What to Expect from the Virtual Seismologist: Delay Times and Uncertainties of Initial Earthquake Alerts in California

Yannik Behr[1], John F Clinton[1], Philipp Kästli[1], Roman Racine[1], Men-Andrin Meier[1], Carlo Cauzzi[1], Georgia Cua[1]

[1]Swiss Seismological Service, ETH Zurich

contact: behr@sed.ethz.ch

1. Introduction

The Virtual Seismologist (VS) method is a Bayesian approach to regional network-based earthquake early warning (EEW) originally formulated by Cua and Heaton (2007). Implementation of VS into real-time EEW codes has been an on-going effort of the Swiss Seismological Service at ETH Zurich since 2006, with support from ETH Zurich, various European projects, and the United States Geological Survey (USGS). VS is one of three EEW algorithms — the other two being the ElarmS [Allen and Kanamori, 2003] and On-Site [Wu and Kanamori, 2005; Boese et al, 2008] algorithms — that form the basis of the California Integrated Seismic Network (CISN) ShakeAlert system, a USGS-funded prototype end-to-end EEW system that could potentially be implemented in California.

The possible use cases for an EEW system will be finally determined by the speed and reliability of the estimates of the earthquake source parameters. A thorough understanding of both is therefore essential to evaluate the usefulness of VS. For California, we present statewide theoretical alert times for hypothetical earthquakes by analyzing time delays introduced by the different components in the VS EEW system. Taking advantage of the fully probabilistic formulation of the VS algorithm we further present an improved way to describe the uncertainties of every magnitude estimate by evaluating the width and shape of the probability density function that describes the relationship between waveform envelope amplitudes and magnitude.

2. Definition of Delays in the Virtual Seismologist and Methodology to Model Expected Alert Times

The overall delay in an early warning alert is defined as the time between the earthquake intitation and the issuance of the first alert which we label Δt_{alert}. For the Virtual Seismologist, we split this into the following parts:

$$\Delta t_{origin} = \max(\overline{\Delta t_p} + \overline{\Delta t_t}) + \Delta t_e \qquad (1)$$

$$\Delta t_{alert} = \max(\min(\overline{\Delta t_p} + \overline{\Delta t_w}), \Delta t_{origin}) + \Delta t_m \qquad (2)$$

where Δt_{origin} is the time between an earthquake initiation and the first available hypocenter estimate; the other components that contribute to the first alert are defined in Table 1. The relationships between the delays are shown in Figure 1
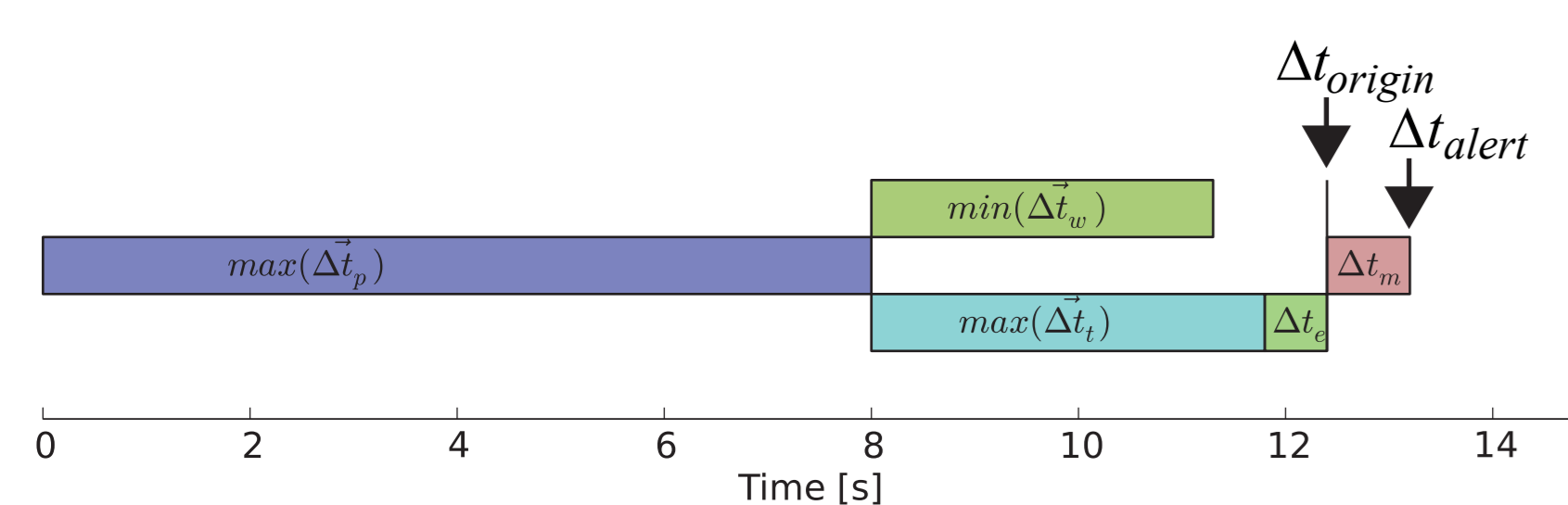

Figure 1: Summary of the delays that are part of an EEW alert.

Table 1: Component delays in the Virtual Seismologist. Only the highlighted delay times are directly measured in this study. Many of the intermediate delay times are either / both close to neglible and /or difficult to independently isolate (Δt_d, Δt_{pk}, Δt_{trans}, Δt_{rec}, Δt_{pk}, Δt_{pw}). Δt_{log} and Δt_{win} are known a priori with high accuracy.

Delay	Description	Dependencies
$\overline{\Delta t_p}$	Time for the P-wave to arrive at each of the first n stations (vector of length n)	network geometry and hypocentral location
$\overline{\Delta t_t}$	Time between the arrival of the P-wave at each of the first n stations and the trigger / detection of the P-wave at the network (vector of length n). Includes:	
$\overline{\Delta t_t} = \overline{\Delta t_{log}} + \overline{\Delta t_{trans}}$	Data latency for each of the first n stations	
$\overline{\Delta t_{log}}$	Processing and packaging at the datalogger at the first n stations	datalogger, sampling rate
$\overline{\Delta t_{trans}}$	Telemetry delay between the first n stations and the datacenter	communications, sampling rate
Δt_{rec}	Receiver queuing and processing at the datacenter	acquisition software
Δt_{pk}	Processing delay of the automatic picker	EEW software
Δt_e	Associator processing delay: time between using all picks to derive a location	EEW software
$\overline{\Delta t_w}$	Time between the arrival of the P-wave at each of the first n stations and when k seconds of waveform data is available in order to estimate magnitude. Includes:	
$\Delta t_{win}(k)$	Waveform window, k seconds long,	datalogger, communications, sampling rate, acquisition software
Δt_{pw}	Pre-processing of waveform data	EEW software
Δt_m	Delay to compute EEW magnitude with available location and waveform data	EEW software
Δt_d	Alert dissemination	messaging system and end-user internet connection

To generate a model of alert times for a given network geometry we first compute Δt_p for every epicenter and then draw a sample of the delay distributions for Δt_w and Δt_t at each of the first n stations. Together with a sample of the distributions for Δt_e and Δt_m this results in one possible scenario for Δt_{alert}. By repeating this analysis 500 times we get a distribution of possible alert times for every potential epicenter.

3. Analysis of Component Delays

Since VS requires at least four P-wave detections to first estimate location, we compute the P-wave travel time delay Δt_p to the first four seismic stations for a constant hypocentral depth of 8 km and a homogeneous velocity model with $V_p = 6.5$ km/s (Figure 2). In Figure 3, we show the distributions of the measured delays across the EEW system. The trigger delay Δt_t and delay in waveform amplitude processing Δt_w are measured during snapshots of time in real-time operations for every seismic station that is monitored by VS. Event association delay Δt_e and final VS magnitude estimation delay Δt_m are determined from off-line playbacks of 119 events with magnitudes ≥3.5 from southern California between January 2010 and August 2012. VS is currently running independently at three different locations each seeing only a subsection of the seismic stations in California.
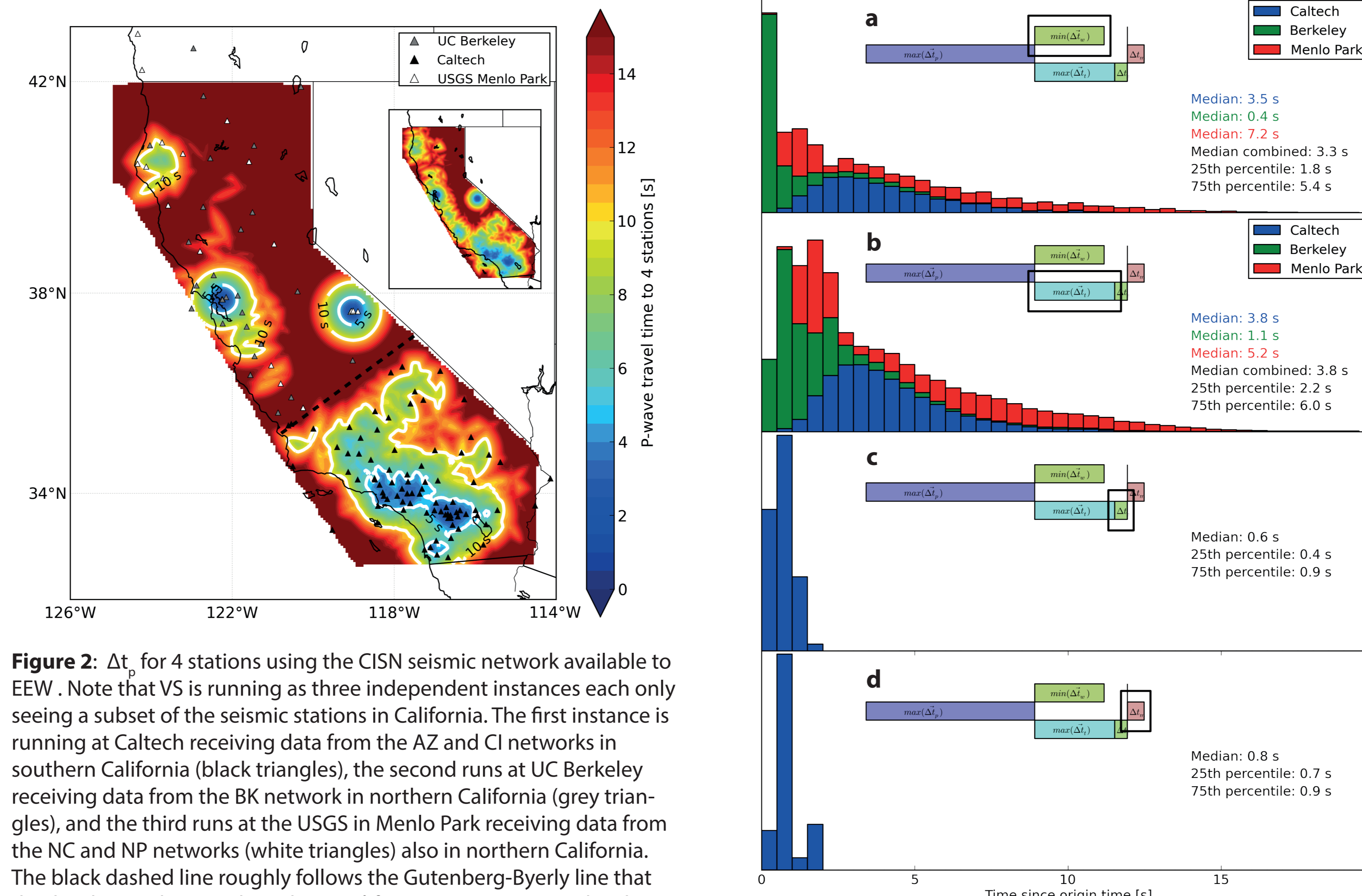

Figure 2: Δt_p for 4 stations using the CISN seismic network available to EEW . Note that VS is running as three independent instances each only seeing a subset of the seismic stations in California. The first instance is running at Caltech receiving data from the AZ and CI networks in southern California (black triangles), the second runs at UC Berkeley receiving data from the BK network in northern California (grey triangles), and the third runs at the USGS in Menlo Park receiving data from the NC and NP networks (white triangles) also in northern California. The black dashed line roughly follows the Gutenberg-Byerly line that divides the northern and southern Californian seismic networks. The white contour lines outline the areas within which Δt_p is less than 5 s and 10 s. The inset shows Δt_p for a single statewide VS installation.

Figure 3: Delay time distributions for a) Δt_w and b) Δt_t at all stations for which VS is receiving data at the three independent installations; c) Δt_e and d) Δt_m measured from off-line event playback.

4. Overall System Delays — Comparison between Observations and Expectations

Using the distributions presented in the previous section, we can model the expected alert times within California for possible epicenters across the state. We compare the resulting map of the median, 16th, and 84th percentile alert times with real alert times as observed using real-time detections from 99 events that occurred within the region of interest between 1 January 2012 and 1 November 2013 and had magnitudes ≥3.5. Almost all observed real-time alert times fall within the 16th and 84th percentile of the predicted alert time distributions which correspond to ±1σ if the alert times were normally distributed.
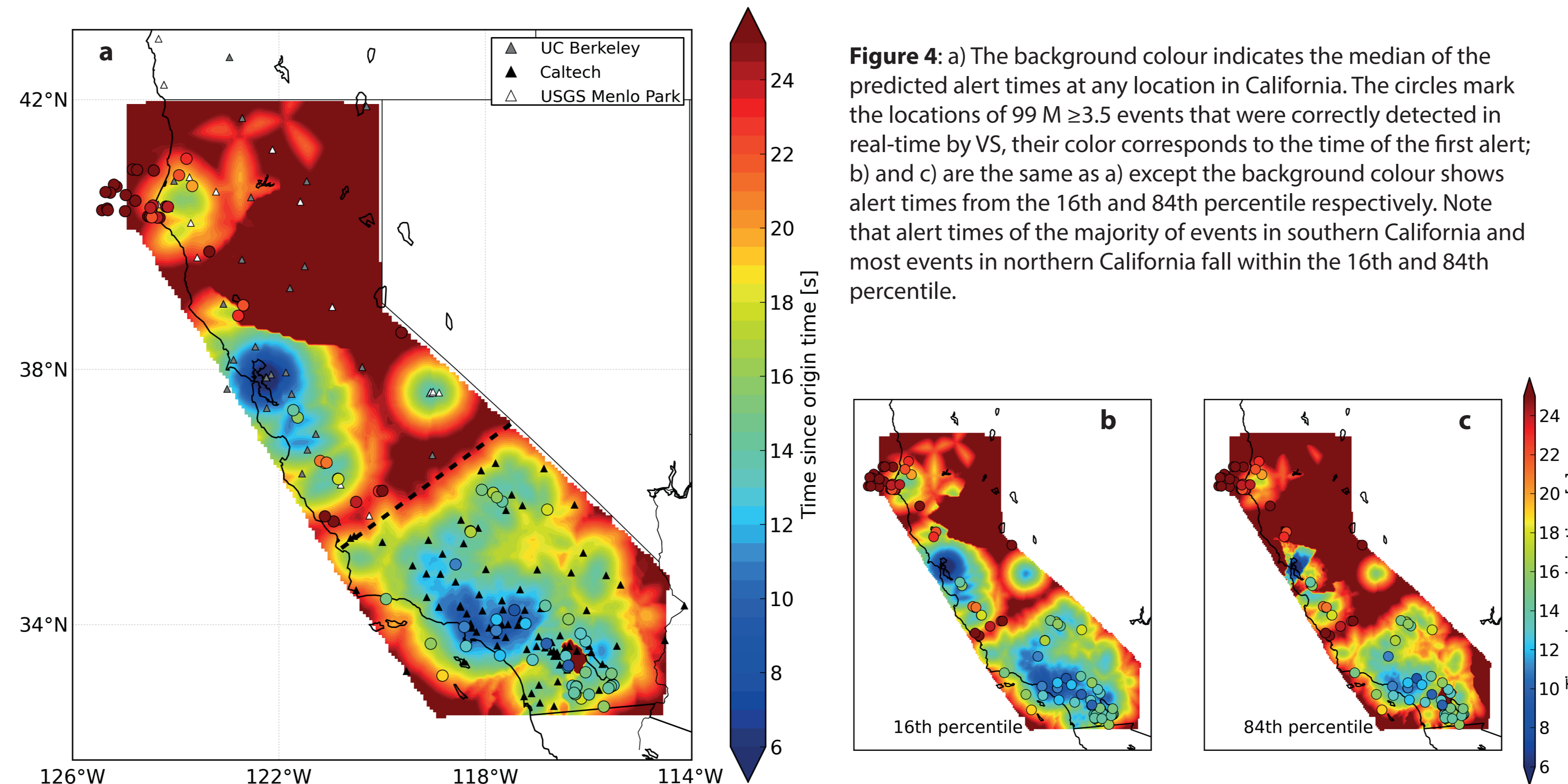

Figure 4: a) The background colour indicates the median of the predicted alert times at any location in California. The circles mark the locations of 99 M ≥3.5 events that were correctly detected in real-time by VS, their color corresponds to the time of the first alert; b) and c) are the same as a) except the background colour shows alert times from the 16th and 84th percentile respectively. Note that alert times of the majority of events in southern California and most events in northern California fall within the 16th and 84th percentile.

5. Improving the Virtual Seismologist: Better Estimation of Uncertainties and Using Priors

5.1 Evaluating the PDF to Estimate Magnitude Uncertainty

An accurate magnitude uncertainty estimate is crucial especially for end-users of EEW alerts for whom the costs of action are high. Uncertainties in magnitude are currently determined from the discrepancies between the VS magnitude estimate and the magnitude from the Advanced National Seismic System (ANSS) catalog for past seismicity. These differences are evaluated for all correctly detected earthquakes with a magnitude ≥3.0 that occurred between May 2012 and January 2013 within California's state boundaries. This method of computing uncertainties has the drawback of not taking into account differences in data quality due to the location and size of the earthquake. We therefore propose a different way of evaluating the magnitude uncertainty for VS by using the shape of the underlying likelihood function.
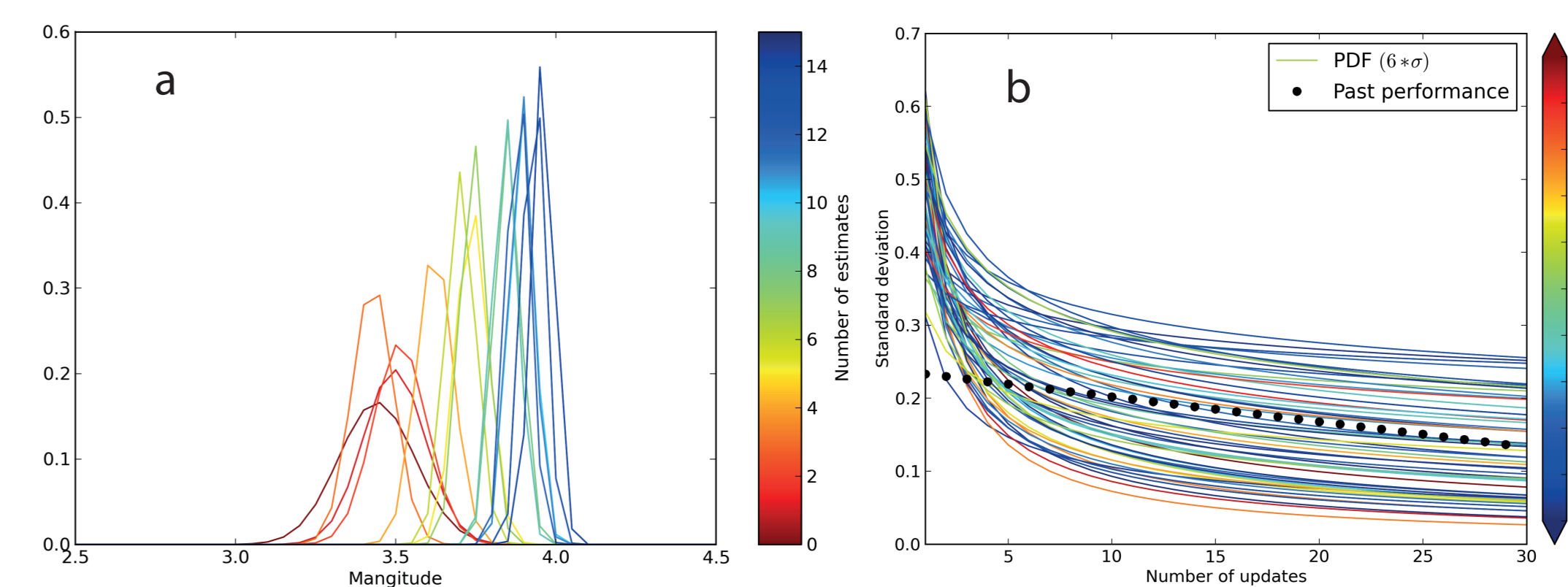

Figure 5: a) Example from a single earthquake of the evolution of VS' magitude likelihood function using the number of updates since the initial earthquake detection (VS updates every second); b) The existing CISN VS function for standard deviation based on past seismicity (dotted black line) and the standard deviation (as shown in a) for several events with magnitudes ≥3.5.

5.2 Using Prior information

The analysis shown in Section 3 and 4 reveals the requirement of waiting for 4 P-wave arrivals for the first detection as the dominant delay in all but the densest parts of the seismic network. To relax this constraint requires inclusion of additional information such as waveform characteristics or a priori information (prior). Although the original formulation of the VS algorithm includes priors, the current real-time implementations only evaluate the likelihood function. Potential candidates for priors are the network geometry, smoothed seismicity maps, smoothed fault-moment release maps or the expected shape of the VS likelihood function.
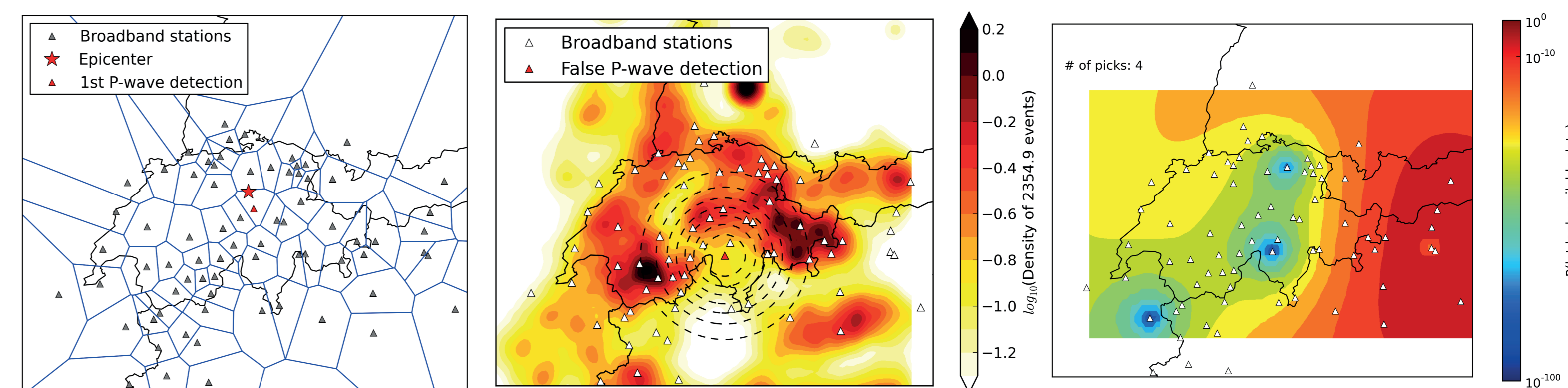

Figure 6: Solid blue lines show the boundaries of the Voronoi cells, i.e. the nearest neighbor regions of every station. The red star marks the epicenter of a magnitude 4.2 event and the red triangle the station that reported the first P-wave arrival. Its Voronoi cell would give a good first location estimate.

Figure 7: The colors show a seismicity density map with redder colors corresponding to higher seismicity. The red triangle marks a station with a false pick and the black dashed lines describe the resulting shape of the VS likelihood function with respect to latitude and longitude. In this case the seismicity density map would downweigh the likelihood of this being a real event.

Figure 8: The colored area shows the shape of the VS likelihood function with respect to latitude and longitude for a false event in Switzerland. The three pronounced minima are centered around three stations with false picks. By describing the overall shape of the likelihood function using, for example, gaussian mixture models, this could be classified as a false event.

6. Conclusions

Analyzing the performance of an EEW algorithm using only on-going realtime seismicity can give confidence in a system but does not provide a complete summary of potential performance in future events. We present here a simple but effective methodology to predict alert times for any network geometry and EEW algorithm. Our method not only reveals which components of the end-to-end EEW system produce the largest delays, but it can also serve as a framework for comparing performance of different EEW algorithms running on the same network.

It is clear that the most significant contributors to the overall delays in EEW are density and geometry of the seismic network, the data latency, and the number of P-wave detections required to locate an event. Improving the first 2 requires significant investment in hardware and communications infrastructure. For VS and other regional EEW algoithms, reducing the number of stations required to locate and estimate magnitude requires accurately characterising the uncertainty and increasing the reliability of these less constrained solutions. This means development of improved EEW-specific association algorithms that include more information than just P-wave arrival times.

References
Cua, G. B. and T. Heaton in Earthquake Early Warning Systems, 2007
Allen, R. M., and H. Kanamori in Science, 2003
Wu, Y.-M., and H. Kanamori in Bulletin of the Seismolgical Society of America, 2005
Böse, M., E. Hauksson, K. Solanki, H. Kanamori, and T. H. Heaton in Geophysical Research Letters, 2009