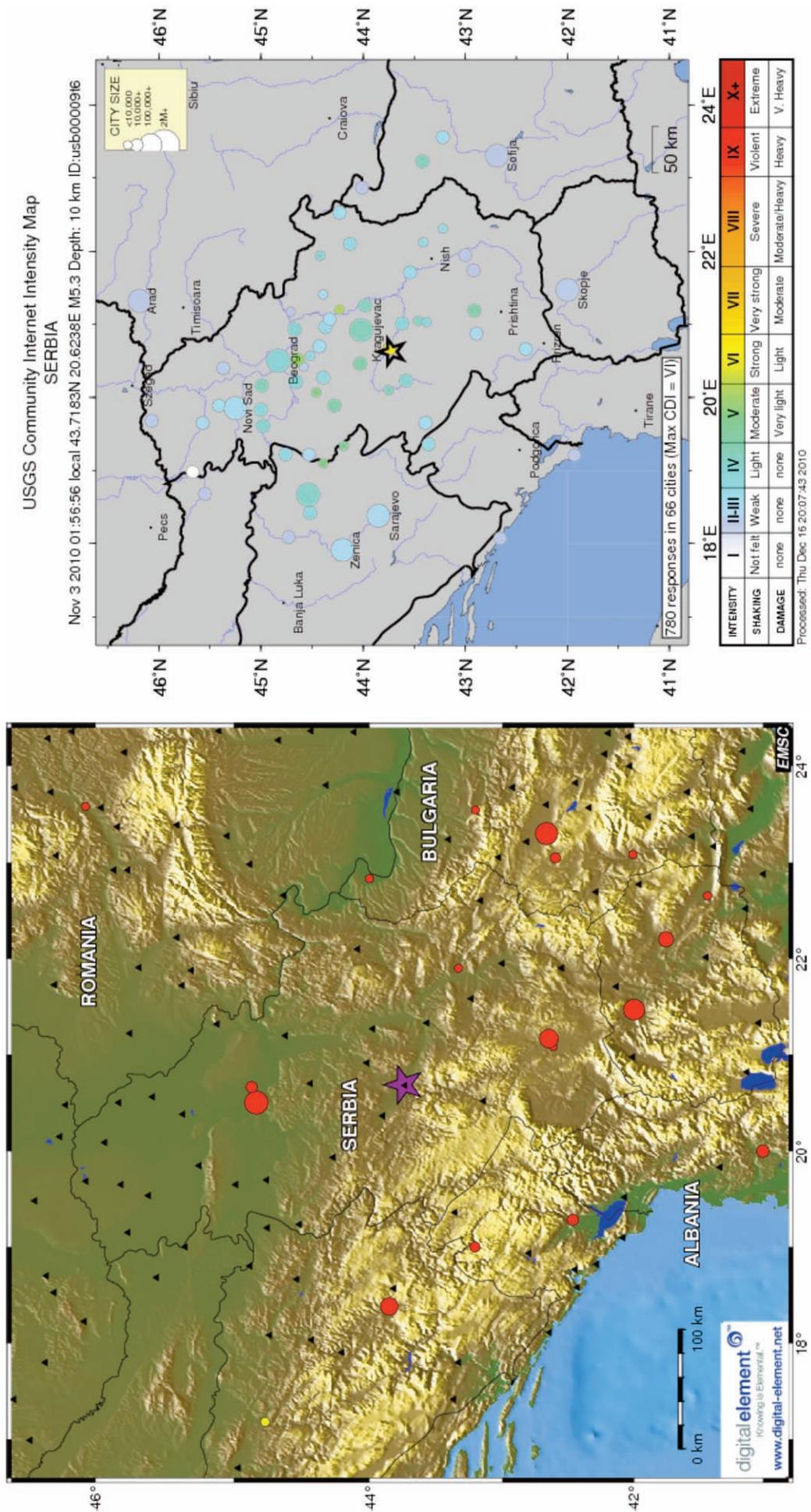
The static maps of the felt area (Figure 5a) use a 5-min time window. A time-sequenced map from 0 to 5 min after an earthquake occurrence and with a 5-s time resolution is provided as an electronic supplement. The reduction in the time window from 10 [Bossu et al. 2011] to 5 min since 2009 has been made possible by the increased visibility and website traffic on the EMSC website. A shorter time window reduces the possibility that the information about the earthquake occurrence has spread beyond the felt area through telephone calls or social networks. Assuming that the information has already spread through these channels, and so beyond eyewitnesses, this is very unlikely to affect the felt area. First of all, because the number of these visitors should be large enough (after the filtering of referrers) to generate a statistically significant increase in the website traffic from the locations they originate from. Even if they are in sufficient numbers in a few locations, they will not be spatially correlated with the epicenter location. A live television broadcast during the shaking caused by an earthquake would, however, irremediably affect the mapping of the felt area, by attracting visitors to the EMSC website from the broadcasting sphere of the program.

An example of a map of a felt area is shown in Figure 5a, for the M 5.4 Kraljevo (Serbia) earthquake mentioned above. The overall felt area compares well with the macroseismic maps that were independently produced by the US Geological Survey (USGS; Figure 5a, b). The comparison with the EMSC macroseismic maps might be influenced by the intrinsic overlap between the eyewitnesses who have filled in the questionnaires and those who have caused the website traffic surge.

Another website traffic increase was observed to originate from Tirana, Albania (Figure 5a). In this case, no questionnaires were collected in this country either by the USGS (Figure 5b) or the EMSC. The possibility that the shaking had been felt in northern Albania is consistent with the macroseismic pattern going south to the epicenter (Figure 5b). Although we have not been able to find evidence to confirm this possibility, the absence of collected questionnaires might be due to the low number of respondents; Bossu et al. [2011] showed from tests on several earthquakes that on average, less than 1% of the eyewitnesses visiting the EMSC website actually complete the online questionnaire.

Detection and mapping of earthquake damage

A striking feature illustrated by Figure 5a is the absence of website visitors who originated from the epicenter area, where the shaking was strong and only light damage occurred. The same phenomenon was observed for the l'Aquila (Italy) M 6.3 earthquake on April 6, 2009 [Bossu et al. 2011]. We believe that this lack of website visitors during strong shaking might result from the eyewitnesses initially being more concerned about personal safety, and/or from

network disruption. However, while the presence of website visitors excludes the possibility of severe damage, with network collapse, an absence of website visitors is only an indication of the possibility of damage, and is not on its own concrete evidence of damage.

To improve the identification of damaged areas, we developed a system based on the fact that in cases of severe damage, the existing website sessions that originate from an affected area will end instantaneously and simultaneously, due to power and/or network failures. The AJAX-based tracking system detects the arrival and departure of each unique visitor to the EMSC website using a periodic client 'ping' to track visitor presence. Nominally, every 30 s, the browser of a website visitor makes an HTTP request to the EMSC servers, which updates the tracking information. The arrival of a new website visitor creates a new record in the tracking system, which is subsequently updated each time their browser pings the server while the website visitor remains on the EMSC website. Departures from the EMSC website are determined by the detection of tracking records that have not been updated within the last 30 s.

A massive power failure on March 14, 2010, that plunged part of Chile into darkness was easily detected by the sudden increase in the number of sessions that originated from Chile that ended at 23:43 GMT (Figure 6). This is the exact timing for this power failure that is reported on Wikipedia (http://es.wikipedia.org/wiki/Apag%C3%B3n_de_Chile_de_2010). According to the same source, in Santiago the power failure affected 6 of the 52 districts, which represents approximately 27% of the city population. The number of sessions that originated from Santiago dropped by a similar proportion (31%), from 147 to 101 (Figure 6). These similar ratios might be coincidental, however, as the majority of the sessions that originated from Santiago at that time remained active, this shows that it was not a total blackout for this city. When applied to an earthquake, this approach can provide vital information on earthquake damage within 30 s of its occurrence.

Flash sourcing for faster public earthquake information

In April 2011, and under the username @LastQuake, the EMSC set up an automatic service for Twitter, the well-known, real-time, information-sharing website, to distribute the flash-sourced information. The first Twitter message, or 'tweet', is published immediately after the automatic detection of a website traffic surge. This indicates the regions where an earthquake appears to have been felt (Figure 7). The names of the regions follow the Flinn–Engdahl naming convention [Young et al. 1996], and they are determined by the regions where the locations with traffic increases are.

Once the causative earthquake has been located, it is automatically associated with the website surge; a second
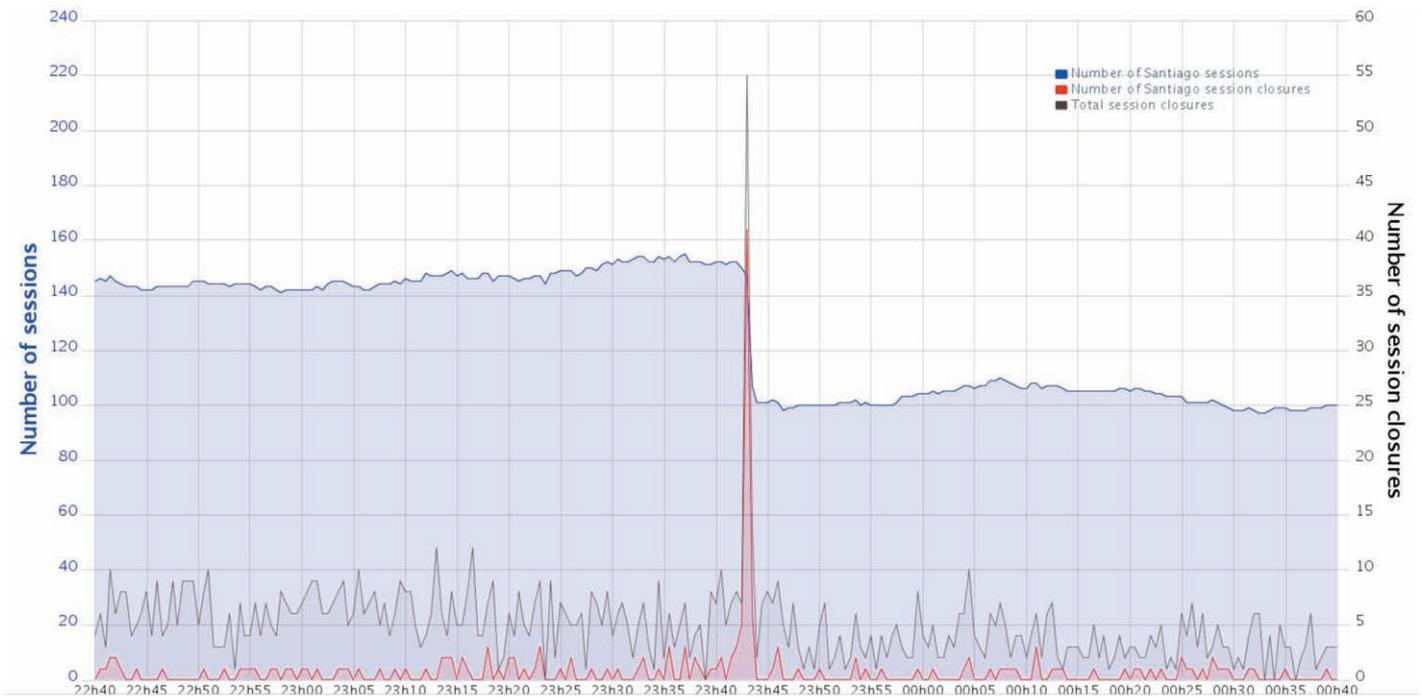
Figure 6. Change over time in the number of sessions (blue) and the number of session closures observed on the EMSC website for a 30 s window on March 14, 2010. The sudden increase in the number of closures (black and red) at 23:43 UTC was due to a massive power failure which affected Chile. At the time of the blackout, the number of sessions originating from Santiago (blue) dropped by 31% which is similar to the estimation of 27% of the affected population in the city.

Figure 7. Example of tweets published by the EMSC under the username @LastQuake for the M 5.8 earthquake in western Turkey on May 19, 2011. The first tweet (bottom) was automatically published once the website traffic surge was detected 3 min and 25 s after the earthquake occurrence. The origin of the website visitors identified western Turkey as the area where the shock had been felt. The first location and magnitude were published 6 min and 44 s after the earthquake, with a preliminary magnitude of M 5.9. This triggered the publication of the second tweet (top). The same information is published in a scrolling banner at the same time on the EMSC website, pointed to by the shortened URL in the tweet. In the first tweet, the 2 min is estimated from the previous trigger, which had a median time of 88 s (Figure 2).

tweet, that indicates the magnitude and origin time of the event, is then published (Figure 7). The same information is published in parallel on the EMSC website as a scrolling banner, with a second notice replacing the first. Earthquakes causing a website traffic surge are automatically labeled as felt in the earthquake list published on the EMSC website.

The services derived from the detection of website traffic surges primarily improves the speed at which the earthquake information is made available to the public via the EMSC website and Twitter. The time benefit is defined as the delay between the automatic detection of the website traffic surge and the publication of the first earthquake location. In a few cases, this can be null, if an earthquake is located before the website surge is detected. The median value is about 2 min (Figure 8). This can be much larger (Figure 8), especially for small magnitude events, which are often reported less rapidly by networks than larger shocks. In the vast majority of cases (95% in the period of January, 2009 to May, 2011; Figure 8), the automatic detection of website

traffic surges speeds up the information for the eyewitnesses at a time when they are actively looking for information on the shaking that they have just felt.

Information on public behavior

Website traffic analysis provides some insight into both the event that has caused a website traffic surge and the public behavior when facing this event. The geographic origin of the television-generated website traffic in Spain (Figure 4) depicts the national extension of the broadcasting sphere of the program, with no website traffic increase from Portugal or France (Figure 9). The rise time was 1 min, which indicates that the majority of the website visitors shifted from their television to their browser very rapidly. Browsing the Internet while watching the television might also be quite typical, at least for some sections of the population. The 5 min width of the website surge (Figure 4) is much shorter than the width when the website surge is caused by an earthquake (e.g. Figure 1), which is on the order
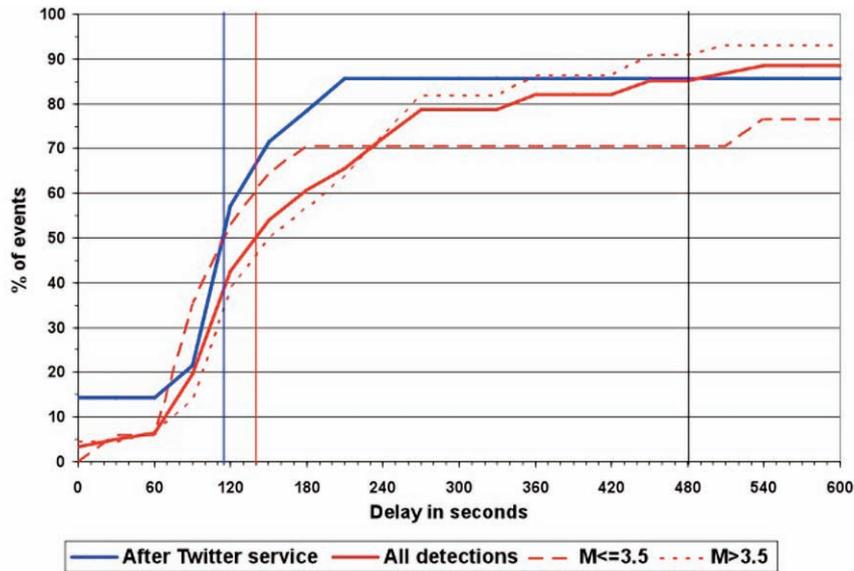
Figure 8. Distribution of the time benefit of using flash-sourcing for earthquake information. The time benefit is defined as the delay between the detection of the felt earthquake through the website traffic surge it generates and the publication of the first preliminary location and magnitude on the website. The time benefit is null for earthquakes that were first reported by the seismological networks. The blue line is the cumulative distribution since the initiation of the Twitter service, and the red line is for the period January 2009 to May 2011. The median time benefit is around 2 min. This can be much longer, especially for small magnitude earthquakes: it exceeds 8 minutes (480 s) for 30% of earthquake below M 3.5, as compared to 10% for larger shocks.
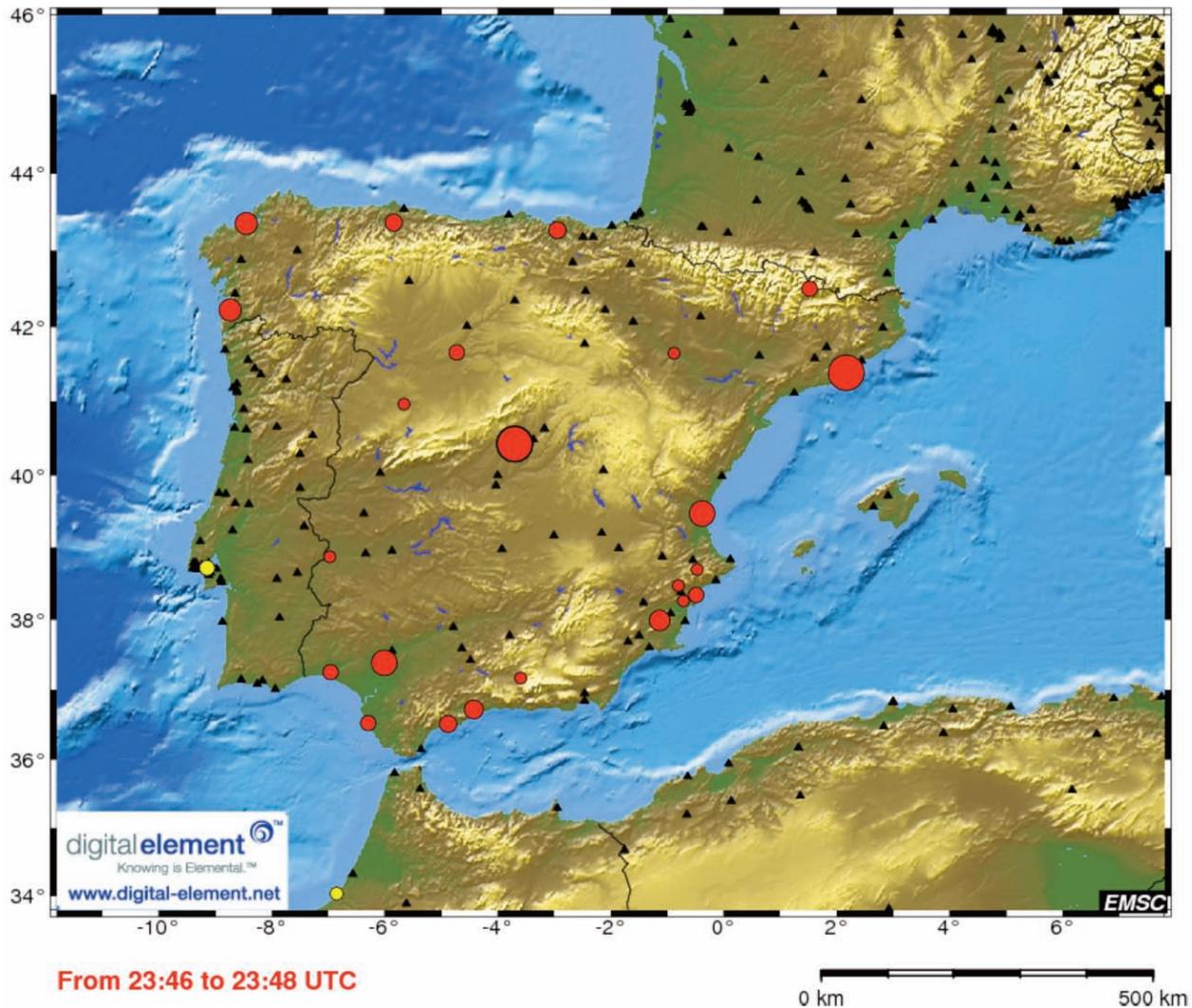

From 23:46 to 23:48 UTC

Figure 9. The television program in Spain that led to a website traffic surge (see Figure 4) generated a significant increase in the number of Spanish visitors of the EMSC website from all over Spain (see Figure 5a for caption).

of 90 min [Bossu et al. 2011]. We believe that the website viewers who did not switch immediately to the EMSC website simply did not visit it following the program.

On January 22, 2010, two earthquakes of M 5.3 and M 4.8 occurred within 3 min and 38 s of each other in the region of Patras, Greece (Figure 10). In agreement with the other Greek networks, the bulletin of Patras University indicates an inter-epicenter distance of less than 3 km for these two shocks, and a similar focal depth. As the two shocks were co-located, and with the first event being the stronger one, eyewitnesses who felt the second earthquake are then very likely to have also felt the first one. Surprisingly, the two shocks are clearly visible on the website traffic curve (Figure 10). Therefore, some of the eyewitnesses who visited the EMSC website only visited it after the second shock. We believe that some of the website visitors might not have consider it worth getting out of the bed after the first shock, but were then sufficiently troubled by the succession of two shocks to finally check what was going on. If this hypothesis is correct, this is probably a more common behavior in areas where felt earthquakes are frequent, than in regions of low seismic hazard.

These two earthquakes occurred in the middle of the night (first shock, 02:46 local time). Nonetheless, the first eyewitnesses reached the EMSC website within 50 s of this first earthquake (Figure 10). For comparison, this delay was of the order of 8 min for the first studied earthquake in 2004 [Bossu et al. 2011]. The increasing Internet penetration rate among the public, as well as the rapid rise in the use of smart phones and other connected devices, are the most likely explanations of these ever-shrinking reaction times.

Future developments

The automatic detection of felt earthquakes is based on significant website traffic surges. Therefore, the number of eyewitnesses who converge to the EMSC website has to be significant compared to the total website traffic. With a doubling of the website traffic every year [Bossu et al. 2011], the number of eyewitnesses required to trigger the automatic detection also needs to double every year, which will impede the detectability threshold of the system. The analysis of the website traffic at different geographic levels, such as per country, region, or city, can dramatically improve the detection performance. This improves the signal-to-noise ratio for more rapid detection (e.g. Figure 10), and can isolate several felt earthquakes that occur only a few minutes apart. On May 1, 2011, a M 4.7 earthquake in Romania was followed 7 min later by a M 5.4 earthquake in Eastern Kazakhstan. The first shock was detected, although the website traffic surge caused by the second shock remained unnoticeable in the global website traffic curve. A simple after-the-fact website traffic filtering at the country level
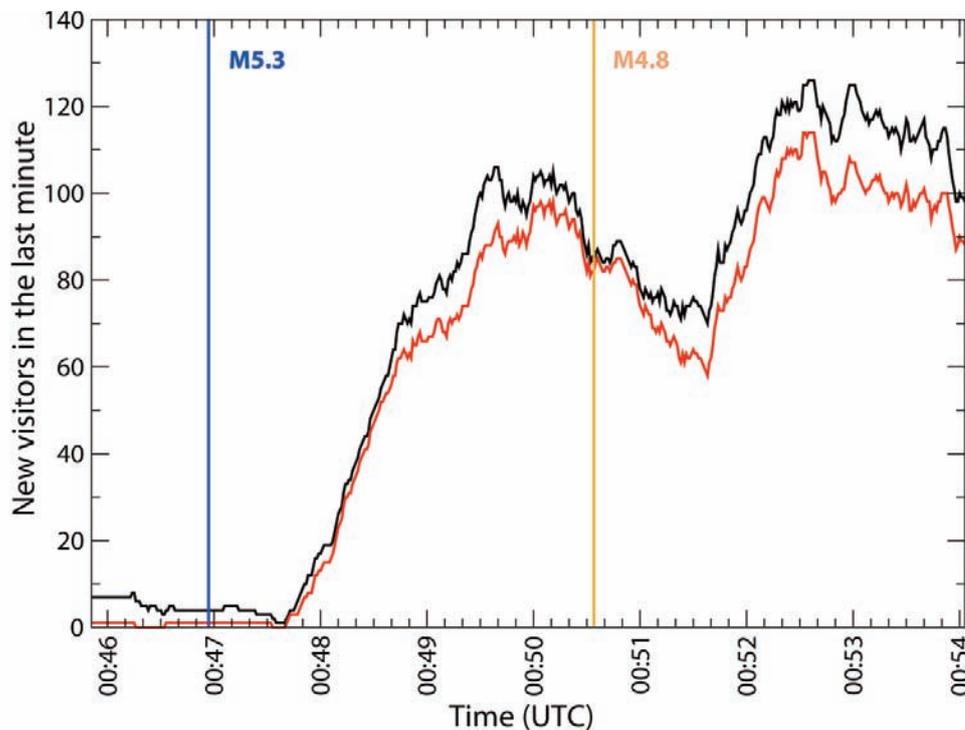

Figure 10. Website traffic surge as a result of two co-located earthquakes of M 5.3 and M 4.8 that occurred 3 min and 30 s apart in the Patras region (Greece) on January 22, 2010. The black curve shows the number of new website visitors in the last minute; the red curve shows the new website visitors originating from Greece, i.e. the country where the event was felt. While both shocks are clearly visible, the automatic detection of the second shock by seismological networks proved tricky, as its seismic waves were mixed with the waves of the mainshock. These shocks happened at about 03:00 local time, and the increase in the number of visitors from Greece is clearly visible within less than 50 s of the earthquake occurrence. Geographic filtering (red curve) improves the signal-to-noise ratio and can lead to faster detection of website traffic surges.
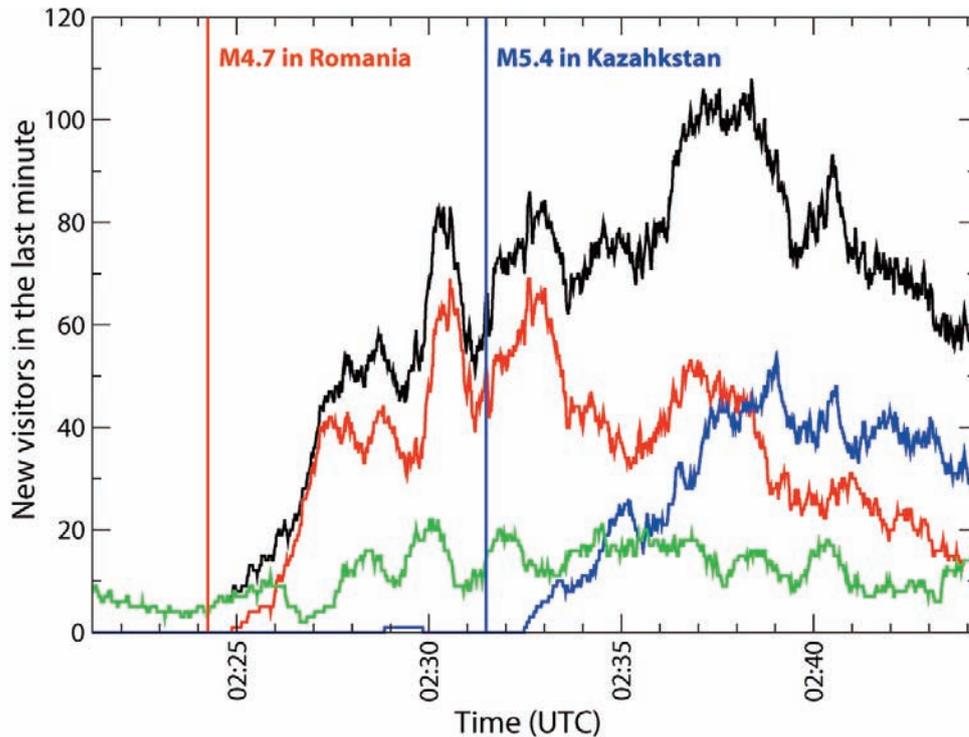
Figure 11. Superposition of website traffic surges caused by two distant felt earthquakes. On May 1, 2011, a M 4.7 earthquake in Romania was followed 7 min later by a M 5.4 earthquake in eastern Kazakhstan. The observed website traffic (black line) is the addition of the website traffic originating from Romania, Bulgaria and Moldova, the three countries where the first shock was felt, according to the EMSC macroseismic map (red line), and the one originating from Kazakhstan, and the baseline website traffic (green line). This demonstrates that the traffic can be broken down in its distinct components by geographic filtering.

demonstrates that both of these earthquakes could have been equally detected (Figure 11).

The ratio of the number of eyewitnesses visiting the EMSC website to the regional population might be an indication of the relative strength of the shaking. It might be expected that for a given earthquake, the number of eyewitnesses normalized by the number of inhabitants will be lower for a city where only a fraction of the population felt the ground shaking, when compared to another city where the event was widely felt. The exclusive use of mobile Internet access, which can be discriminated through the user-agent of the HTTP requests, might also characterize areas where the citizens have rushed out of the buildings in panic, leaving land-line access unused. When all sessions using land-line access close down at the same time, it can be inferred that the electric grid might be down. The full potential of analyzing mobile *versus* land-line access remains to be evaluated. This will require more data, which should be available in the coming year on the basis of the rapid rise in smart-phone use in the European–Mediterranean region.

In summary, the joint analysis of the absence, the partial or total loss of sessions at the time of an earthquake, the increase of the number of visitors in relation to the local number of inhabitants, and the type of Internet access is expected to deliver at least in some cases a more progressive classification of local earthquake effects from 'not felt' to 'heavy damage'.

Discussion

Flash sourcing provides automatic detection of widely felt earthquakes within an average of 2 min of their occurrence. Earle et al. [2010] reported similar performances through monitoring the number of Twitter messages that contain keywords such as "earthquake" or "quakes". These two examples demonstrate that social networks can offer fast detection of rapid-onset events. Flash sourcing can currently map the area where an earthquake has been felt in less than 5 min. It can exclude the possibility of widespread damage, and in certain cases it can detect and map regions affected by network failures in near real time. This makes flash sourcing the fastest tool for the gathering of information on earthquake damage.

For comparison, the very first online questionnaire that related to the Kraljevo earthquake (Figure 5) was collected within 7 min at the USGS (Quitoriano and Wald, personal communication), and within 9 min at the EMSC. Twenty minutes after the earthquake, the USGS had collected 50 questionnaires and the EMSC had collected 3. As the website traffic analysis is based on implicit data, as opposed to declarative data such as questionnaires, it is complementary to crowd-sourcing approaches, such as online macroseismic questionnaires [e.g. Wald et. al. 1999], or collection of geo-located pictures of earthquake damage [Bossu et al. 2011], each of which provide different constraints on the actual earthquake impact in a different time frame. If successful, the planned improvements should reduce the trigger time to

below a 1 min threshold and characterize up to 7 different levels of effects: not felt, weak, moderate or strong shaking, panic, light or heavy damage.

Ultimately, we aim to integrate flash-sourced information, answers to online macroseismic questionnaires, and collected geo-located pictures of earthquake damage into a single and time-evolving map of earthquake effects. This will allow us to fill the information gap in a time window ranging from the first minutes to the first few hours following the occurrence of an earthquake. Reliability assessment will be based on consistency between the different types of collected information, as well as on consistency with ground-motion prediction derived from earthquake parameters.

Flash sourcing is cheap and preserves the privacy of the website visitors. No individual behavior is identified or tracked. An analogy would be a road traffic study that records at the geographic origin of the drivers through their license plates. The uniqueness and personal nature of the plate number is thus disregarded, by considering only the state indication of the plate.

Flash sourcing has one strict constraint: it can only be implemented on a website where eyewitnesses naturally converge after an event. They do not visit the EMSC website to fill in a questionnaire or to share their pictures, but rather to find the information they are looking for about their felt event. This implies that the website must be a reference information point for the population for the type of events considered. This approach can potentially be extended to other rapid-onset phenomena, such as flash floods [Bossu and Walker 2009], as long as the time for the eyewitnesses to reach the website is much shorter than the spread of information about the causal event through the media. Although the first website traffic surge that resulted from an earthquake was reported for the 1999 Hector Mine event [Wald and Schwartz 2000], we are not aware of other studies that are aimed at characterizing an event through the analysis of the flash-crowd characteristics that it generates.

Seismologists also benefit from flash-sourcing information. In the case of the two earthquakes in Patras (Figure 10), many real-time processing tools implemented in the seismological networks that contribute to the EMSC real-time services [Mazet-Roux et al. 2011] failed to detect the second shock, as its waves were mixed with those from the first event. Flash-sourcing can be a heads-up to initiate interactive processing in such tricky cases.

The benefit for the public is first more rapid information about earthquakes that they consider to be significant, regardless the magnitude and the use of alternative communication channels, such as Twitter. This shifts the traditional policy of delivering earthquake information when available to delivering this information when requested. The amplitude of a website traffic surge is a direct measurement

of the public desire for information. A large public information request that is ignored or simply not promptly answered by the seismologists might alter the public trust and the scientific credibility in the operating organization, and hence turn eyewitnesses to less-reliable websites; an official agency or an authoritative organization just cannot afford to be the last one informed when a significant event strikes.

Flash sourcing sheds light on the public reactions following an earthquake. It opens ways to evaluate the influence of culture and of the level of seismic hazard on the way people react during an earthquake. Website traffic surges identify clear teachable moments. They can trigger timely geographically targeted online information on seismic hazard and prevention measures, to complement the existing public-information tools.

Flash sourcing is part of an overall scientific evolution that promotes the engagement of the citizens and the development of community-based environmental monitoring and information systems. In seismology, this started with online macroseismic questionnaires [e.g., Wald et al. 1999]. The EMSC has been collecting geo-located pictures of earthquake effects on its website since 2006 [Bossu et al. 2011]. More recently, in cases of damaging earthquakes, the EMSC makes available online the RICHTER (Rapid geo-Images for Collaborative Help Targeting Earthquake Response) application for Android mobile phones, which was developed by AnsuR Technologies and which allows earthquake eyewitnesses to easily and freely share their pictures with the EMSC. The Quake Catcher Network [Cochran et al. 2009] was the first citizen-operated seismic network; based on cheap micro-electromechanical systems sensors, this is well suited for the deployment of dense networks [Cochran et al. 2009]. As mentioned above, the real-time analysis of the use of social networks, such as Twitter, offers rapid detection of felt earthquakes [Earle et al. 2010]. Such initiatives (and the others contained in this special issue) can change the way the seismological community interacts with the public, from simple recipients of information to a more active role. The understanding of the public demands and motivations, and clear communication on project benefits and outcomes are a prerequisite to citizen engagement. In turn, improved public communication can contribute to more efficient risk awareness programs and develop public ownership of risk reduction policies.

The benefit for the scientific community is a cheap and complementary approach to document earthquake-related phenomena that can contribute to both scientific knowledge and improved earthquake response. In the longer term, rapid two-way exchange of information between the scientific community and earthquake eyewitnesses through Internet and mobile applications or social networks will open the way

to resident-to-resident assistance in the critical first hours of a damaging earthquake. Desirable or not, such a possibility raises many complex issues, especially on liability aspects. These cannot be addressed by the scientific community alone, but it will have to contribute to these discussions.

Conclusions

Flash sourcing is the fastest tool for the gathering of information on earthquake damage. Information is derived from the analysis of the use of the EMSC website by the earthquake eyewitnesses, who are the first informed and the first concerned by an earthquake occurrence. Flash sourcing offers an inexpensive alternative to dense real-time accelerometric networks, the implementation of which is limited due to high operational costs, and it complements crowd-sourcing techniques, such as online questionnaires. Flash-sourcing improves public earthquake communication by taking into account the public desire for information, rather than by focusing solely on the magnitude of an earthquake.

If successful, future developments should offer a more detailed picture of earthquake damage, which in turn can contribute to filling in the information gap in the immediate aftermath of an earthquake, and thereby contributing to improved earthquake response.

References

Bossu, R., V. Douet, S. Godey, G. Mazet-Roux and S. Rives (2007). On the use of Internet to rapidly collect earthquake impact information, EMSC Newsletter 22, 31-34.

Bossu, R., G. Mazet-Roux, V. Douet, S. Rives, S. Marin and M. Aupetit (2008). Internet users as seismic sensors for improved earthquake response, EOS, Transactions, 89, 225-226.

Bossu, R. and A. Walker (2009). New tools for alerts and capturing immediate effects of rapid-onset events across Europe, Chemical Hazards and Poisons report, Health Protection Agency, 14, 35-37.

Bossu, R., S. Gilles, G. Mazet-Roux. and F. Roussel (2011). Citizen Seismology or How to Involve the Public in Earthquake Response, In: D.M. Miller and J. Rivera (eds.), Comparative Emergency Management: Examining Global and Regional Responses to Disasters, Auerbach/Taylor and Francis Publishers, 237-259.

Cochran, E.S., J.F. Lawrence, C. Christensen and R.S. Jakka (2009). The Quake-Catcher Network: citizen science ex-

panding seismic horizons, Seismol. Res. Lett., 80, 26-30.

Earle, P., M. Guy, R. Buckmaster, C. Ostrum, S. Horvath and A. Vaughan (2010). OMG Earthquake! Can Twitter improve earthquake response?, Seismol. Res. Lett., 81, 246-251.

Grünthal, G., ed. (1998). European Macroseismic Scale 1998 (EMS-98), Cahiers du Centre Européen de Géodynamique et de Séismologie 15, Centre Européen de Géodynamique et de Séismologie, Luxembourg, 99 pp.

Mazet-Roux, G., R. Bossu, E. Carreno and J. Guilbert (2011). Report on 2010 real-time activities, EMSC report, 38 pp.; URL: http://www.emsc-csem.org/Doc/EMSC_DOCS/EMSC_RT_activities_2010.pdf (last accessed June 23, 2011).

Wald, D.J., V. Quitoriano, L. Dengler and J.W. Dewey (1999). Utilization of the Internet for Rapid Community Intensity Maps, Seismol. Res. Lett. 70, 87-102.

Wald, L.A. and S. Schwartz (2000). The 1999 Southern California Seismic Network Bulletin, Seismol. Res. Lett., 71, 401-422.

Wessel, P. and W.H.F. Smith (1998), New, improved version of the Generic Mapping Tools Released, EOS Trans. AGU, 79, 579.

Wikipedia report: Apagón de Chile de 2010 (in Spanish); URL: http://es.wikipedia.org/wiki/Apag%C3%B3n_de_Chile_de_2010 (last accessed June 23, 2011).

Young, J.B., B.W. Presgraveb, H. Aichelec, D.A. Wiensd and E.A. Flinn (1996). The Flinn–Engdahl Regionalisation Scheme: The 1995 revision, Phys. Earth Planet. Interiors, 96, 221-300.

---

*Corresponding author: Rémy Bossu,
European Mediterranean Seismological Centre (EMSC),
c/o CEA, Bât. Sâbles, Centre DAM - Île de France, Bruyères le Châtel, France; email: bossu@emsc-csem.org.