Deliverable

D5.8

Near real time estimate of parameters of significant European earthquakes based on web

Deliverable information	
Work package	[D5.8 WP5]
Lead	[EMSC]
Authors	[Julien Roch, EMSC; Robert Steed, EMSC; Remy Bossu, EMSC]
Reviewers	[Name, Institution]
Approval	[Management Board]
Status	[Draft/final]
Dissemination level	[Public/internal]
Delivery deadline	[31.08.2022]
Submission date	[30.06.2021]
Intranet path	[DOCUMENTS/DELIVERABLES/File Name]

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821115

Table of contents

Summary

Rapid public earthquake information is essential for both the public and authorities and can contribute to a more efficient earthquake response. CsLoc (Crowdseeded Location) is a global service for fast and reliable location of felt earthquakes combining crowdsourced detections and P-wave arrival picks data. In the Rise framework, we increase the number of seismic stations and added the magnitude computation even though the latter one was not planned. As CsLoc performs in the first 5 minutes of an origin time, it needs an extremely dense network around the crowdsourced detections to locate the earthquake. Increasing the number of seismic stations allow CsLoc to locate lower magnitude earthquakes. CsLoc results are based on 1239 felt earthquakes seen by crowdsourced detections recorded between September the 3rd 2020 and May the 15th 2021 at the EMSC. 550 out of 1239 earthquakes have been located by CsLoc and 478 out of the 550 have also a magnitude estimation. CsLoc is very fast with a median publication delay of the earthquake location of 98 sec, while it is around 270 sec (4 min 30 sec) for the EMSC. The median publication time of the CsLoc magnitude, when available, is 148 sec after origin time. CsLoc tends to overestimate the magnitude by 0.18±0.32 in average. To increase the number of events located by CsLoc and particularly the low magnitude ones, the raspberry shakes will be added. Until the next release of CsLoc by the end of 2022, CsLoc is already a service available on EMSC website since October 2020.

1. Introduction

EMSC has developed a system call CsLoc (*Steed et al., 2019)*, improved in *(Bondár et al., 2020*) for the detection of felt earthquakes via crowdsourcing, which is capable to determine fast and reliable seismic locations with only few seismic stations.

In the RISE framework, we describe the CsLoc methodology and the latest improvements since the TURNkey deliverable D3.6, in particular how increasing the number of seismic stations helps CsLoc to locate lower magnitude felt earthquakes and even though it was not planned, we present the CsLoc magnitude. Finally, we show an update of the results presented in TURNkey D3.6 deliverable based on 8 months of data.

2. CsLoc Methodology

The CsLoc procedure is illustrated in Figure 1. Upon notification of a crowdsourced detection, CsLoc searches for phases from stations within 1000 km (increased to 2000 km if not enough station are available) of the crowdsourced barycenter and close in time to the crowdsourced detection time. A "phase" or "arrival time" is the measured time that a seismic wavefront arrives at a seismic station. A simple phase association procedure is then used. The chosen phases are located in a window centered on a regression to the ak135 seismic propagation model (Figure 2) (*Kennett et al., 1995*). These phases are then analyzed by the location software iLoc (Bondár and Storchak, 2011). If a location is found, then the result must also satisfy publication criteria defined in terms of standard seismological parameters (Bondár et al.; 2020). Until the criteria are met, the process runs iteratively at 15-s intervals; each iteration may add newly received arrival times and starts from the solution obtained in the previous iteration. If people seem to react either to the P-wave or S-wave front, CsLoc collects only P-wave arrival times from 1417 seismic stations. Since the beginning of TURNkey and Rise projects, amplitudes are also collected, which allowed us to compute a magnitude. In near real time, arrival picks and amplitudes data are received via the httpmsgbus (HMB) protocol (*Heinloo, 2016*), which is essential for the system's response time.

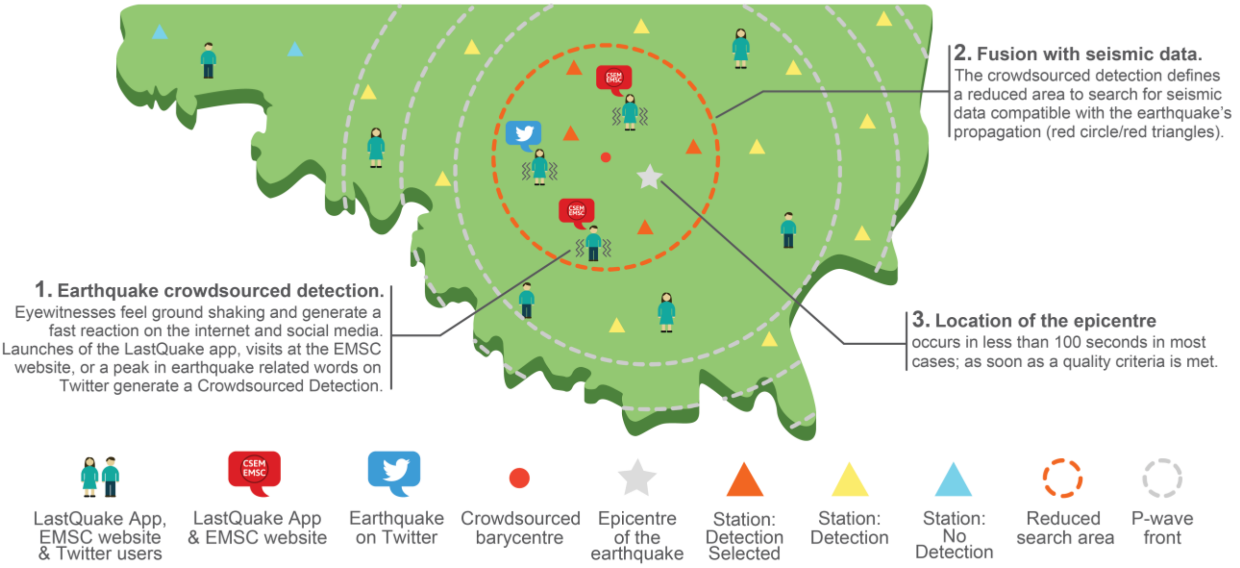

Figure 1: CsLoc fuses crowdsourced and seismic detection of earthquakes. Crowdsourced detections are quick but do not yield the physical properties of an event, and some detections are not related to seismic events. Seismic networks need strong quality criteria for the automatic publication of seismic events to avoid false detections. The fusion of the two sources of data improves the reliability of crowdsourced detections and reduces the response time of a seismic network for the rapid location of felt events.

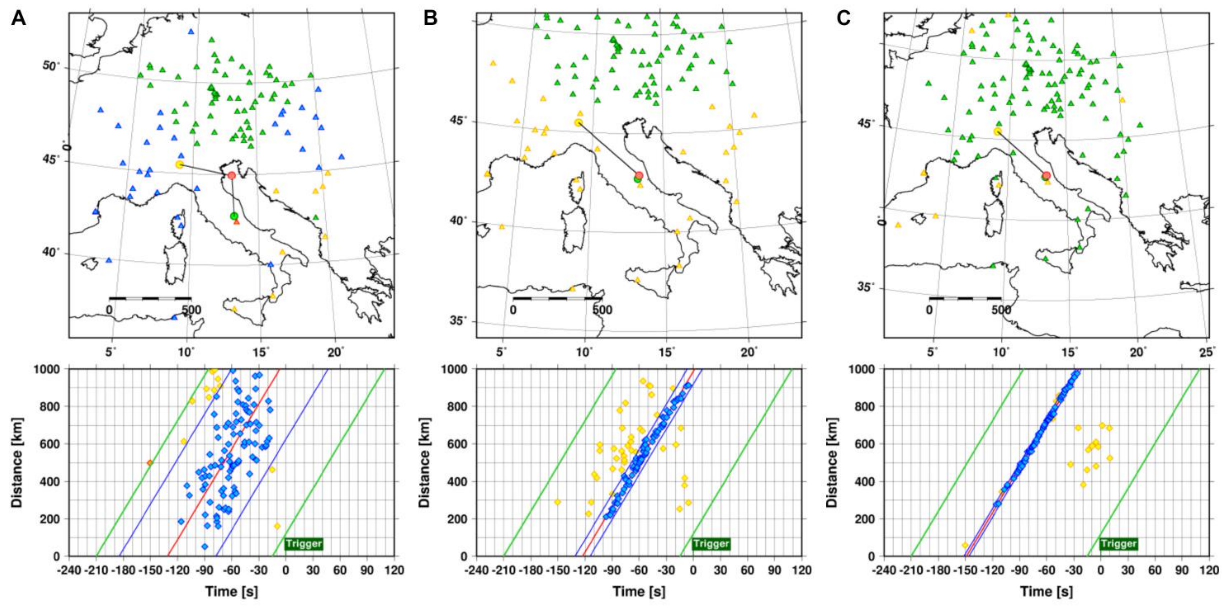

Figure 2: The CsLoc association and location cycle, for iterations (A) 0, (B) 1, and (C) 2. Top row: The initial crowdsourced trigger (yellow circle) may be far away from the EMSC seismic location (green circle), but iLoc (red circle) converges fast to the traditional seismic location. Yellow, blue and green triangles show the seismic stations considered, associated and used in the locations, respectively. Bottom row: First-arriving P phase picks are considered in a time window (green lines) before the crowdsourced trigger. Those within three times the median absolute deviation (MAD) (blue lines and blue diamonds) of the best fitting travel time curve (red line) with the slope of the ak135 Pn velocity, 8.04 km/s, are passed to iLoc.

Because the same earthquake can lead to several crowdsourced detections (Twitter, the app, website, and recently EQN, in several countries) (Figure 3), we have to determine on the fly that these detections are related to the same earthquake. It is very essential to avoid duplication of event in a fully automatic service (Bondár et al., 2020). Events with a large number of seismic arrivals require separate logic from those with only few seismic arrivals. We rely on the assumption that if two solutions share a fair amount of common seismic arrival picks then the events are likely to be the same. For candidate events for multiple triggers we check the number of common seismic arrivals for each event pair. If the number of common seismic arrival picks is larger than 20, we declare the two events as common. For events with just a few picks, we require at least three common seismic arrival picks and that 20% of the seismic phases be shared between the events to declare them the same. We publish the first location obtained by CsLoc, whatever the type of trigger and country, and because the magnitude is often obtained after more iteration, we only update the first location when a magnitude is available.

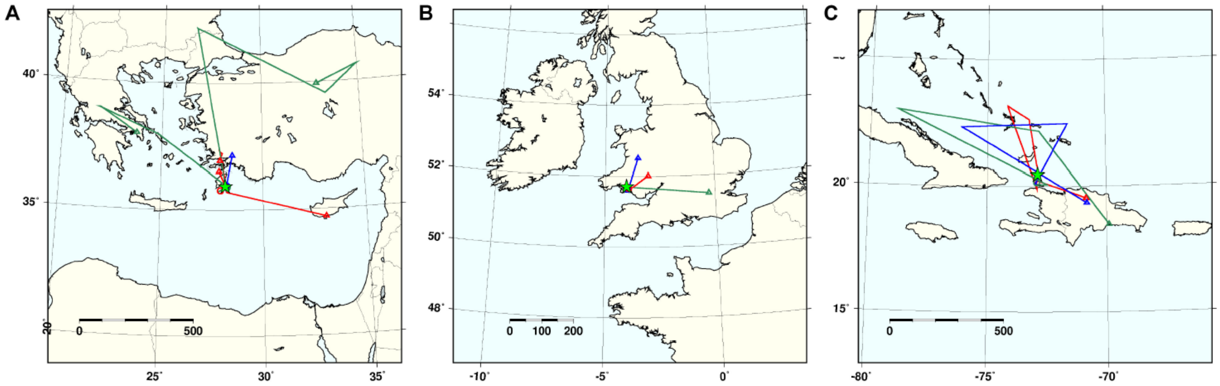

Figure 3: Multiple strains for the same event (star) triggered by various country-based website traffic (green triangle) and TED triggers (blue triangle), as well as the LastQuake app (red triangle) crowdsourced detections in (A) Turkey, (B) Great Britain, and (C) Haiti. Corresponding color lines show the trajectory of CsLoc locations during the iterations. CsLoc shows a robust performance against the position of the initial crowdsourced triggers.

3. Magnitude estimation

Previous CsLoc studies (Steed et al, 2018; Bondár et al, 2019) focus on the speed and accuracy of the seismic locations. Although not planned, the magnitude estimation has been added to CsLoc.
CsLoc is based on the iLoc location algorithm to locate the event and estimate the magnitude (Bondár and Storchak, 2011; Bondár et al., 2018). Since we collect phase picks up to 1000 km (for sparse networks up to 2000 km) this would allow us to calculate mainly the local magnitude MLv and in some rare cases the mB magnitudes. Of course the MLv magnitude starts to saturate relatively early at medium moment magnitudes, therefore for some cases ML would be an underestimated magnitude. For these events we will not publish ML at all.

4. Seismic stations network

In Steed et al. (2018) and Bondár et al. (2019) studies, only seismic stations provided by the GEOFON program have been used. The GEOFON Network is a virtual network with contributions from many different agencies and institutions. Although lots of seismic stations are already used, we added new seismic stations to fill up the gap in the GEOFON network. These new seismic stations are processed by a SeisComP server operating at EMSC. All picks and amplitudes are sent to CsLoc on the fly via the HMB protocol (Heinloo, 2016).

Steed et al. (2019) study show that almost all non-located earthquakes are outside of the seismic stations network. To enhance the coverage, 880 seismic stations have been added from 105 networks for a total of 1417 seismic stations (Figure 4). As CsLoc performs in the first 5 minutes of an origin time, it needs an extremely dense network around the crowdsourced detections to locate the earthquake.

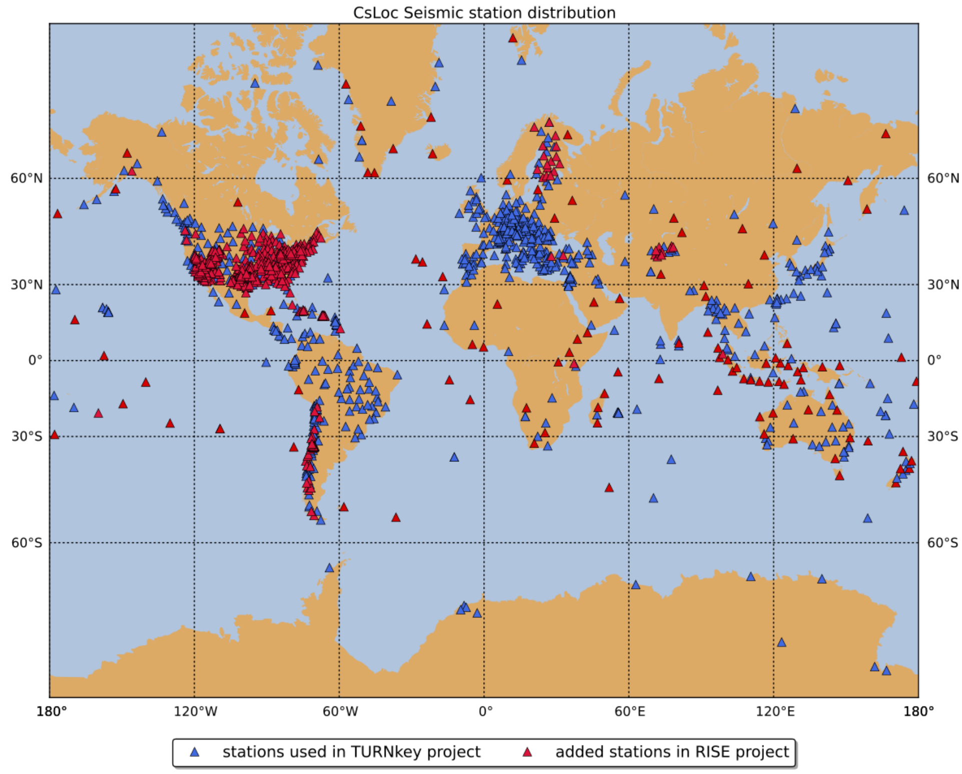

Figure 4: Seismic station distribution currently in use. The blue ones correspond to the station used until December 2020 and the red ones are 880 stations added for RISE in January 2021.

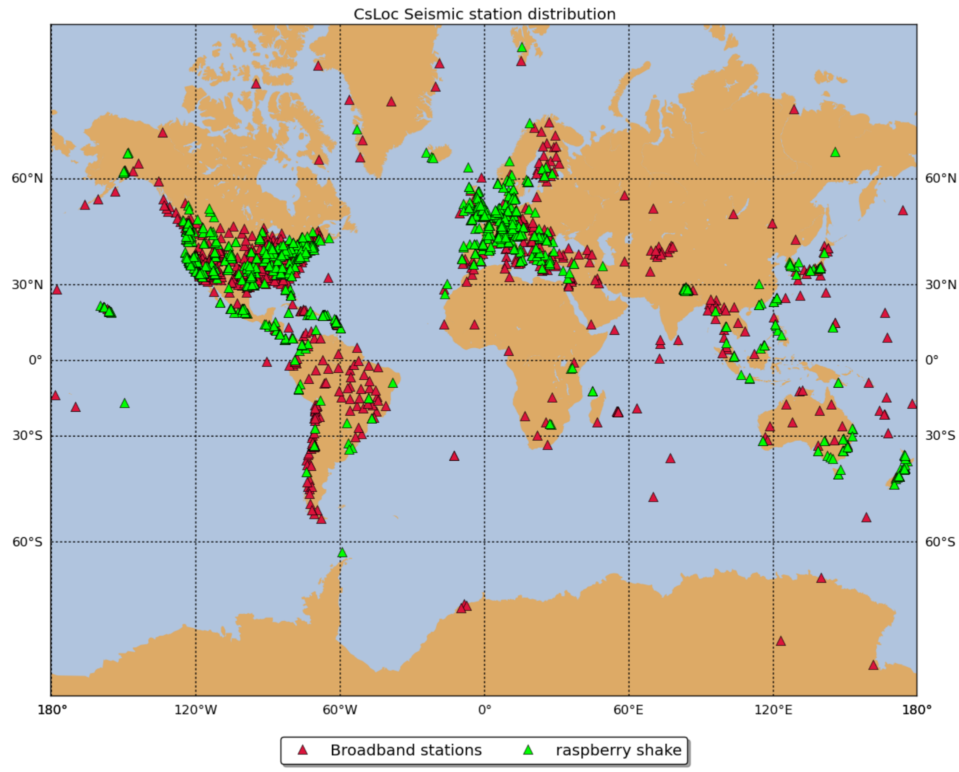

Figure 5: Same seismic station distributions in red as **Figure 4**, with the 1347 raspberry shakes in green.

Although very dense our seismic network still presents some gaps which are very crucial to fill to increase CsLoc performance and accuracy. To do so, we are currently working on the integration of the raspberry shakes (Figure 5). The raspberry shakes picks are send by the raspberry shake company without delay using scimport to our seiscomp at EMSC. The Figure 5 shows the contribution of the raspberry shakes to CsLoc. 1345 raspberry shakes have been added to another instance of CsLoc (not published yet). One can expect more located event by CsLoc in Hawaii, Nepal, New Zealand, and Central America and should bring a better constraint in Europa and US.

5. Results

CsLoc needs good seismic stations coverage around crowdsource location to locate an earthquake, especially for small magnitude events which need very close seismic stations. Steed et al. (2019) and Bondár et al. (2020) studies have used restricted seismic stations in Indonesia. As CsLoc is now a public service, we can no longer use these stations which can induce a decrease in efficiency of CsLoc in this area.

The evaluation of performances is based on 1239 felt earthquakes seen by crowdsourced detections recorded between September the 3rd 2020 and May the 15th 2021 at EMSC (Figure 6). As one earthquake can lead to several crowdsourced detections, these 1239 earthquakes correspond to 1952 crowdsourced detections (Figure 8). There has been a number of false crowdsourced detections (235), mainly from TED, none of them led to the creation of a fake seismic event. Among these 1239 earthquakes, 550 have been located by CsLoc and 478 out of the 550 have also a magnitude estimation.

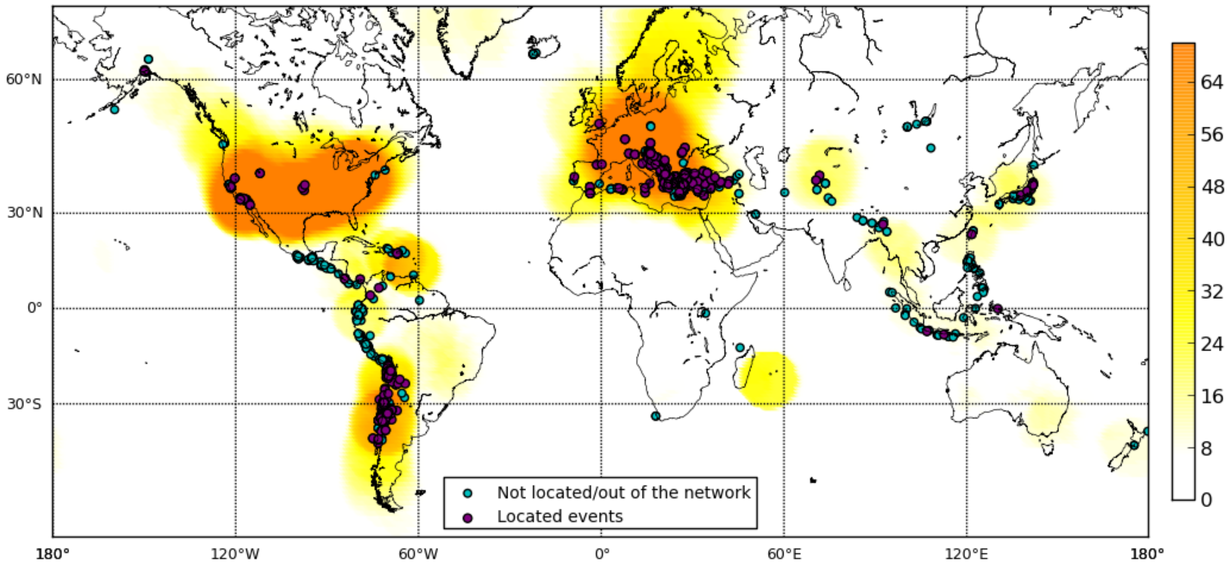

Figure 6: Results of CsLoc analyses overlaid on a density plot of the number of seismic stations within 1000 km of each position. Successful locations are related to local network density: Almost all non-located events are out of the network.

The Figure 6 shows the results of CsLoc for the 1952 triggers overlaid on a density plot of the number of seismic stations within 1000 km of each position and clearly show that most of the non-located earthquakes lay outside the network, especially in Central and South America (except Chili), Indonesia, India and Nepal. Using the Raspberry Shake P arrival time in these regions should help us to locate more earthquakes.

The Figure 7 shows the distribution of magnitudes of our earthquakes dataset. It shows that even though our dataset includes lots of small magnitudes (<4), which are difficult to locate, CsLoc convergence doesn't not depend on the magnitude. For larger magnitudes, better seismic station coverage is needed (Figure 6). Even though 880 seismic stations have been added since September 2020, more stations are required. The Raspberry shake should help to locate these events.

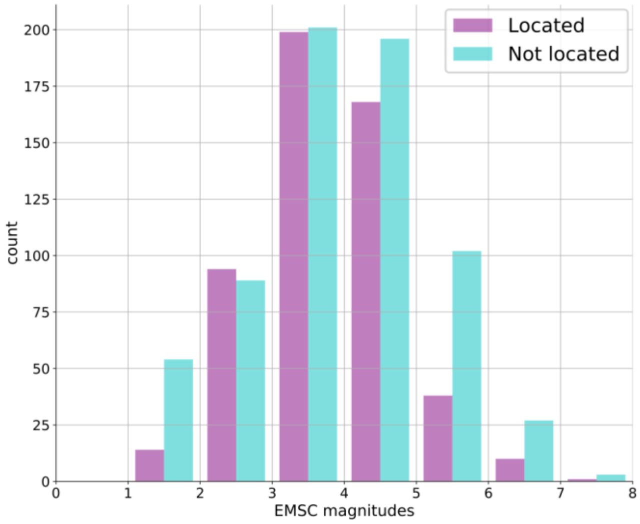

Figure 7: Distribution of magnitude of the 1239 felt earthquakes. The violet bars correspond to the located earthquakes by CsLoc and the cyan bars correspond to the non-located earthquakes.

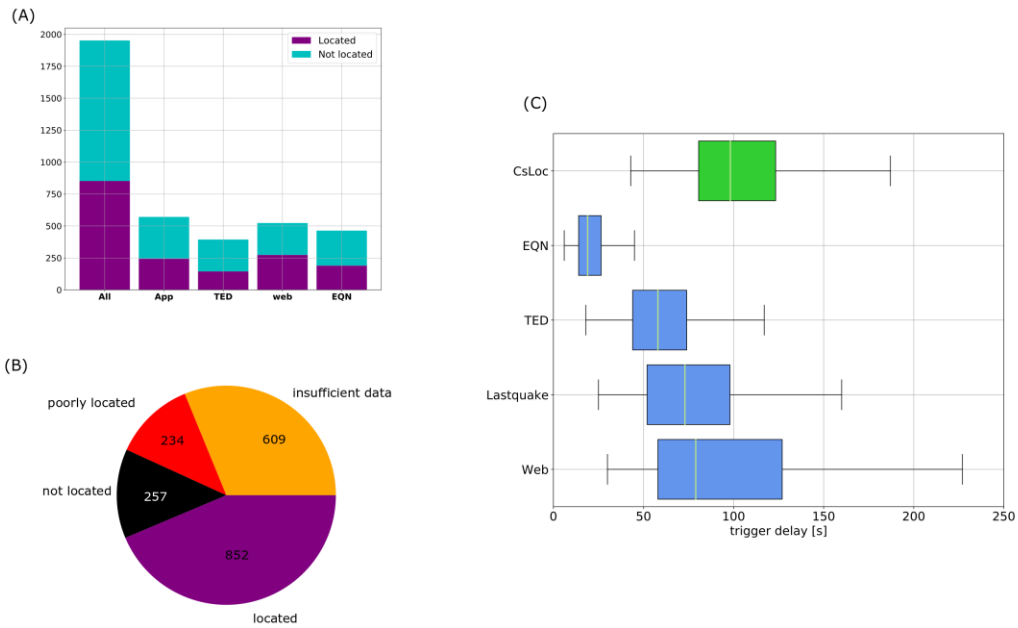

Figure 8: (A): Number of crowdsourced detections by systems. Some Earthquakes were detected by multiple systems. (B): Results of CsLoc for 1952 triggers. (C): trigger delay after origin time for the different 852 located crowdsourced detections. The red lines inside the box represent the median and the box is limited by the first and third quartile.

CsLoc is based on the location algorithm iLoc (Bondár and Storchak, 2011) and needs at least 4 seismic stations to produce a reliable location. Insufficient data means that 4 or less seismic stations were available. When CsLoc location does not satisfy the publication criteria defined in terms of standard seismological parameters (Bondár et al., 2020) the solution is poorly located and not published. The number of insufficient data and poorly located solutions in Figure 8-B) is directly related to a lack of seismological data in real time. Adding the raspberry shake should decrease the ratio of poorly located or insufficient data solutions.

The Figure 8 A) show that the CsLoc location does not depend on the type of triggers (LastQuake, Web, EQN, TED). In terms of speed, the Figure 8-C) show that triggers from EQN (Finazzi, 2016) are the fastest with a median trigger in 19 sec, while the triggers from the website are the slowest (median = 79 sec).

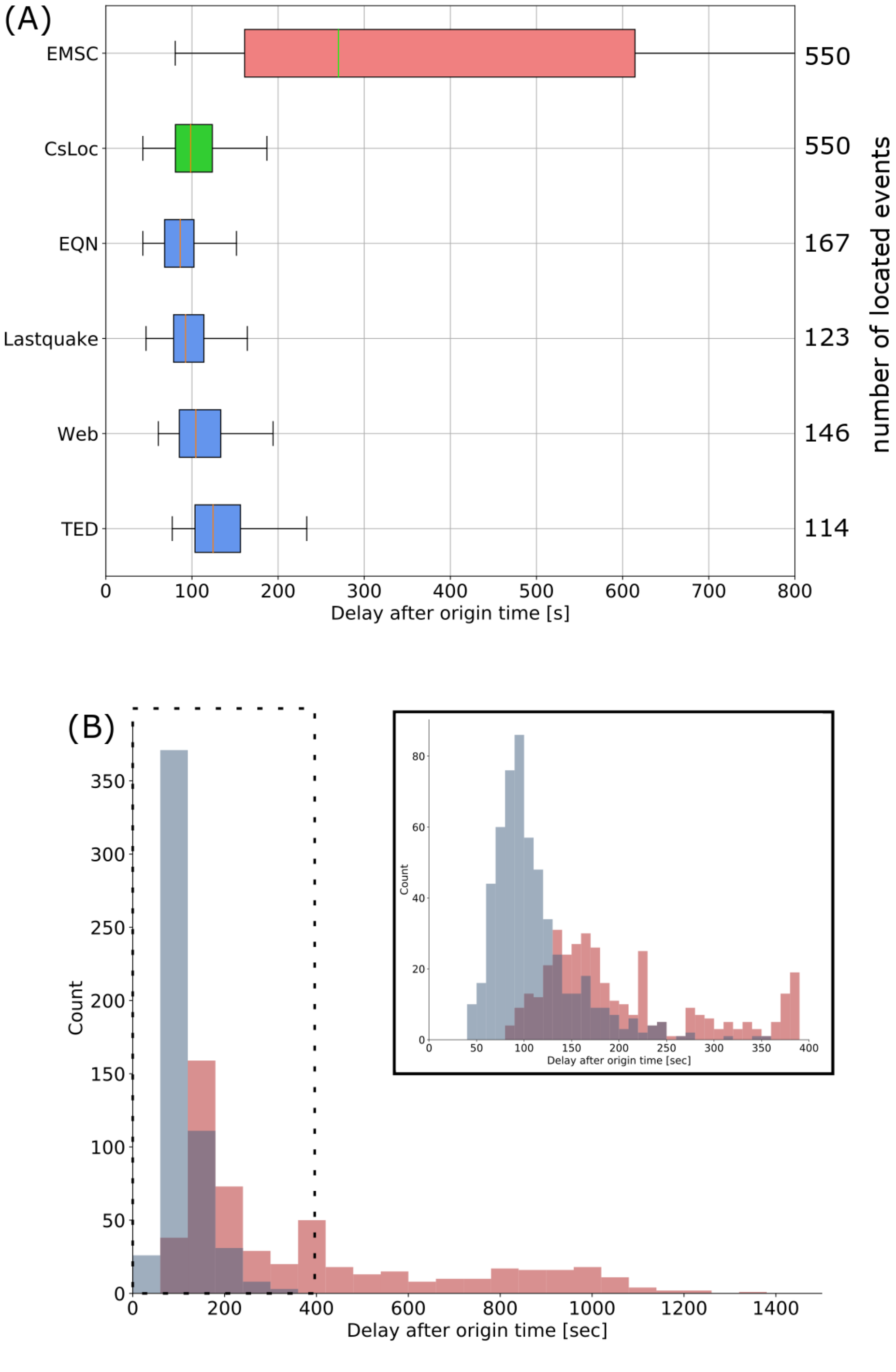

Figure 9: Publication delay after the origin time for the 550 events located by CsLoc. (A): Box-and-whisker plot of publication delay for each crowdsource trigger types (blue), all CsLoc locations and the EMSC. The red lines inside the box represent the median and the box is limited by the first and third quartile. (B): Histogram of CsLoc (blue) and EMSC (red) publication delay after the origin time in seconds.

From the 1239 earthquakes, 550 distinct earthquakes have been located by CsLoc and published by the deduplication algorithm. No duplicate event has been published, indicating that the deduplication criteria are well tuned. Nevertheless, CsLoc publication is very fast and is often the first seismic message published on the "for seismologist" EMSC website section (Figure 9). The median publication delays for the CsLoc locations are 98 sec, while it is around 270 sec (4 min 30 sec) for the EMSC publication delay. The median publication delay is 86, 92, 104, 124 sec for EQN, LastQuake, website and TED, respectively. The fastest solution published by CsLoc was 43 sec after origin time.

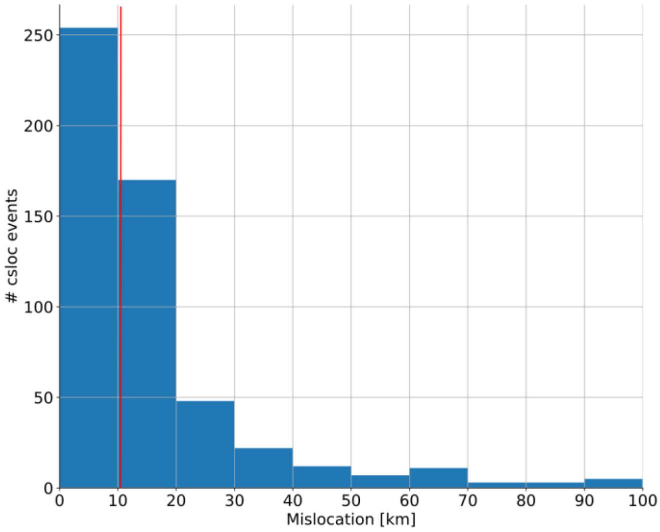

Figure 10: Histogram of separation of final publishable CsLoc result for each earthquake with respect to the final EMSC-published epicenter with a median mislocation of 11 km (red line).

92% of the CsLoc locations were located at less than 50 km of the final EMSC-published epicenter (Figure 10) with a median mislocation of 11km. 11 CsLoc locations were located at more than 100 km. Locations affected by uncertainties greater than 50 km do not seem to follow a specific pattern in terms of number of stations or in terms of magnitudes. Large mislocations (> 100 km) can occur also during aftershocks sequence and are very difficult to deal with.

The automated magnitude computation by CsLoc is the most challenging, which considers the delay of publication (less than 2 min) and the number of seismic stations to handle. CsLoc computes only ML and mB magnitudes and are compared to EMSC magnitudes which can be Mw, mb, and ML. 478 seismic events located by CsLoc out of the 550 were published with a magnitude. The Figure 11-A) shows that CsLoc magnitude is available very quickly after origin time. The median of the magnitude publication distribution is 148 sec after origin time. The fastest published CsLoc magnitude was 49 sec after origin time. CsLoc tends to overestimate the magnitude by 0.18 in average (Figure 11-B) with a standard deviation of 0.3. In the TURNkey deliverable D3.6, we show that the difference between the EMSC and CsLoc magnitudes does not depend on the event location, or the earthquake magnitude. Nevertheless, after investigations, we found several issues regarding the magnitude computed by CsLoc.

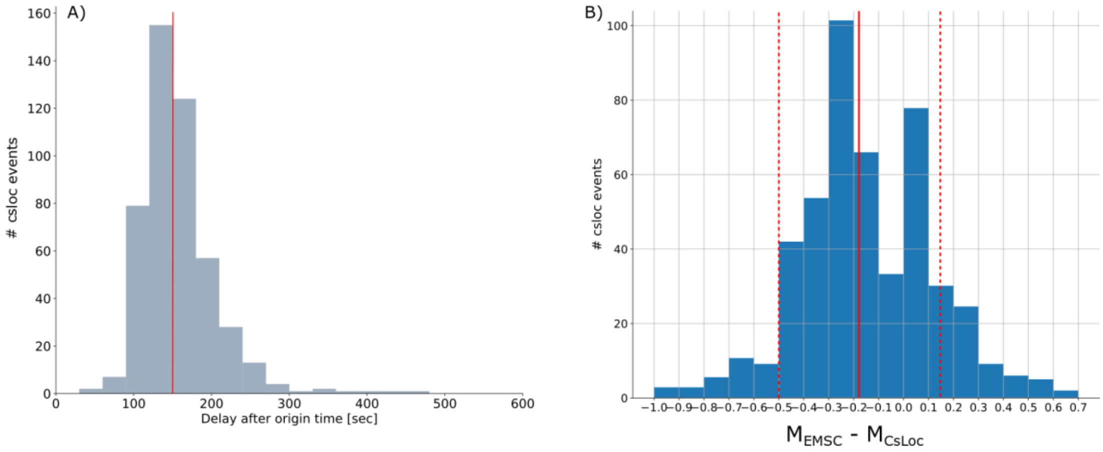

Figure 11: A) publication delay of the CsLoc magnitude. The median publication delay (in red) is 148 sec. B) Histogram of the deviation of CsLoc magnitude determinations from EMSC values. The average deviation is -0.18 (red line) with a standard deviation of 0.32 (dashed red lines).

The first issue that explains the overestimated CsLoc magnitude and particularly for $|M_{EMSC} – M_{csLoc}| > 0.6$, is related to a bad station reference in the station book. Due to the large number of available stations worldwide, many stations have the same name and without using the network code, we can't make the difference between one or the other.
The second issue is related to bad instrumental responses which lead to magnitude overestimates or underestimates by CsLoc. This issue appears more likely when few station magnitude contributions are used. To overcome this issue, we setup a station magnitude contribution monitoring to check which stations often provide incorrect magnitude estimations.

When removing from the dataset the bad CsLoc magnitude due to bad instrumental responses or station misidentifications, the average magnitude deviation is -0.15 and the standard deviation drops to 0.25.

CsLoc is now a service, and its seismic locations and magnitudes are published since October 16th, 2020 (Figure 12).

2021-05-01	**20:30:47.7**	36.99 N	27.36 E	13	ML	2.6	A	DODECANESE IS.-TURKEY BORDER REG	CSLC
2021-05-01	**20:30:47.5**	**36.95 N**	**27.40 E**	**15f**	**ML**	**2.6**	**M+**	**DODECANESE IS.-TURKEY BORDER REG**	**INFO**
2021-05-01	**20:30:47.0**	36.95 N	27.50 E	12	ML	2.9	A	DODECANESE IS.-TURKEY BORDER REG	THE
2021-05-01	**20:30:46.9**	36.91 N	27.35 E	4	ML	2.6	M	DODECANESE IS.-TURKEY BORDER REG	KAN
2021-05-01	**20:30:46.2**	36.88 N	27.29 E	7	ML	2.5	M	DODECANESE IS.-TURKEY BORDER REG	DDA
2021-05-01	**20:30:24.2**	36.89 N	27.24 E	20	ML	2.5	M	DODECANESE IS.-TURKEY BORDER REG	THE

Figure 12: CsLoc seismic locations are published on the EMSC website in the For Seismologist section. The full message with the date and time of publication is available inside the CsLoc message.

6. Conclusion and perspectives

CsLoc methodology has been significantly improved since the start of the RISE/TURNkey projects over three main aspects. In the RISE framework, we took great care of the seismic stations distribution, adding 880 seismic stations for a total of 1417. The magnitude has been added, although not planned. CsLoc is now a service and is published on the EMSC website.

The performance analysis is based on 1239 felt earthquakes seen by crowdsourced detections recorded between September the 3rd 2020 and May the 15th 2021 at the EMSC which correspond to 1952 crowdsourced detections. We show that CsLoc is very fast and accurate. The median publication delays for the CsLoc locations are 98 sec, while it is around 270 sec (4 min 30 sec) for the EMSC. The median time of publication of the CsLoc magnitude is 148 sec. CsLoc tends to overestimate the magnitude by 0.18±0.32.

CsLoc failed to find a solution for 1100 out of 1952 crowdsourced detections and 609 out of the 1100 are due to a lack of seismic data (< 4 seismic stations). This, clearly, shows the need of increasing the number of seismic stations in some regions.

The next step will be to add the raspberry shakes to our current seismic network and to publish the CsLoc solution on LastQuake and Twitter by the end of the project.

References

Bondár I, Storchak D (2011). Improved location procedures at the International Seismological Centre. *Geophys. J. Int.* **186**, 1220–1244.

Bondár I, Steed R, Roch J, Bossu R, Heinloo A, Saul J, Strollo A (2020). Accurate locations of felt earthquakes using crowdsource detections. Frontiers in Earth Science 8, 272.

Bossu R, Roussel F, Fallou L, Landès M, Steed R, Mazet-Roux G, Dupont A, Frobert L, Petersen L (2018). LastQuake: From rapid information to global seismic risk reduction. *Int. J. Disaster Rick Reduct.* **28**, 32–42.

Finazzi F (2016). The earthquake network project: Toward a crowdsourced smartphone-based earthquake early warning system. *Bull. Seismol. Soc. Am.* **106**, 1088–1099.

GEOFON Data Center: P-phase arrival times used in steed et al. 2018 for crowdseeded Seismic locations (CsLoc). GFZ Data Services, 2018, http://doi.org/10.5880/GFZ.2.4.2018.002

Heinloo A (2016). The httpmsgbus general-purpose HTTP-based message bus. *GFZ Data Services*.

Kennett B L N, Engdahl E R, Buland R (1995). Constraints on seismic velocities in the Earth from traveltimes. *Geophys. J. Int.* **122**, 108–124.

Steed J R, Fuenzalida A, Bossu R, Bondár I, Heinloo A, Dupont A, Saul J, Strollo A (2019). Crowdsourcing triggers rapid, reliable earthquake locations. Sci. Adv.s,5, eaau9824, doi:10.1126/sciadv.aau9824.