

USGS “Did You Feel It?” Internet-based macroseismic intensity maps

David J. Wald^{1,*}, Vincent Quitoriano¹, Bruce Worden², Margaret Hopper¹, James W. Dewey¹

¹ U.S. Geological Survey, National Earthquake Information Center, Golden, Colorado, USA

² Synergetics, Inc., Fort Collins, Colorado, USA

Article history

Received August 5, 2011; accepted September 12, 2011.

Subject classification:

Seismology, Macroseismology, Ground motions, Internet seismology.

ABSTRACT

The U.S. Geological Survey (USGS) “Did You Feel It?” (DYFI) system is an automated approach for rapidly collecting macroseismic intensity data from Internet users’ shaking and damage reports and generating intensity maps immediately following earthquakes; it has been operating for over a decade (1999–2011). DYFI-based intensity maps made rapidly available through the DYFI system fundamentally depart from more traditional maps made available in the past. The maps are made more quickly, provide more complete coverage and higher resolution, provide for citizen input and interaction, and allow data collection at rates and quantities never before considered. These aspects of Internet data collection, in turn, allow for data analyses, graphics, and ways to communicate with the public, opportunities not possible with traditional data-collection approaches. Yet web-based contributions also pose considerable challenges, as discussed herein. After a decade of operational experience with the DYFI system and users, we document refinements to the processing and algorithmic procedures since DYFI was first conceived. We also describe a number of automatic post-processing tools, operations, applications, and research directions, all of which utilize the extensive DYFI intensity datasets now gathered in near-real time. DYFI can be found online at the website <http://earthquake.usgs.gov/dyfi/>.

INTRODUCTION

Over the past decade, the U.S. Geological Survey’s “Did You Feel It?”[®] (DYFI) system has automatically collected shaking and damage reports from Internet users immediately following earthquakes. DYFI is now a rapid and vast source of macroseismic data, providing quantitative and qualitative information about shaking intensities for earthquakes in the USA and around the globe. Our systematic collection of citizen-provided data preceded the use of the formal concept of ‘crowdsourcing’ by more than a decade.

DYFI has become vital for automatically collecting macroseismic data for all felt seismic events in the United States; it is also one of the most popular, interactive websites within the Federal Government. For earthquakes occurring

outside the USA, the worldwide DYFI data rapidly indicate or confirm earthquake occurrence for seismic analysts and scientists at the USGS National Earthquake Information Center, giving a quick indication of the extent and nature of shaking effects. The global intensity data from DYFI, and in some cases its international counterparts, can be automatically used as constraints in our Global ShakeMap system (GSM) (Figure 1) [Wald et al. 2006a], which is the hazard input for the USGS Prompt Assessment of Global Earthquakes for Response (PAGER) system [Wald et al. 2006a].

We first provide background on the DYFI system, and then discuss how Internet-based data collection has changed the approach, coverage, and usefulness of intensity observations. We then promote the advantages and note the limitations of online intensity data collection. DYFI is fundamentally a citizen-based science endeavor, and this affords opportunities to both educate and analyze societal response and earthquake awareness. We then discuss how user feedback and scientific considerations have led to a number of important, iterative, and continuing improvements to the DYFI system, and we document these changes. Finally, we describe recently developed tools and analyses of the DYFI data that yield other benefits from online macroseismic intensity-data collection.

INTENSITY-DATA GATHERING IN THE INTERNET AGE

The Community Internet Intensity Map (CIIM, more commonly known as “Did You Feel It?”) system went online informally in California in 1997, and was first fully described by Wald et al. [1999a]. In the subsequent years, significant refinements to the system and its use have ensued. DYFI expanded to the entire USA in 2000, and global macroseismic data collection via DYFI began in 2004.

The CIIM system was an adaptation of a system to process questionnaires collected by telephone interviews, developed by Dengler and Moley [1994] and Dengler and Dewey [1998]. In an effort to process larger quantities of

questionnaires, these researchers recognized the need to assign intensities in a quantitative manner, eliminating the need for exhaustive and subjective assignments of individual intensity reports. They did this by assigning numerical values to answers of individual questions based on the Modified Mercalli Intensity (MMI) questionnaire, and then relating the cumulative sum of the numerical values (weighted differently within varying shaking indicator categories) to independently assigned MMI values. For a specified areal extent (normally, a postal ZIP code), the intensity assignment was done by averaging the numerical values associated with the answers to each question on the questionnaire and weighting different indicated categories to determine a “community weighted sum” (CWS) [Dengler and Dewey 1998]. The CWS value is related to traditional MMI numerically through linear regression and is assigned a Community Decimal Intensity (CDI).

Once Dengler and co-workers had related CWS values to traditional USGS MMI [Stover and Coffman 1993], intensities could be assigned objectively and numerically from their telephone-based macroseismic questionnaires. Dengler and Dewey [1998] referred to the resulting intensities as Community Decimal Intensities (CDIs), to distinguish them from traditional MMIs. Following closely on their work, automating the data collection and processing system to take advantage of the growing popularity of the Internet was the natural course of evolution. Wald et al. [1999a] refined Dengler’s approach by regressing additional questionnaire and historical macroseismic intensity data, expanding the range of applicable intensities to lower and higher values. They also began to automatically and rapidly compute, map, and update the CDI values in what they referred to as a Community Internet Intensity Map (CIIM). Subsequent analyses confirmed the consistency and compatibility of CDI and MMI values (discussed later), and over time CIIM gave way to the more popular reference “DYFI”. The authors initially preferred the use of “CIIM” over “DYFI”, given that data would be collected not just for felt events but also for earthquake disasters, but the DYFI nomenclature appears to be more successful at attracting contributions from non-seismologists.

Of particular note, the Dengler and Dewey [1998] strategy, and the Wald et al. [1999a] Internet-based implementation of it, resulted in decimal rather than ordinal values of assigned intensities. The authors find significant advantages in the use of decimal intensities, though we acknowledge that real-valued representation of intensities also has potential for misuse [e.g., Richter 1958, Musson and Cécic 2002]. We do not know of another group that uses a similar decimal intensity strategy to Wald et al. [1999a], and the alternative approaches discussed below for numerically assigning intensities at a variety of institutions all use logic-based strategies and integer intensity assignments rather than

numerically regressed intensity values. This distinction will become evident in subsequent analyses.

National and international systems

The evolution from manual/postal, to emailed, to web-based macroseismic surveys has been accomplished in many regions of the world. Many countries maintain the manual approach as either the primary strategy, or reserve the option to augment their web-based approaches with traditional assignments. Several very successful Internet-based macroseismic survey systems are now implemented in several countries and regions, and a non-comprehensive survey of a few of these systems follows for comparative purposes.

In New Zealand, Coppola et al. [2010] describe the online system at GNS Science (<http://www.geonet.org.nz>) available since 2004. The GNS web interface, part of the GeoNet project, is an interactive and zoomable interface showing both instrumental and macroseismic assignments. GNS uses what Coppola et al. [2010] describe as a logic-based pyramid strategy to assign intensities, with subsequently higher intensities requiring more ‘positive’ answers than lower values. Intensities are integers, employing New Zealand’s customization of the MMI scale [Dowrick 1996]. The GNS automatic system only allows intensity assignments up to intensity VIII, although Dowrick’s version of the MMI scale goes up to XII. GNS allows web users to retrieve felt reports via interactive queries based on event or date and time.

Sbarra et al. [2010] summarize the Istituto Nazionale di Geofisica e Vulcanologia (INGV) web-based macroseismic collection system for Italy (“Did you feel the quake?”/“Hai Sentito il Terremoto”: <http://www.haisentitoilterremoto.it>), which has been online since 2007. Their 2010 analyses of the web-contributed data in comparison to their traditional macroseismic surveys indicated the reliability of the former strategy. Individual entries are assigned the most probable macroseismic degree (an integer) by statistical analyses of the likelihood of the collection of responses being associated with any intensity. For any entry, they also use the variance associated to the weighted mean intensity in their database to cull entries that appear inconsistent [Sbarra et al. 2010]. In addition, as done with DYFI, entries are removed if their value is outside preset bounds about a selected intensity prediction equation as a function of magnitude and distance; Sbarra et al. remove entries more than three intensity units above or below the prediction. Assigned intensities are averaged for each town or village, unlike the USGS system, for which the community decimal intensity (CDI) results from the average response to each questionnaire question (community weighted sum, see above, and electronic Appendices III and IV). The INGV system supports assignments of both Mercalli-Cancani-Sieberg (MCS) [Seiberg 1930] scale and European Macroseismic Scale (EMS) [Grünthal 1998] intensities. As an effort to try to avoid

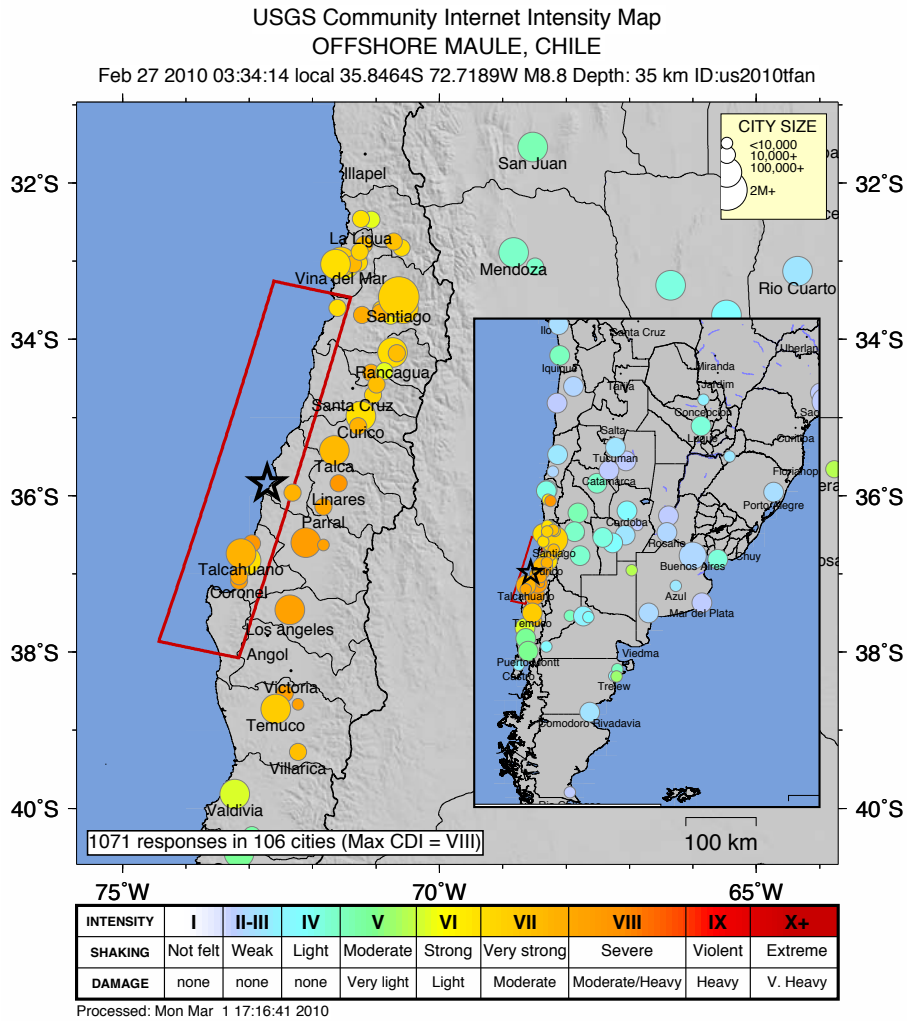


Figure 1. Felt area and distribution of DYFI reported intensities about two days after the February 27, 2008, M 8.8 Maule, Chile, earthquake. The red polygon indicates the surface projection of the approximate rupture area used in the USGS ShakeMap. Intensities do not exceed VIII probably because they are areal averages over the extent of each city rather than over postal codes or districts within cities, and because the subduction rupture surface is over 20 km distant at its closest proximity to coastal cities. The inset shows a smaller-scale map, showing more observations from low-intensity regions.

sampling bias, INGV is also building a large group of spatially distributed volunteers who are alerted to the occurrence of an earthquake; their group is rapidly growing given recent significant earthquakes in Italy.

The European-Mediterranean Seismological Centre (EMSC) has embarked on a project to cover its region with an online questionnaire (<http://www.emsc.eu/Earthquake/felt.php>). The system has been operating since 2004, but has not settled on a permanent solution for automatic intensity assignments (R. Bossu, personal communication, 2010); their logic-based algorithm, written by Roger Musson at the British Geological Survey (BGS), is being analyzed and modified by Sebastian Gilles at EMSC. The EMSC is multilingual, translated into 32 languages. Currently, it provides integer EMS assignments for cities with a minimum of five responses [Mazet-Roux and Bossu 2010].

Additional systems are in place at the British Geological Survey, Le Bureau Central Sismologique Français, Natural Resources of Canada, the Swiss Seismological Service, and the Royal Observatory of Belgium, among several others.

These systems employ a variety of approaches, but both the Canadian and Belgian approaches employ the same questionnaire as DYFI (Wald et al. [1999a]; Appendices I-IV, this report).

As can be inferred from the above descriptions of various national systems, there is very little uniformity among different groups' choices of the macroseismic scales, the nature of the questionnaires, and in the approach for assignments of intensities from these forms. (For a good comparison of macroseismic scales, see the review by Musson et al. [2010], and its supplemental materials.) Recognizing the importance of uniform data collection, the ESC Working Group for Internet Macroseismology was established at the 2008 General Assembly of the ESC in Heronissos, Crete. The Working Group was charged with developing common methods for collecting and disseminating macroseismic data using online methods [Musson 2010]. A primary goal of that group, to which the present authors contribute, is to determine a suite of internationally agreeable questionnaires and an exchange

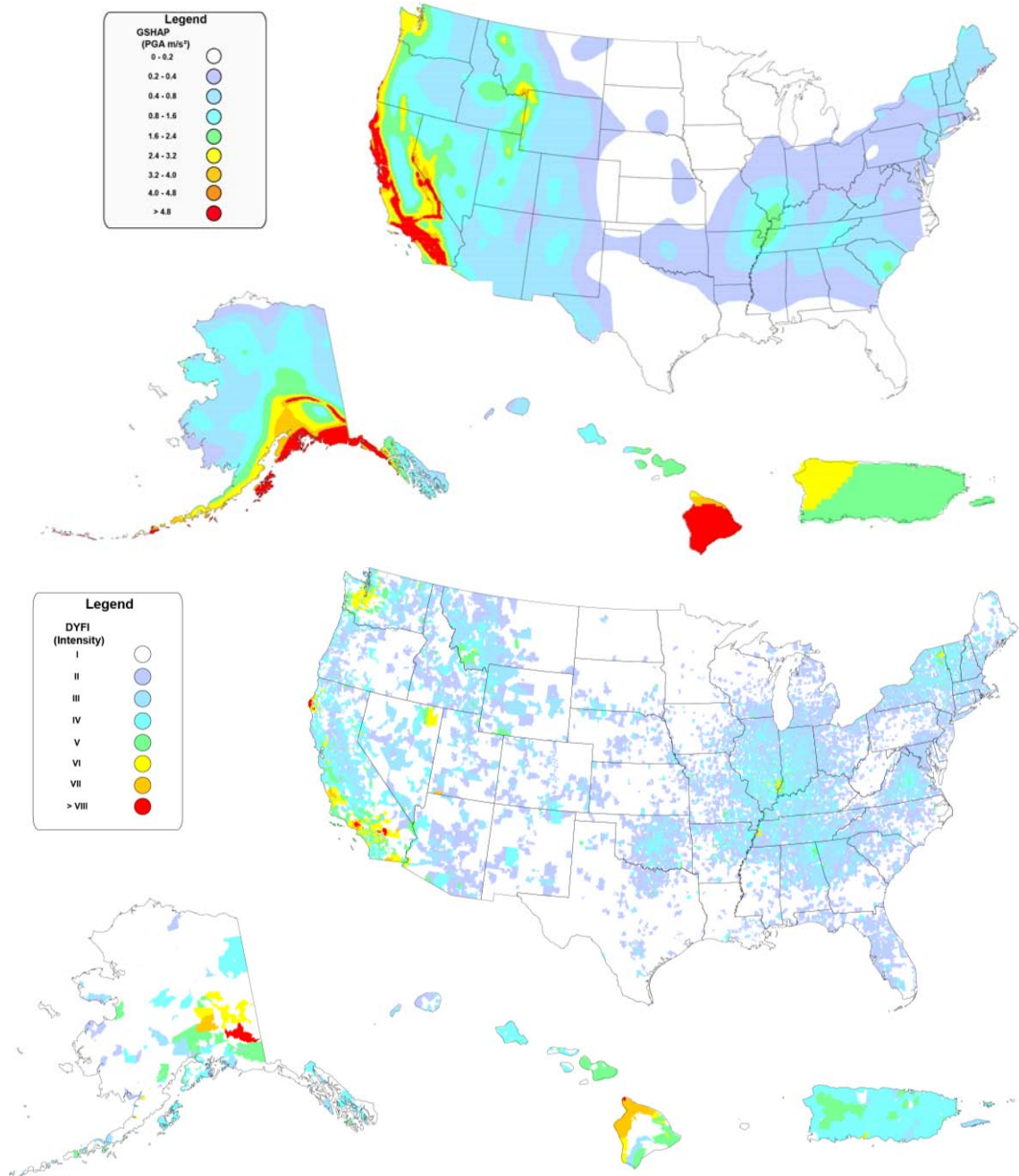


Figure 2. Comparison of the USGS National Seismic Hazards Map (NSHM) [Petersen et al. 2008], top, with a decade of DYFI responses (1999-2011), bottom. For each postal ZIP code, the maximum intensity reported during that time period is shown. DYFI intensity color coding is the standard intensity palette used by USGS for ShakeMap and DYFI; for NSHM the 10% probability of exceedence of peak acceleration in 50 yrs ground motions is scaled approximately to the DYFI color palette. During this period there were over 1.6 million individual responses in over 25,000 ZIP code areas.

format for distribution. It is anticipated that processing of the responses can still be left to different strategies unless a Global Macroseismic Scale (GMS) is established.

Advantages of online macroseismic data collection

Traditionally, intensities are assigned by a classification process and are assigned integer values (or, less frequently, a range of integer values such as “6-7” or “>6” for less certain assignments; see Grünthal [1998], Section 4.5). The CIIM process of assigning numerical values to macroseismic observations and then calculating real-valued intensities thus represents a philosophical as well as a procedural departure

from traditional intensity-assigning methods. Over the years, analyses of the DYFI data strongly suggest that reporting intensities to higher precision is warranted [e.g., Atkinson and Wald 2007, Wald et al. 2006b, Worden et al. 2011]. USGS now carries a single decimal place, and the discrete (I to X) intensity scale is replaced with a continuous, real-valued, scale. When the DYFI and similar decimal intensities for widelyobserved earthquakes are plotted as a function of distance, the retention of information to tenths of an intensity unit results in lower scatter than is the case when intensities are assigned by classification and reported to the nearest integer [e.g., Dengler and Dewey 1998, Dewey et al.

2002, Wald et al. 2006b]. For map representation of decimal intensities, a continuous palette of colors is chosen using algorithms that automatically interpolate between discrete color values, using a similar technique to the one used by ShakeMap [Wald et al. 1999c]. This allows us to present more subtle variations of intensity than previously achievable.

Designed to work in conjunction with rapid epicenter and magnitude determinations that are provided by regional and national seismic networks, the DYFI system is now triggered automatically; individuals can respond to and view maps for a particular earthquake within minutes and watch as maps are continuously updated with new data (every 1-2 minutes). Within the USA, intensity observations are grouped, averaged, and plotted according to postal codes. Postal ZIP code polygons are color-coded according to their computed intensity values. We can now also automatically geocode the users’ locations if they correctly provide their street addresses (most do), and we can thus provide refined, spatially aggregated intensity assignments as needed.

Starting in 2004, we implemented the DYFI system for international data collection, with observations grouped, averaged, and plotted by city. Outside of the USA, we enable users to select their country and city from pull-down menus. Thus, the resolution for automatic intensity assignment outside of the USA is at the level of individual cities, which

we color code to the intensity value and map as a circle (Figure 1). Currently we have approximately 140,000 cities in our database, which was culled from the open-source Geonames database [Geonames 2011]. As of this writing, we are augmenting the global city-based system with user-geocoded locations.

Rate of responses

Data for widely felt earthquakes come in at a rate that allows confirmation of the earthquake’s occurrence within minutes. The DYFI maps evolve rapidly to informative macro-seismic intensity distributions that are useful to the public, media, scientists, and even emergency managers typically within tens of minutes. Statistics attest to the abundance and rapid availability of these Internet-based macroseismic data: nearly two million entries have been amassed over the decade (Figures 1 and 2); there are 33 events each with more than 10,000 responses; 250 events have over 1,000 entries (Table 1). The greatest number of responses for an earthquake is more than 77,000 for the April 2010, M 7.2 Baja California, Mexico, event. Table 1 summarizes some of the notable statistics associated with DYFI data collection.

Outside the USA, DYFI has gathered over 145,000 entries in 7,600 cities covering 192 countries since its global

Within the USA and territories

Top 3 earthquakes by number of entries	77,758 entries for M 7.2 near Baja California (April 4, 2010) 58,362 entries for M 5.0 near Ontario-Quebec Border (June 23, 2010) 47,763 entries for M 4.7 near Los Angeles, CA (May 18, 2009)
Number of US postal ZIP codes with entries	25,909 (out of a total of 41,558)
Total Number of Individual Entries	1,603,100
Largest rate of responses	73/s for M 5.7 Southern California (June 15, 2010) 51,624/h for M 5.6 Northern California (Oct 31, 2007)

Outside of the USA

Number of cities with entries	7,603 (out of 145,286)
Total number of individual entries	140,263
Number of countries with entries	192 (out of 234)
Top 3 cities by number of entries	Santiago, Chile, 6,690; Toronto, Canada, 5,755; Melbourne, Australia, 3,250
Top 3 countries by number of entries	Canada, 25,859; Chile, 14,839; Japan, 10,566
Top 3 earthquakes by number of entries (outside N. America)	M 5.0 England, U.K., (February 27, 2008), 19,044 M 4.7 Australia, (March 6, 2009), 3,794 M 5.5 near Spain (December 17, 2009), 3,298.

Table 1. Sample USGS “Did You Feel It?” Statistics (as of July 2011).

inception in late 2004 (Figure 3). The rapid intensity data are automatically used in the Global ShakeMap system (GSM) [Wald et al. 2006a], providing intensity constraints near population centers (Figure 1) and in places without strong-motion instrument coverage (most of the world), and allowing for bias correction to the empirical prediction equations employed in ShakeMap. In practice, we automatically incorporate DYFI intensities into GSM at present for a minimum number of responses per location (currently set at three or more). ShakeMap has also been recently refined to automatically use macroseismic input data in their native form, and treat uncertainties rigorously in

concert with the more standard-use recorded ground-motion data (ShakeMap Version 3.5) [Worden et al. 2010]. DYFI contributions to GSM have two important aspects. First, they provide ground truth intensity assignments, predominantly at sites with significant populations (such as cities). Second, with 5-10 intensity assignments in the near-source area, these data can allow GSM to compute a bias correction term to be applied to the ground-motion-prediction equations, effectively removing the inter-event variability, or correcting for an incorrect initial magnitude calculation [Wald et al. 2006a, Worden et al. 2010, Worden et al. 2011].

The Internet makes it possible to rapidly gather larger,

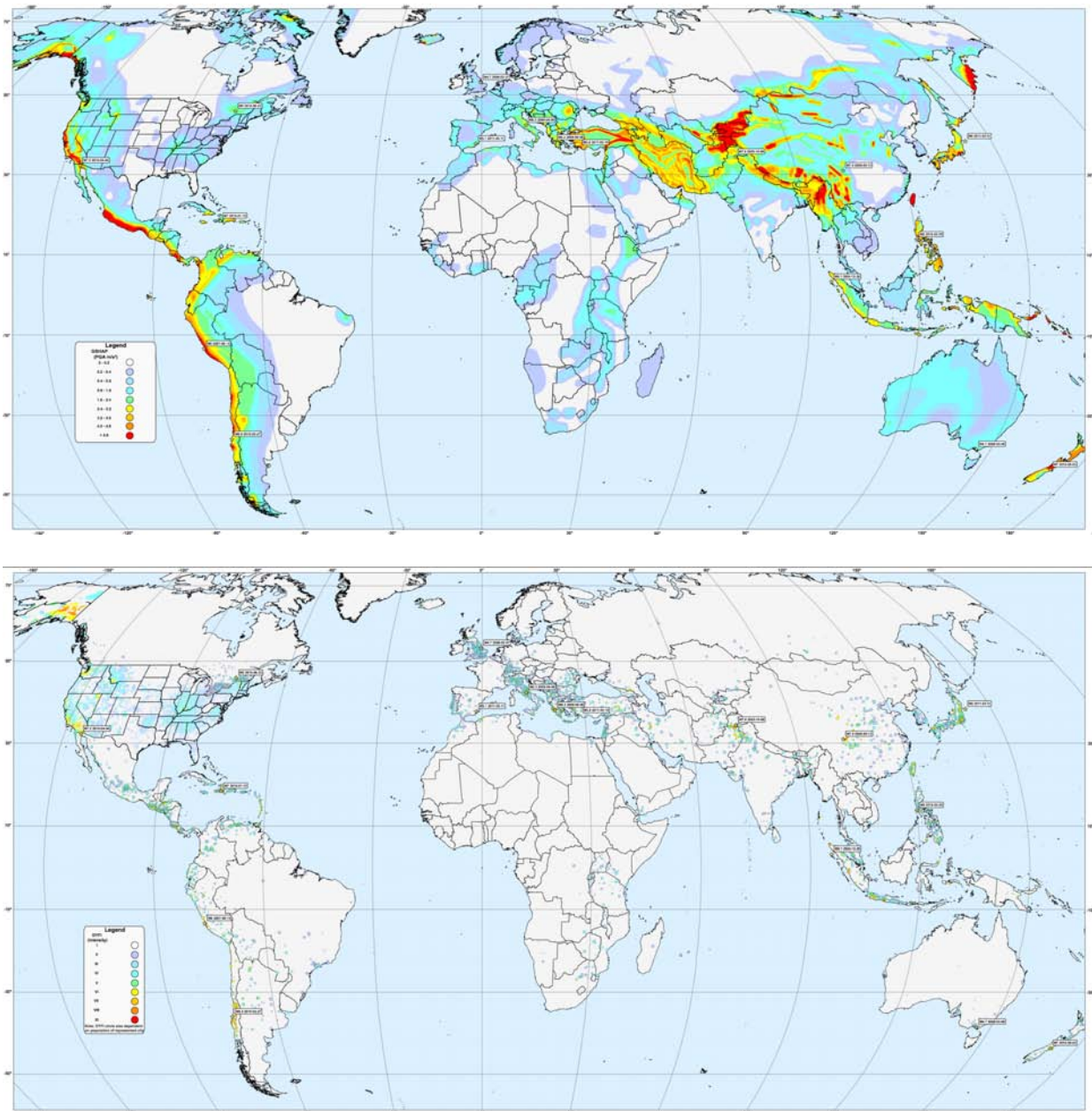


Figure 3. Comparison of the Global Seismic Hazards Map (GSHAP) [Giardini et al. 1999], top, with six years of DYFI responses (late 2004-2011), bottom. For each city or postal ZIP code, the maximum intensity reported during that time period is shown. DYFI intensity color coding is the standard intensity palette used by USGS for ShakeMap and DYFI; for GSHAP the 2% probability of exceedence of peak acceleration in 50 yrs ground motions is scaled approximately to the DYFI color palette. Earthquake dates and epicenters are shown for selected significant events. During this period there were 140,000 individual responses in over 7,000 cities outside of the USA.

more comprehensive datasets than ever before, and at minimal cost. Prior to this system, intensity maps were rarely made for US earthquakes of magnitude less than about 5.5; now intensities are routinely reported for the smallest felt earthquakes nationwide (Figure 2). In addition, thousands of reports are available for moderate to large events, often tens of thousands for those in densely populated areas. The greatly expanded datasets allow for post-processing and analysis in ways that were not before possible. To date, over 250 earthquakes having over 1,000 entries present a substantial data resource for portraying shaking distributions and for quantitative analyses.

Figure 2 shows the intensity of shaking reported over the entire USA for more than a decade (Figure 2, bottom panel). As far as we are aware, nearly every felt earthquake is or can be reported now (although separating contributed reports for multiple shocks occurring close together remains problematic). Thus, this map represents something we have never been able to show before: the actual distribution of shaking intensity over the entire nation for a decade. This is actually quite an extraordinary contribution of the DYFI system. In comparison with USGS National Seismic Hazard Map (Figure 2, top panel), one can analyze the overall consistency in many areas, and identify those that have not experienced events in a decade despite high longer-term probabilities of shaking. In the comparison in Figure 2, we have scaled the thirty-year, 2% exceedance probability in terms of peak acceleration to the same range of intensities (colors) as for the MMI scale used for DYFI maps. While these maps should not be compared quantitatively, the

qualitative comparison essentially implies that the range of likelihoods of ground motions span the range of intensities. As discussed below, such maps provide significant earthquake awareness and educational opportunities. Figure 3 provides the analogous comparison for the global DYFI data over a six-year period (2005-2011) with the Global Seismic Hazard Map (GSHAP) [Giardini et al. 1999]. Rates and completeness of responses, as well as the total time period, are not as impressive as for the USA, but nonetheless the opportunity for comparison with a long-term hazard maps is a good one, and it will improve with time.

The impressive rate of responses and feedback from users prompted us to routinely plot entries contributed as a function of time (Figure 4). Questionnaire response rates have reached 62,000 per hour (ca. 1,000 per min; Figure 4), requiring substantial web resource allocation and capacity (see Appendix 1). These plots are provided online for each event, and they show logical patterns of immediate post-earthquake surges followed by decays; late-night lulls followed by morning surges. Occasionally, aftershocks and media attention result in belated surges. Continuous plots of the entry rates allow operators to track system performance and gauge future bandwidth requirements.

The data quality and quantity depend primarily on population density and prevalence of Internet access, but not necessarily on earthquake awareness or the overall hazard of the region. Surprisingly, events in the eastern and western USA have comparable response rates, despite significantly different historical rates of earthquake occurrence in the two regions (e.g., examples in Figures 5, 6, 7).

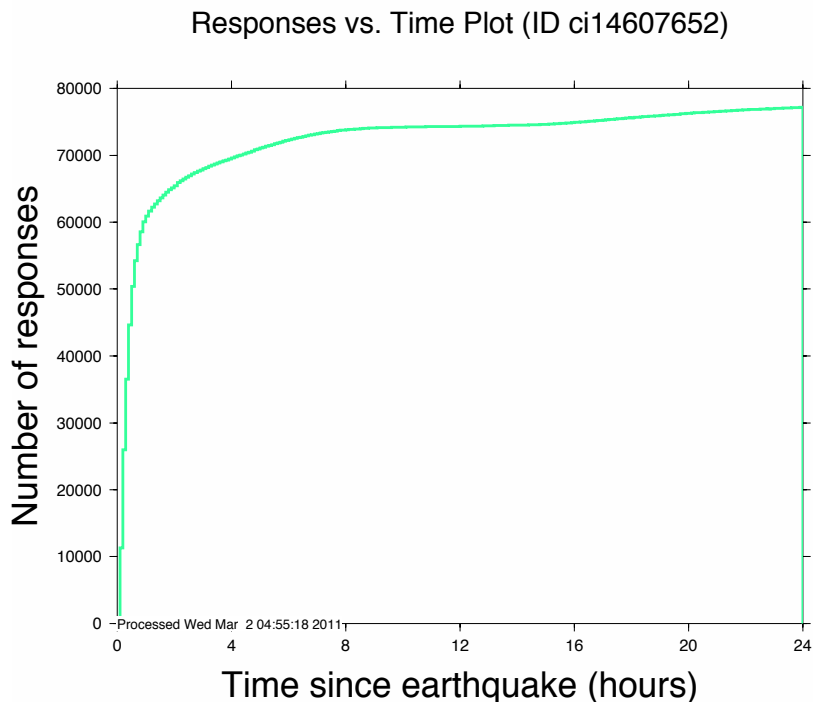


Figure 4. Plot of individual questionnaire responses versus time for the April 2010, M 7.2 Baja California, Mexico, event. Over 62,000 entries were received in the first hour, or roughly 1,000 per minute. Similar response plots are generated for every event. The earthquake occurred at 15:40 local time.

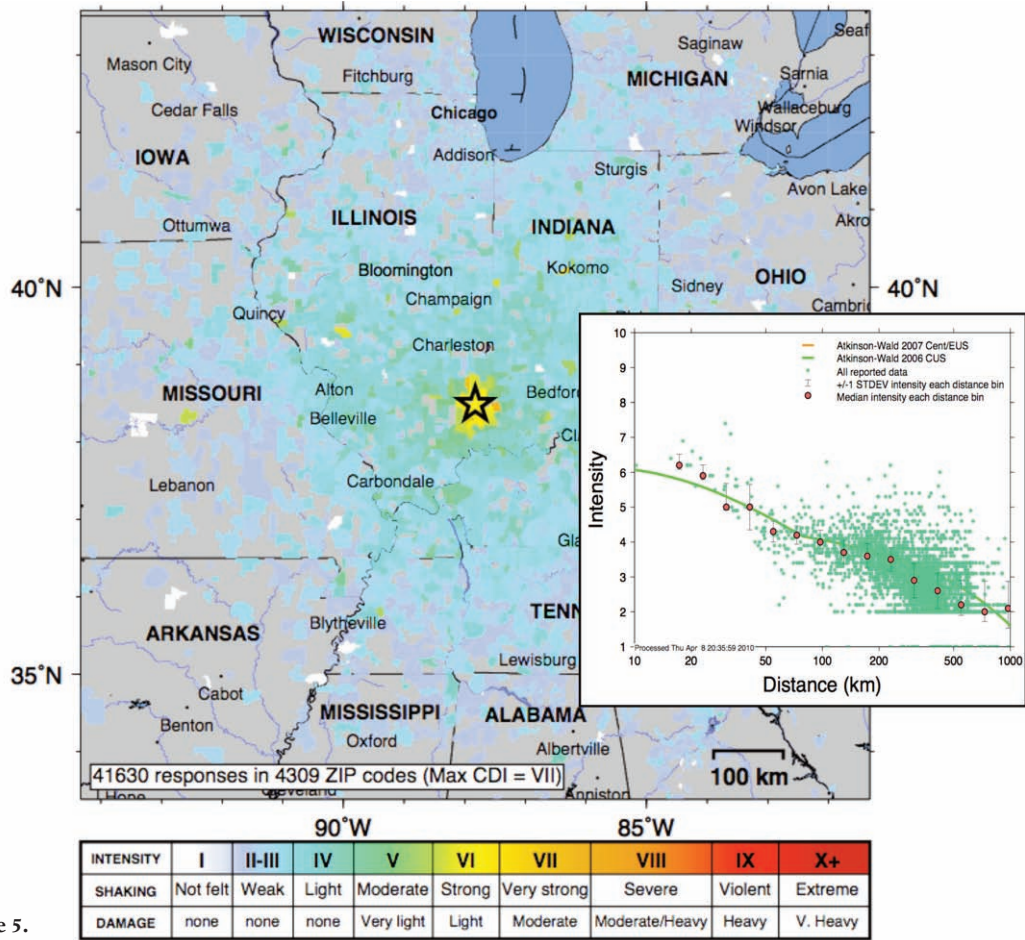


Figure 5.

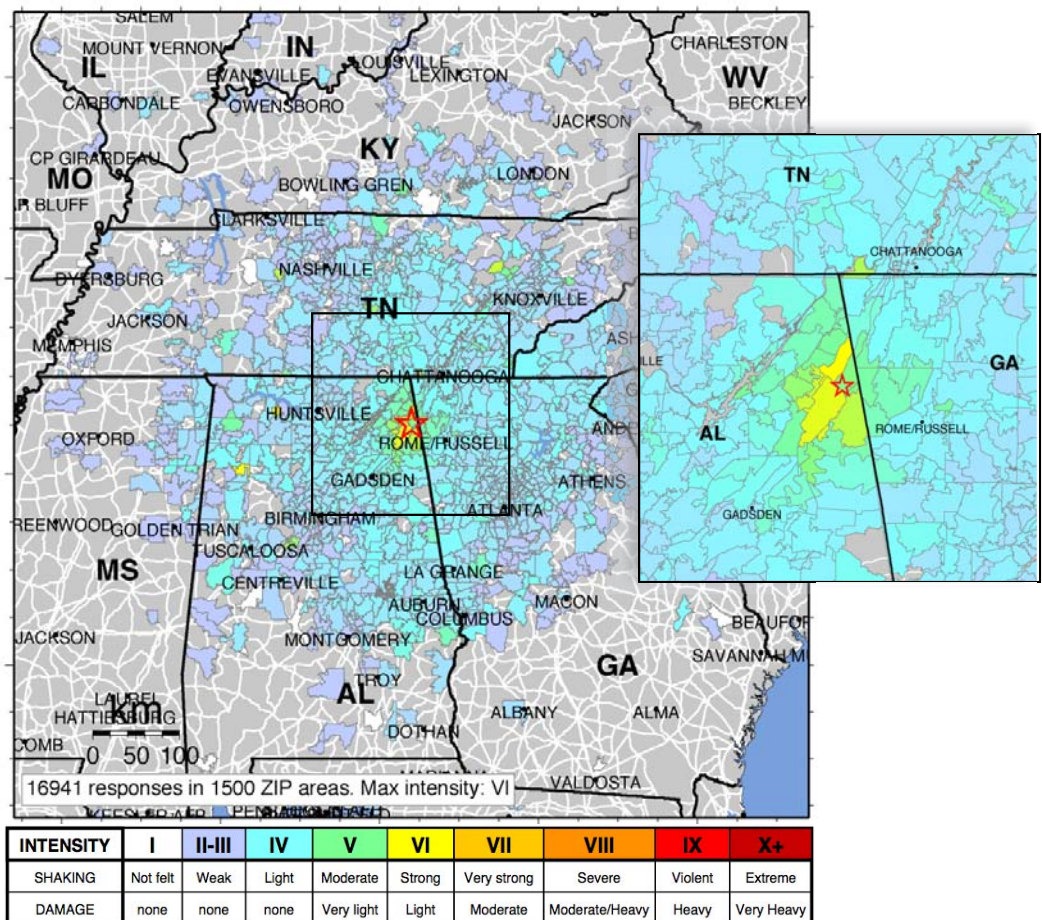


Figure 6.

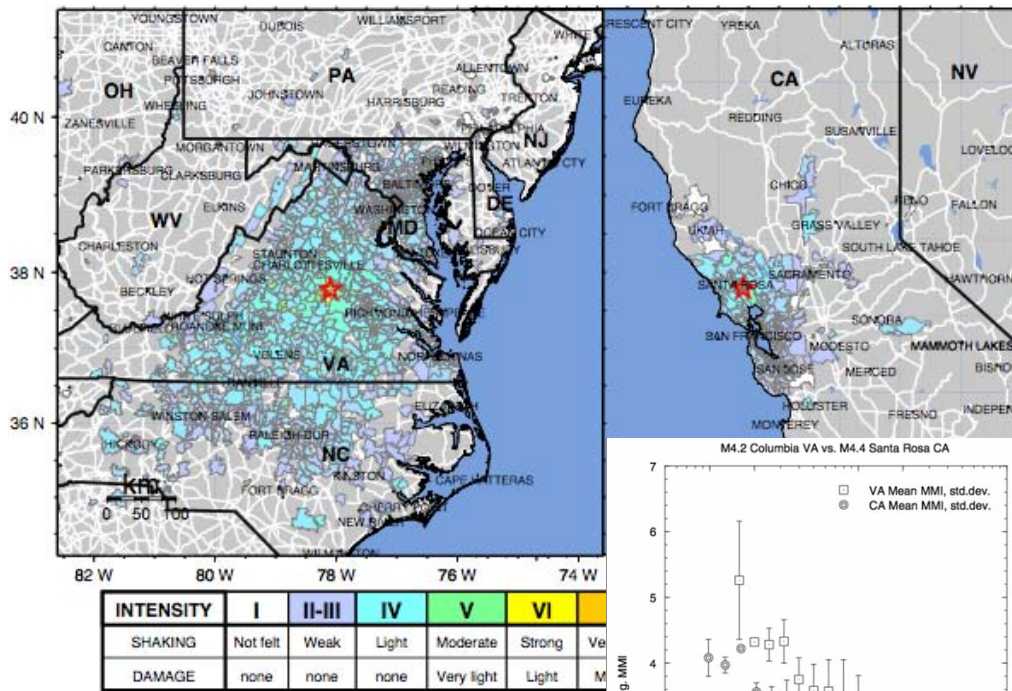


Figure 7. Comparison of felt area and intensity–distance attenuation for the December 9, 2003, M 4.2 Columbia, Virginia, earthquake (left) with the August 2, 2006, M 4.4 Santa Rosa, California, earthquake (right). Note the dramatic difference in the overall felt area and difference in epicentral intensity. Maps scales are approximately the same. The inset plot shows DYFI intensities as a function of distance for the two events. Symbols show mean and standard deviation. Slight offset of distance values used for California (CA) dataset for plotting clarity [after Atkinson and Wald 2007].

Figure 5 (top of previous page). Felt area and distribution of DYFI reported intensities for the April 18, 2008, M 5.2 Illinois earthquake. The inset shows the decay of intensities with hypocentral distance: green dots are ZIP code intensities, red circles are median intensities in log-distance bins, and the green line is the prediction equation of Atkinson and Wald [2007]. The prominent inflection at a distance of about 100 km is typical of most central and eastern USA earthquakes and results from amplification from post-critical reflections from the Mohorovičić discontinuity.

Figure 6 (bottom of previous page). Felt area and distribution of DYFI reported intensities for the April 29, 2003, M 4.6 Alabama earthquake. Inset figures show smooth variations of intensities near the epicenter for the rectangle on the main map (top), and smooth attenuation with distance (green curve, bottom), consistent with Atkinson and Wald’s [2007] predictions; red dots are distance bin averages of individual entries (green dots). There were 16,941 individual responses for this earthquake in 1,500 ZIP code areas. The prominent inflection at a distance of about 100 km is typical of most central and eastern USA earthquakes, and results from amplification from post-critical reflections from the Mohorovičić discontinuity.

Accuracy of responses

Given that DYFI intensities are assigned by responses to questionnaires by the general public, one might expect them to be subjective, or even unreliable. However, the large numbers of responses from most communities make the resulting intensity approximations surprisingly robust. Qualitatively, for the most part, DYFI intensity maps show smooth variations of intensity in areas from which there are many observations (e.g., Figures 6, 7, 8). DYFI macroseismic intensity maps also generally agree with instrumental intensity maps (ShakeMaps) that are based solely on seismographically measured peak ground motions, at least in areas where both can be made with sufficient data-sampling density (e.g., California). Examples of earthquakes with both human-and instrumentally generated intensity maps abound on the USGS ShakeMap and DYFI websites.

There is a wide variety of potential sampling biases associated with any macroseismic survey. With the DYFI

system, we know from comparison with the traditional USGS MMI, that at the low end of the scale (I-III), Internet-based questionnaires actually reach a greater potential sample than postal surveys made in parallel, and thus felt reports can be more widely distributed than traditionally sampled MMI [Dewey et al. 2002]. At middle range of intensities (IV-VII) we see of no obvious differences between traditional MMI and DYFI. DYFI limitations and sampling biases at the high end of the scale are discussed in a later section.

More quantitatively, Worden et al. [2011] used the large volumes of DYFI data available in California to show that the standard deviation for ten responses at a given location is about 0.25 intensity units (Figure 8); additional responses lower repeatability differences to less than 0.10 units. Since adjacent integer intensity levels correspond to roughly a factor of two increase in peak ground motion [Wald et al. 1999b, Worden et al. 2011], the fact that maps show intensities that vary smoothly between integer values means that a

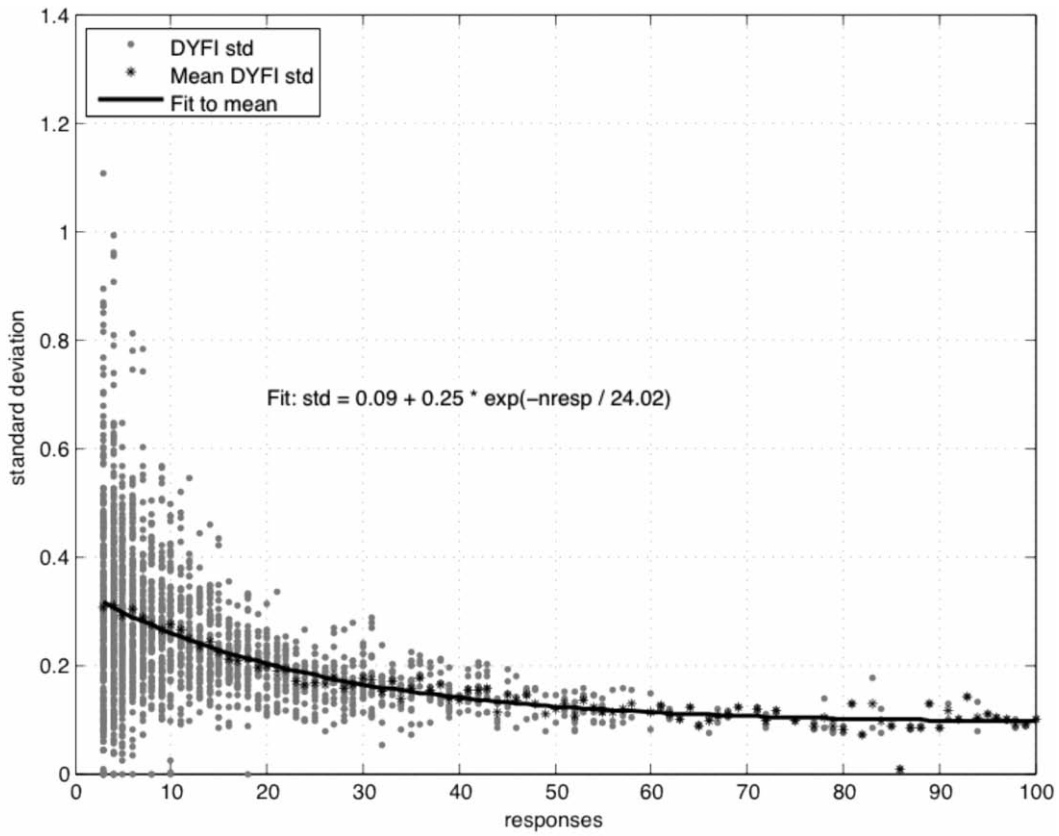


Figure 8. Standard deviation of bootstrapped DYFI responses as a function of the number of responses within an area of 2 km radius about a ground motion observation. The mean within each bin (binned by number of responses) is shown, along with an exponential function fit to the means. From Worden et al. [2011].

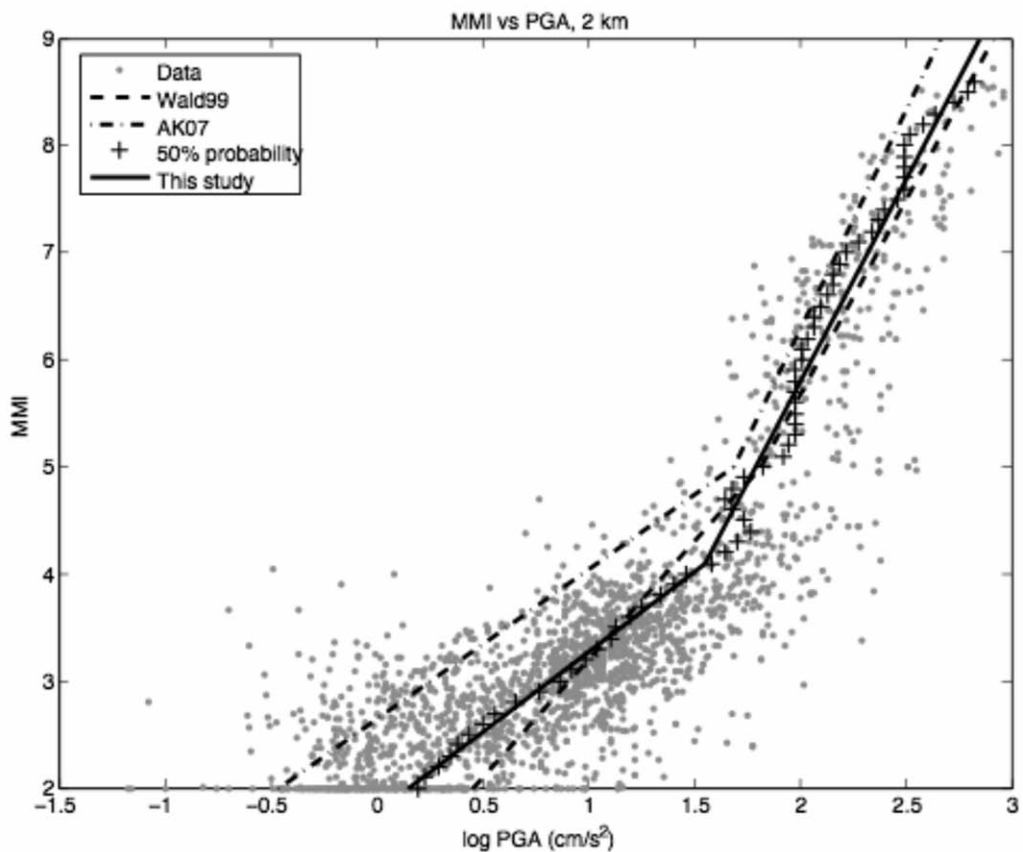


Figure 9. The PGA vs. MMI bilinear total least squares fit to the 50% probability picks (shown) is overlain upon the PGA/MMI pairs. The Wald et al. [1999b] and Atkinson and Kaka [2007] relations are also shown. From Worden et al. [2011].

‘community’, at least on average, is capable of distinguishing rather small differences in ground motions, potentially within about $\pm 20\%$. From a seismological perspective, this observation is rather remarkable: a consensus opinion of the general public (e.g., human perceptions) on ground shaking, as determined from average responses to a series of simple questions, implies the assignment of the *absolute level* of shaking to considerably better than a factor of two, and repeatedly to within about 20% with sufficient numbers of contributors. Worden et al. [2011] detail the process and uncertainties associated with directly transforming DYFI intensity to peak ground motion estimates, but based on the repeatability seen with these data, we consider that process to be particularly informative for a number of real-time (ShakeMap and magnitude determinations) and other seismological analyses.

The uncertainties that we estimate for macroseismic intensities may seem surprising. For traditionally assigned intensities, disagreements about integer assignments can occur due to different experts’ subjective evaluations, differences in the observations included at a particular location, the assumed extent of an area used for the assignment, and many other reasons. For example, EMS-98 [Grünthal 1998], the most comprehensive macroseismic scale to date, can produce fundamental ambiguities in integer assignments if damage grades for two different structure types indicate conflicting values. Given the limitations of intensity assignments (summarizing multiple observations over an unspecific area), the numerical consistencies and low scatter of DYFI assignments found by Worden et al. [2010] and seen in the smooth intensity gradients on the maps shown herein (Figures 5, 6, 7), are rather remarkable. That said, there is now general agreement within the loss-modeling communities that highly detailed damage assessments for modern earthquakes require not only aggregated shaking and impact assignments implied by macroseismic data, but also comprehensive, detailed geospatial collections of individual structures, their structural descriptions, and their damage states [e.g., Coburn and Spence 2002].

In addition to uncertainties in intensity assignment, there is also the uncertainty in the spatial location and extent of our users’ observations. DYFI users are asked to provide their ZIP code (or city, for non-US locations), and also their street address for geocoding (see “Geolocation” below). ZIP code extents could be less than a kilometer across (for densely populated cities) or tens of kilometers for sparsely populated areas. For non-US events, a user’s “city” can be an area of several kilometers or more. On the other hand, geocoded locations (when available) are accurate down to the scale of city blocks.

As discussed previously, DYFI intensities are an aggregate of user responses over a certain area, rather than point observations. Thus the spatial resolution of our

intensity maps and the scale at which they are meaningful depend on the spatial precision of the underlying data. For widely felt events outside the USA, we make maps with intensity assignments at the city level; for small events felt within populated US urban areas, we can create maps with neighborhood-scale intensity levels, if there is a sufficient number of responses.

Catching missed events

DYFI has proven to be of particular value for the USGS’s National Earthquake Information Center (NEIC) when earthquakes occur in areas of the USA that have sparse seismographic stations. Some earthquakes that are identified with DYFI macroseismic data often have not been automatically detected by domestic seismographic networks, yet their locations can subsequently be confirmed by analysts with scrutiny (and the timing and location guidance provided by DYFI) and thus their magnitudes determined with seismographic data from several stations. Likewise, the DYFI system routinely provides an immediate heads-up for destructive events around the globe within minutes of the event. Unlike more informal information flowing from social media outlets, which are difficult to quantify, the rapid DYFI reports provide data on location, time, and valid macroseismic values; these data can be used both qualitatively and quantitatively.

Remarkably, first-arriving intensity observations from earthquakes often precede the seismic network automatic determination of earthquake location and magnitude. We refer to these earliest, as-yet unassociated responses as “unknown event” responses, since at the time we receive them we do not have other information on the events that produced the reports. We monitor the unknown event reports by email and algorithmically. When the number of reports from a region exceeds 20 in a five-minute period, we alert operators and seismic analysts of a likely event (times, locations, and intensities). Such a rate of unknown event reports is almost inevitably associated with an earthquake from which we have not yet received seismograph-based source information. Once the event is located, the earlier unassociated reports are associated with it based on appropriate space-time windows.

With DYFI the USGS is now capable of monitoring and automatically collecting intensity data for *all felt earthquakes* in the United States. That is a remarkable statement. Events of magnitude less than 2.0, well below routine reporting level for most U.S. domestic seismic networks, are not uncommonly reported on the DYFI website. What’s more, DYFI can capture felt reports for non-earthquake-related shaking; DYFI maps are routinely made for sonic booms from supersonic aircraft, bolides, and even mining and other explosions. Thus, DYFI allows comprehensive, quantitative analyses of macroseismic events that would otherwise go

uncollected, and had for decades prior.

For example, some very widely felt events, such as the unusual magnitude 6.0 event of 10 September 2006 in the Gulf of Mexico (responsible for many of the felt reports in the Gulf Coast region, Figure 2), would not have been canvassed in the past due to the overall low intensities of shaking. With DYFI, over 6,000 entries were tallied, providing unique intensity data for this normally seismically quiet region. In that particular case, shaking likely exceeded the 50-year shaking probabilities shown by the USGS National Seismic Hazard Map (Figure 2a). Likewise, the number of DYFI responses generated by a descending space shuttle passing over Los Angeles warranted a separate event webpage dedicated to the reports, and the reports were sufficient to map out the re-entry trajectory. Sonic booms from aircraft, particularly in earthquake country, rattle residents and rattle nerves, and yet prior to DYFI, most could not be confirmed, since seismic networks do not routinely locate and report sonic events. DYFI reports of sonic booms provide quantitative data for long-observed characteristics that are diagnostic of the phenomena: spatially and temporally correlated low-intensity reports from a rather broad geographic area, the *lack* of an associated instrumentally recorded earthquake that is large enough to be felt over the area from which reports are received, and the nature of user entries: rattled windows are noted nearly ubiquitously in the response or comments provided. USGS's ability to document and map the occurrence of such events serves a public service.

Uniformity and flexibility

One long-term objective of the USGS in gathering Internet-based macroseismic data was to do so uniformly throughout the USA. This was particularly important as regional networks and universities began collecting email notifications of earthquake effects by employing a wide variety of questions that only approximated those required for systematic and official intensity assignments. Now, almost all Internet-based macroseismic data collection in the USA is done via the USGS DYFI portal, providing uniformity of intensity assignments and, critically, a single collection point and long-term archive of these data. Regional network operators linking their felt reports to DYFI benefit directly by not having to support local macroseismic data collection and processing efforts, yet can receive DYFI data as it is generated by USGS. We also customize event triggering and provide automatic notifications of felt events for the regional seismic network operators.

Interactivity and citizen-based science

Millions of annual web visitors and hundreds of thousands of individual questionnaire entries are a testimonial to the outreach potential and benefit of DYFI. DYFI also provides a unique opportunity for citizen-based science. In

some scientific disciplines, data collection has long depended on the citizenry, e.g., backyard rain and temperature measurements for the US National Weather Service, and the National Audubon Society's hundred-year plus annual Christmas Bird Count; both have provided long-trending, spatially distributed data collection not possible by groups of scientists or instruments in the field.

More recently, a number of other systems have been developed with similar 'citizen science' applications, allowing scientists to obtain vast data collections in ways not previously or otherwise possible. One example is Cornell University's bird-surveying program Project Feeder-Watch (<http://birds.cornell.edu/pfw/>), in which citizens contribute via Internet a wealth of observations that help constrain migratory bird habits, quantify variations in population density, and establish species ranges. Another example is the CoCoRaHS, a grassroots aggregator of backyard weather observations, which allows citizen to report rain, hail, and snowfall data online (<http://www.cocorahs.org/>). Originally serving Colorado, over 5,000 CoCoRaHS observers now report valuable observations daily from all 50 US states. Such observations would not be possible without this informal organization of thousands of volunteer 'citizen scientists'. The Internet facilitates the rapid and extensive collection of such datasets. What's more, automatic data collection and processing provides instant feedback to the participants, or 'field assistants', further connecting science to citizen and creating a feedback loop for further participation. Participatory science not only expands the observational base for data collection, an obvious advantage for intensity observations, but it also empowers the community to take ownership and allow better understanding of important scientific issues of the day [e.g., Trumbull et al. 2000]. The DYFI system takes full advantage of online data collection by being fully interactive, providing users' intensity assignments instantaneously, and by showing the effects of their entries on the updated intensity maps.

An important subtlety with respect to DYFI contributors is that they are not trained for the task, and as such they are not 'citizen scientists'; rather, DYFI is more accurately described as 'citizen-based' science. Other citizen science systems do, in fact, require substantial expertise on the part of the contributor, for example, being able to recognize the species of birds that their feeder attracts for Project Feeder-Watch, or knowing the phenophases of flowering plants for the USGS Phenology Program (<http://www.pwrc.usgs.gov/bpp/>). Yet the well-established nature of macroseismic collection, with exception of the highest intensities, has been reliant on casual observers for many decades; only of late has the Internet facilitated the collection of these data more rapidly and systematically.

The DYFI system also educates the public on oft-misunderstood seismological concepts like the spatial

variations of shaking intensity, and it provides a basis for clearing up confusion of macroseismic intensity with instrumental magnitude. As described in a later section, we have built educational tools that further this goal. Greater awareness of these concepts allows for more rational decision-making for both long-term mitigation and emergency response in the immediate aftermath of an earthquake [e.g., Goltz 2003]. DYFI also provides an important human perspective on earthquakes, providing sociological documentation of the way people behave and respond, and how they perceive risk [e.g., Celsi et al. 2005]. Finally, and perhaps most rewarding to the authors, DYFI seems to provide emotional help to citizens who have just had a frightening or even traumatic experience. By allowing citizens to share their experiences and by enabling them to contribute their observations towards a general public understanding of the phenomenon they have experienced, DYFI provides many with a form of catharsis at an opportune time, and in a rapid fashion. Often users describe the desire to confirm their experience with others, or in the case of DYFI, with the collective community.

In short, the DYFI website provides a two-way information conduit, since citizens coming to the USGS for information are empowered to become data providers themselves by contributing valuable observations that benefit the USGS as well as the observers, their local communities, and earthquake responders.

Challenges and limitations of online macroseismic data collection

The voluntary nature of macroseismic questionnaires collected online and the open-environment character of the Internet itself, pose problems in data collection that force us to continually modify aspects our DYFI processing system. Likewise, we must also retain realistic expectations of the performance of the DYFI system in the immediate aftermath of moderate magnitude events in areas with poor Internet coverage, and different expectations for severely destructive earthquakes. Limited access to the DYFI system due to damage and assignments of high intensities with DYFI both require special consideration.

With DYFI, the highest contribution rates to date have come from events with a combination of dense population, ubiquitous Internet access, and a *lack* of significant damage, power or Internet outages, or otherwise disrupting influences. It is for these events that we have slowly grown processing capacity at the USGS. In contrast, and without specifics, several countries’ webpages have had near-immediate loss of service even after relatively minor but widely-felt events, and systems like DYFI can certainly be largely culpable in contributing to that traffic spike without proper precautions. Several mitigating strategies are outlined in the section below on robustness and system specifications.

For very widespread damaging events, which to date have not been experienced in the USA (where response rates are normally the highest) since DYFI’s inception, we anticipate that response-related traffic will be limited by some of the disruptive factors mentioned above. In addition, providing maps for emergency response was never the intended main goal or service of the DYFI system. Rather, where risks dictate and resources permit, the procedure of choice for more robust post-earthquake hazard evaluation is the ShakeMap system [Wald et al. 1999c]. ShakeMap, in conjunction with the related, downstream systems like ShakeCast and PAGER [e.g., Wald and Bausch 2009], provides situational awareness to inform and initiate appropriate response activities. As a tool for earthquake response, DYFI has several severe limitations immediately after a major earthquake. In the high-intensity regions we cannot expect Internet and power connections to be available; equipment may be damaged, and there is no expectation that getting online to report to the USGS should or will be a priority for citizens. In contrast, ShakeMap depends on hardened (yet not overly robust) seismic networks, often with redundant or alternate communication channels; it does not rely on the availability of Internet users.

For regions with sparse strong-motion seismographic coverage, DYFI does provide a partial substitute for instrumentally generated ShakeMaps, again with the caveat of a lack of robustness for serious events. That said, DYFI macroseismic maps *are* more representative of earthquake felt areas and impacts than ShakeMaps. ShakeMap, with intensity derived from conversion of recorded ground motions [e.g., Wald et al. 1999c, Worden et al. 2011], serves only as a proxy for what macroseismic effects actually occurred; DYFI directly reports them and thus *defines* the macroseismic intensities.

Outliers

Automated data collection from the public via the Internet inevitably results in data outliers, some unintentional and others deliberate. We have developed two automated filters that remove the bulk of such outliers. No data are discarded in the process of filtering; rather, outliers are flagged as “suspect” and bypassed in the processing of results.

The first filter works on checking the self-consistency of each individual entry. We reject responses that have contradictory answers, such as a response of “not felt” and “frightened”. In particular, the damage portion of the questionnaire (see Appendix I) is the one most likely to be falsified; for some reason, checking all possible damage options is not uncommon, yet it is an invalid selection.

The second filter is based on the individual intensity computed from each entry and the distance of that user from the epicenter. In the past, this was compared to a simple, empirically derived intensity-distance curve that was

based on the event's magnitude. Entries whose computed intensity was more than two intensity units above the empirical curve were rejected. Now, we use Intensity Prediction Equations or IPEs [e.g., Atkinson and Wald 2007], which are regionally dependent and were derived using archived DYFI data explicitly.

Aside from deliberate false entries, we also get simple user errors, like typos. Most users show genuine concern over the potential impact of their errors, often sending email to the operators when they realize that they have inadvertently caused an error (e.g., entering their home rather than their work ZIP code). Users occasionally attempt to correct an error by sending in additional reports; since the last entry is normally the one with which they are satisfied, we automatically filter out duplicate entries and keep the latest. Another common error is a mistyped ZIP code, which could place a user hundreds of kilometers away and is easily caught by the intensity-distance filter.

We save the time that users submit their forms and the time they indicated they believe the earthquake was felt for later sorting, filtering, and analyses. Often the users' reported times aid in determining if the response was for a mainshock or an aftershock and help to resolve other ambiguities.

Conveniently, most errors and intentional mischief are obvious and can thus be filtered or sorted out. On the other hand, more subtle attempts at pranks have little impact on our results due to the overwhelming majority of quality reports. It is notable that mischievous responses tend to occur in times without earthquakes; in the post-earthquake time period the vast majority of users are responsible and the numbers of valid responses overwhelm any potential pranksters. The result is high-quality data when it is most needed. Finally, the DYFI operators maintain ultimate data quality control over the data and results of the system. Analysts can flag suspected entries based on obvious errors, specific comments, or other indicators. Finally, USGS also reserves the option to assign intensities at specific locations as described below.

Limitations at high macroseismic intensities

Higher macroseismic intensity values (typically, VIII and greater) primarily describe observed structural damage to buildings [e.g., Musson 2010]. Given the lack of engineering expertise among the usual DYFI contributor, one could argue that assigning intensities to intensity VIII and higher should formally be done by professionals. That said, DYFI has been calibrated to achieve a maximum of MMI IX by relating community responses to the traditional MMI questionnaire. In comparison with MMI data for high-intensity earthquakes in California and elsewhere, the consistency of community-based intensity VIII and IX assignments is quite convincing [e.g., Dewey et al. 2000, 2002]. For this reason, we assume equivalence of DYFI and MMI assignments unless there is

specific evidence to the contrary.

Nonetheless, for destructive earthquakes, it is necessary to have the capability to accommodate reliable alternative macroseismic observations in addition to DYFI as well as to minimize the influence of errors in web questionnaire responses. Dewey et al. [2002] introduced the notion of a Reviewed Community Internet Intensity (RCII) for just this purpose. RCII entails three elements that go beyond completely automated online data collection. Initially, by default, RCII is the CII computed at a two-week cutoff time, so that the CII value is not allowed to continue changing with additional later responses. Second, RCII data are more thoroughly reviewed for errors or inconsistencies. Third, RCII may be adjusted from CII or assigned to communities lacking CIIs on the basis of other types of macroseismic observations such as engineering reports, press reports, and field reconnaissance (see Dewey et al. [2002] for more details). Hence, professional review by seismologists and field-based macroseismic observations will continue to be important for augmenting higher intensity DYFI observations in the future to enable seismologists to calibrate and fully document what the DYFI data represent. In such cases we can assign a ZIP code intensity based on the independent observations, overwriting the community decimal intensity. To date, this has only been done in rare cases. As previously noted, the authors feel that there are diminishing returns to highly detailed, field-based macroseismic surveys which assign integer macroseismic values to locales, since GIS-based geospatial archiving of damage distributions has become the norm and the expectation following significant earthquake disasters.

Operational robustness, web traffic, and system specifications

A natural concern with a system like DYFI is the challenge of accommodating the post-earthquake deluge of input data and web traffic. As DYFI has grown in popularity, we have continuously improved capacity by making both hardware and software improvements in order to handle the spike in Internet traffic following a widely felt earthquake (an issue known as the "Slashdot effect", see a decade-old description at <http://pasadena.wr.usgs.gov/office/stans/slashdot.html>). To date, the largest such spike is 78 questionnaires submitted in one second recorded for a M 5.7 event in southern California, and there have been four other events that exceeded 50 questionnaires per second; sustained rates at greater than 1,000 forms per minute are likely in the future.

In order to handle this load, we currently have two separate, redundant back-end servers for triggering, event processing, and map creation. Three additional public servers are dedicated to serving the questionnaire itself and handling user input. We cache all DYFI web content, relegating webpage generation to another layer of servers maintained by the Earthquake Hazards Program Web Team. Currently,

there are two intermediate product servers and four web servers that store content from the entire Earthquake Hazards Program. For robustness, these servers are stored in different USGS locations: Golden, CO; Denver, CO; Pasadena, CA; Menlo Park, CA; and Reston, VA. Finally, we commercially contract to a web delivery service provider (Level3 Communications), which redistributes our cached webpage content to thousands of servers distributed globally.

In spite of these precautions, we have no grand expectations for the performance of the DYFI system for the areas hard-hit by damaging ground motions. It is possible, if not likely, that power outages, damage to users’ networks and computers, and limited Internet access will lead to significant data gaps. Thus, the low likelihood of DYFI reliability for retrieving data from the most heavily damaged regions in the immediate aftermaths of destructive earthquakes necessitates a separate, robust post-earthquake response tool like ShakeMap [Wald et al. 1999c]. Unlike DYFI, ShakeMap does not depend on Internet-based human input to place ground motion and intensity maps online immediately.

On the other hand, the DYFI approach does not require expensive, high quality real-time seismic stations – it can be implemented anywhere in the world where there are people with Internet connections, and DYFI data provide direct intensity observations as opposed to intensity values inferred from peak ground motions as in the ShakeMap system [Wald et al. 1999c]. Moreover, we hope the DYFI system will perform well for much of the affected region even for damaging events, and that data will later be available from areas that were not able to respond in the immediate aftermath. In conjunction with ShakeMap and abundant seismic recordings, DYFI provides a constantly improving database for calibrating relations from recorded ground motions to intensities. For these and other noted reasons, the two very different DYFI and ShakeMap approaches to rapidly mapping ground shaking and intensities naturally complement each other: a “man *versus* machine” challenge.

New and ongoing developments with Internet intensity data

After several years of DYFI data collection and processing, and from beneficial advice from our contributors, data users, and other scientists, we have made several modifications to the DYFI system and developed numerous post-processing tools and products for use with the DYFI data. All DYFI data collected to date are now in an archival database; this greatly facilitates research by streamlining the selection, organization, and exportation of data. The DYFI system includes a graphical user interface that allows seismic analysts to perform common functions, including map triggering and resizing, as well as sorting, searching, geocoding, and flagging entries. New web-based geolocation

and geocoding services are being incorporated into DYFI for improving the accuracy of the users’ locations. All these improvements are leading to additional ways to utilize the vast quantities of DYFI data.

DYFI database and graphical user interface

In a recent advancement over command-line interaction, we now allow DYFI to be operated by the NEIC staff of seismic analysts. We have developed a web-based graphical user interface (GUI) that allows our seismic analysts to trigger, delete, resize, or re-center maps, as well as view, supplement, or flag intensity observations. Easy interactive searching and manipulation of the DYFI database allow for additional manual quality control, including flagging obvious outliers or suspected entries and then regenerating maps and webpages for that event. The MySQL database for DYFI contains approximately 77,000 events, with a total of nearly two million individual entries; it supports common queries allowing database reports for annual summaries and use statistics. All completed entries and summary data (ZIP code averaged intensities) are permanently archived.

In order to facilitate the further use of the DYFI data for research, analysis, and visualization, we also provide summaries of DYFI data in a variety of formats online. One simple form of the data available is a tab-delimited summary of the ZIP (geocode) intensity, ZIP code centroid latitude and longitude, epicentral distance, and the number of responses contributed to that ZIP code. Due to privacy considerations, we cannot redistribute personal information or users’ comments. However, upon request, we can provide a sanitized version for research purposes, stripped of identifying data, yet allowing more detailed analyses of the nature of individual responses. We have also begun to produce Google KML files of city and ZIP code intensities for visualization in Google Earth and Google Maps.

Since systems comparable or complementary to DYFI now operate in several countries, collaborative efforts to uniformly collect and exchange data in near-real time are underway [Musson and the ESC Working Group 2009, Musson 2010]. To contribute to this effort, USGS now generates and exports XML-formatted versions of each report, ultimately for near-real time exchange of data with EMSC and other global partners. Systems comparable or complementary to DYFI now operate in several countries, and collaborative efforts to uniformly collect and exchange data in near-real time, particularly in a standardized XML schema, are underway [Musson and ESC Working Group 2009, Musson 2010].

Currently, DYFI is served and collected in both English and Spanish, the choice of the majority of our domestic constituents. Web-based translators provide some additional access to other users, and these capabilities will undoubtedly

be enhanced in the future. However, since the majority of USGS users are English speakers, and due to limited development resources, we have not expanded language options at this time. The EMSC system does provide more extensive multilingual options.

Geolocation

For data collected within the USA, we can manually trigger a DYFI-specific geocoding algorithm, which turns street addresses into latitude and longitude coordinates with precision enough to distinguish adjacent street blocks. Even though we require that users supply only their postal ZIP codes in order for their observations to be processed, up to 90% of respondents to the USGS DYFI website questionnaire volunteer their street addresses. Since intensity assignments optimally make use of numerous entries for robustness [e.g., Worden et al. 2011], we select spatial domains over which to average responses rather than using the ZIP code polygons. To date this has been achieved by subdividing the map extent into a grid of 25 × 25 or 50 × 50 equally spaced boxes and averaging entries in each box. Averaging is done as described earlier and in Wald et al. [1999a], where community averages for each question are determined, rather than simply averaging intensities. In the future we are considering using coordinates of the National Grid domestically and the Universal Transverse Mercator coordinate system internationally rather than map-centric boxes, in order to subdivide our geocoded maps on standard, repeatable domains.

From experience, we have found the additional geocoding processing to be unnecessary in populated areas where ZIP

codes sizes are small, but the spatial refinement provided by geocoding is important where ZIP code areas are larger. For this reason, geocoding is manually triggered as deemed necessary. Since not all respondents provide an address, some information is lost, and there is an inherent trade-off between spatial resolution and the number of responses available for geocoded maps. Examples of the geocoded maps and data can be found online for many of the larger magnitude US events for which thousands of responses have been received. We are currently augmenting the global city-based system with user-supplied, geocoded locations, allowing us to apply our geocoding tools for global events. DYFI promotes web form access via mobile web devices, and we are facilitating mobile use via the development of smart-phone apps; however, automatic locations provided by such devices can complicate processing since the user’s instantaneous position is not necessarily the site where the earthquake was experienced.

DYFI tools and products

In addition to community intensity maps, we produce several complementary products for each DYFI event. We publish a list of the aggregated intensities by ZIP code or city (see Figure 10), sortable by location or intensity, and downloadable in CSV or XML format. In addition, it is convenient to portray these intensity data as a function of distance, thereby showing at a glance the attenuation of intensity from the hypocenter. We now systematically generate and update plots of intensity *versus* distance as the DYFI data are processed (e.g., Figure 5). Following Bakun and Wentworth [1997], the individual intensities are plotted

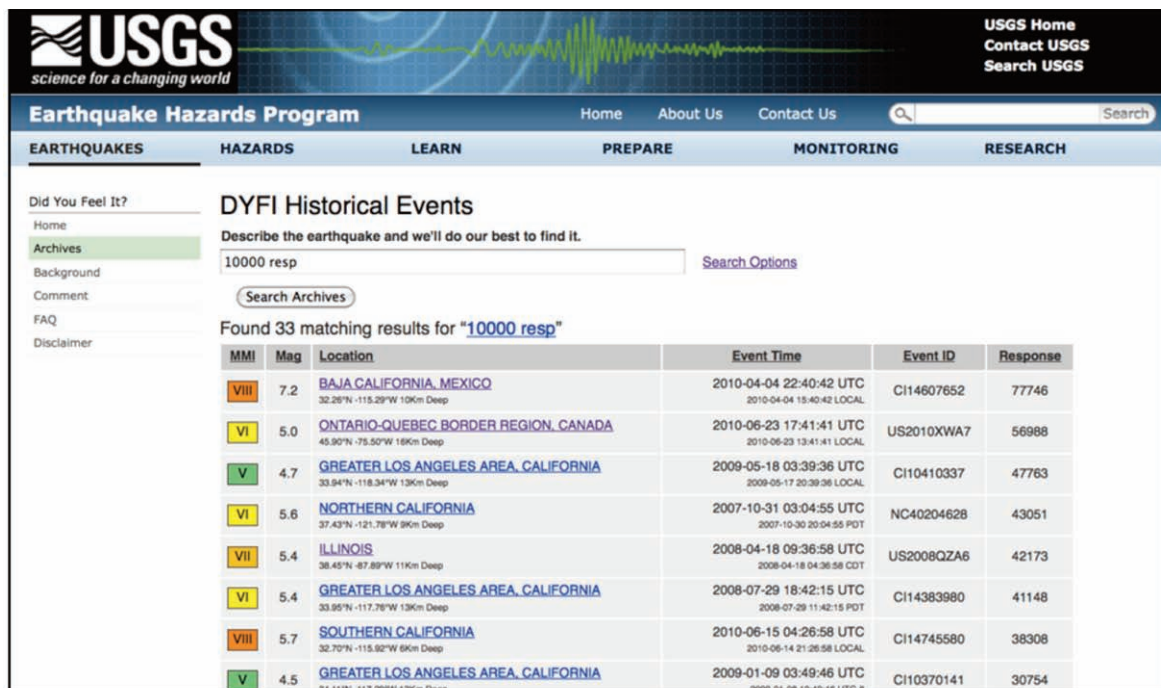


Figure 10. Search of the DYFI Archive Webpage allows queries based on event names, number of responses, magnitude, and intensity ranges. Sortable columns allow listing by any field.

as well as the average intensities calculated for bins in increments of distance. The curves recovered from the DYFI data generally follow regional intensity attenuation functions well, and in fact, an obvious systematic offset of the observed data from previous attenuation curves has provided an independent incentive to re-examine the seismic network magnitudes and/or location assignments for individual events. The intensity–distance plots show clear differences between East and West within the USA: the central and eastern USA typically show higher epicentral intensities for a given magnitude, slower rates of attenuation with distance, and clear indications of the amplifying effects of post-critical reflections from the Mohorovičić discontinuity (e.g., Figures 5, 6, 7; also, see Atkinson and Wald [2007]).

Finally, as discussed above, we publish plots of the incoming DYFI responses as a function of time after the earthquake origin (see Figure 4). This feature provides the system operators the ability to see various effects on the rate of user responses, such as time of day or news coverage. Also, network slowdowns or other problems can be easily inferred from interruptions in the time history.

Research utilizing DYFI data

An important goal in seismology is rapidly estimating shaking and damage after an earthquake. Such impact assessments require calibration against past damage and associated ground motions. Oftentimes only macroseismic data are available to characterize the shaking, and it is of great benefit to be able to infer ground motion peak amplitudes from such data. Given the DYFI data for an event, we have developed tools to estimate ground motions in the absence of data from seismic instruments.

With such available conversion equations, the global DYFI data are also now automatically used as constraints in our global predictive ShakeMap system, which in turn is used as the hazard input for the USGS prototype Prompt Assessment of Global Earthquakes for Response (PAGER) system [Wald et al. 2006a]. While numerous relations now relate peak ground motions to intensities [e.g., Wald et al. 1999b, Atkinson and Kaka 2007], these equations suffer two limitations. First, they can only be used to estimate intensity from recorded peak ground motions, or *vice versa*, but not both; second, they were typically regressed with integer intensity values.

Worden et al. [2011] fully utilize DYFI data to derive new relations among peak ground motion parameters and intensity data. The development of these relations sets a new standard for ground-motion-to-intensity relations in that the DYFI intensity data used are decimal intensities and reverse relations are provided; prior regressions were limited to the use of ordinal intensity values, which complicate the formulation and limit the potential resolution and were regressed only in the direction of estimating intensity from

ground motions. Similarly, DYFI data have recently been used to derive direct intensity prediction equations or IPEs [e.g., Atkinson and Wald 2007, Allen et al. 2011]; these equations are important for robust ShakeMap generation [Worden et al. 2010] and hazard evaluations [e.g., Cua et al. 2010], as well as for improved DYFI filtering.

We are also testing the use of unassociated responses: those to compute earthquake magnitude and location in the case when the instrumentally derived centroid is delayed or unavailable. Bakun and Wentworth [1997] developed a method for deriving these parameters from the historical intensity data. They perform a grid search centered on the area with the highest intensity responses, treat each grid node as a “trial epicenter,” and determine the magnitude and intensity centroid that best fits the DYFI observation points according to a region-dependent intensity-distance attenuation relation. More sophisticated systems have been developed since (e.g., Boxer) [Gasperini et al. 2010]. We will test these algorithms with the new relations derived specifically for DYFI-based IPEs to continuously solve for earthquake magnitude and intensity centroid as unassociated DYFI data are received. The intensity centroid and ground motions determined from the DYFI data correlate well with instrumentally derived parameters. With further efforts at calibrating regional variations of intensity attenuation, this approach could be used to automatically determine location and magnitude globally, independently from seismic network operations, with the added capability of doing so for events below many regional seismic networks’ reporting thresholds.

Education and outreach

DYFI has been recognized by a number of educators as a natural format and opportunity for earthquake-hazard education. For example, the USGS produced an educational exercise using the Northridge earthquake DYFI map as a children’s coloring map to help explain the difference between magnitude and intensity. This exercise was later adopted by the National Geographic Society for their educational webpages and is used routinely to help explain a related question recently added to the State of California Education Standards which requires an understanding of the difference between magnitude and intensity (L. Wald, USGS, personal communication, 2005). Independently, Haase and Park [2006] held a series of grade-school-teacher workshops following a widely felt Indiana earthquake, training teachers to use DYFI for explaining magnitude and intensity, as well as getting students to do online submissions (thus further improving the map for that event). The authors are aware of several elementary schools that routinely use DYFI for a ‘teachable moment’, right after students experience an earthquake firsthand at home or at school.

DYFI maps greatly facilitate communication of earthquake hazards by allowing concrete examples of seismic

intensity. A long-standing limitation in communicating hazards to the public has often been the difficult challenge of explaining hazard in terms of earthquake magnitude. The resurgence of the use of macroseismic intensity for describing hazard is a welcome reversal. ShakeMap, and more so DYFI, intensity maps have contributed to these discussions.

Wald and Dewey [2005] published a public-service-based USGS fact sheet describing DYFI, which was originally printed with 10,000 copies. It was later printed at 50,000 additional copies, primarily to promote citizen-based science during Earth Science Week in 2005 in the USA. One of the fact sheet figures, an earlier version of distribution of felt intensities over the nation for the past decade shown in Figure 2, has been found to be a much more intuitive and effective way of communicating the extent of the earthquake problem throughout the country. It is much easier to remind the public about the national earthquake hazard by showing them what actually happened than by telling them what shaking level has a 10% probability of being exceeded in 50 years (e.g., the USGS National Seismic Hazard Map) [Petersen et al. 2008]. Similar DYFI maps have been used for two Congressional briefings, and the USGS and many related organizations recognize the advantages of communicating with maps of actual felt reports rather than with probabilistic seismic hazard maps.

DISCUSSION AND CONCLUSIONS

The U.S. Geological Survey's "Did You Feel It?" (DYFI) system, relying on Internet data collection after earthquakes, has significant advantages over earlier macroseismic intensity data collection approaches, yet there are some notable limitations arising from its web dependence. Awareness of these limitations reduces potential detrimental impacts, and we are continuing to improve the system as new tools and approaches become apparent. DYFI has always been an evolving system.

Among the recent developments, we have described a number of post-processing tools, applications, and studies that make use of the extensive intensity data sets now gathered, including automatic location and magnitude determination, estimating ground motions from the intensity observations, automatic geocoding to allow for more refined intensity localization, recovering higher-precision decimal intensities rather than limiting intensities to integer values, and social-science analyses of earthquake response and risk perception. We have also expanded DYFI data collection from USA ZIP codes to the rest of the globe, and all indications show the usefulness of the global data. We nonetheless see potential for improving the current international DYFI through new automatic geocoding tools.

The DYFI procedure also has limitations, as discussed. It is strongly conditioned by US traditions of macroseismic data interpretation; we envision collaboration with non-US

macroseismologists to make the product more useful in non-US contexts. Questionnaires in the native language of the source region would clearly facilitate collection of data globally. Ultimately, we expect the global intensity database will prove useful for regional attenuation as well as other seismological studies. Data flow rates are extraordinary, and precautions have been put in place through the use of redundant, hardened, high-capacity processing systems. Nonetheless, data flow after major damaging earthquakes may be limited by power outages, excessive Internet traffic, infrastructure damage, and the more important priorities of users.

From our experience with DYFI, essential components of an Internet-based citizen-science portal include: i) easy-to-use forms, ii) instant feedback so that users may see their contributions (validating their experiences), iii) open space for first-person accounts and discussion of effects not covered in the questionnaire, and iv) routinely addressing user comments and questions. In addition, online user-friendly tools now include common searches, statistics, sorting of responses, time-entry histories, comparisons of data with empirical intensity estimates, and data that are easily downloadable for researchers (Figure 10).

A major advantage of the DYFI system is that its contributors do not need to be trained for the task, and as such they are not 'citizen scientists'; rather, DYFI is 'citizen-based' science. Other citizen-science and crowd-sourced systems do, in fact, require substantial expertise on the part of the contributor. Many new web-based data aggregators face rather daunting challenges in that data quality is more directly tied to or limited by the level of expertise of the available 'crowd'.

The macrointensity maps shown in Figures 1-7 fundamentally depart from those produced with more conventional means in the past. We have described the limitations of Internet-based data collection in the macroseismic realm. Yet, despite the limitations of data collection via the Internet outline, the advantages are both numerous and remarkable:

1. Unprecedented macroseismic data collection rates and consistency provide USGS with the capacity to rapidly map out intensity distribution for all felt earthquakes in the United States and its territories. The cost per observation is orders of magnitude lower than traditional macroseismic data collection efforts.
2. Global data collection provides USGS National Earthquake Information Center with rapid first indications of earthquake occurrence and the potential degree and extent of impact, providing constraints for the Global ShakeMap and thus the PAGER system.
3. Macroseismic data quality is sufficient for: i) portraying and making quantitative use decimal intensities for response as well as scientific purposes, ii) revising otherwise poorly constrained earthquake locations and

magnitude, and iii) distinguishing seismic events from other phenomena (such as sonic booms).

4. Millions of macroseismic observations are available for social-science and seismological analyses.

5. The citizen-based science of the DYFI portal provides an unmatched opportunity for interaction between the scientists of a government agency and the community that they serve. DYFI provides a two-way flow of post-earthquake information providing the USGS with quality macroseismic data, as well as an avenue of information for concerned citizens, and a form of reassurance for those who experienced frightening ground shaking. DYFI maps also greatly facilitate USGS communication about earthquake hazards.

Data and sharing resources

The DYFI data we used in this study are freely available for downloading from the archive at <http://earthquake.usgs.gov/dyfi/>. Geocoded locations can be acquired for many well-reported events online or by contacting the authors, and aggregated datasets can be acquired from the DYFI operators.

Acknowledgements. We are thankful for helpful reviews from Daniel Amorese and Roger Musson. Greg Smoczyk produced the US and global maps of intensity and hazards; other maps were made using GMT [Wessel and Smith 1995]. DYFI fundamentally relies on the capabilities put into place on and beneath the hood of the USGS Earthquake Program Webpages; for this we thank the Earthquake Program Web Team.

REFERENCES

- Allen, T.I., D.J. Wald and C.B. Worden (2011). Prediction of macroseismic intensities for global active crustal earthquakes, submitted to *Journal of Seismology*, in press.
- Atkinson, G.M. and S.I. Kaka (2007). Relationships between Felt Intensity and Instrumental Ground Motion in the Central United States and California, *B. Seismol. Soc. Am.*, 97, 497-510.
- Atkinson, G.M. and D.J. Wald (2007). “Did You Feel It?” Intensity Data: A surprisingly good measure of earthquake ground motion, *Seismol. Res. Lett.*, 78, 362-368.
- Bakun, W.H. and C.M. Wentworth (1997). Estimating earthquake location and magnitude from seismic intensity data, *B. Seismol. Soc. Am.*, 87, 1502-1521.
- Celsi, R., M. Wolfenbarger and D.J. Wald (2005). The Effects of Earthquake Measurement Concepts and Magnitude Anchoring on Individuals’ Perceptions of Earthquake Risk, *Earthq. Spectra*, 21 (4), 987-1008.
- Coburn, A.W. and R.J.S. Spence (2002). *Earthquake Protection*, 2nd edition, Chichester, John Wiley & Sons Ltd.
- Coppola, J.M., X. Lorena, X. Cowan, G.L. Downes, K.F. Feenaghty, P. Grimwood, P. Leach and E. de Robertson (2010). Felt Earthquake Reporting via the Internet in New Zealand, *Seismol. Res. Lett.*, 81 (6), 984-991.
- Cua, G., D.J. Wald, T.I. Allen, D. Garcia, C.B. Worden, M. Gerstenberger, K. Lin and K. Marano (2010). “Best Practices” for Using Macroseismic Intensity and Ground Motion to Intensity Conversion Equations for Hazard and Loss Models, GEM Technical Report 2010-4, Report Series, 69 pp.; <http://www.globalquakemodel.org/node/747>.
- Dengler, L.A. and K. Moley (1994). Toward a quantitative, rapid response estimation of intensities, *Seismol. Res. Lett.*, 65, p. 48.
- Dengler, L.A. and J.W. Dewey (1998). An intensity survey of households affected by the Northridge, California, earthquake of 17 January 1994, *B. Seismol. Soc. Am.*, 88, 441-462.
- Dewey, J., D. Wald and L. Dengler (2000). Relating conventional USGS Modified Mercalli Intensities to intensities assigned with data collected via the Internet, *Seismol. Res. Lett.*, 71, p. 264.
- Dewey, J.W., M.G. Hopper, D.J. Wald, V. Quitariano and E.R. Adams (2002) Intensity distribution and isoseismal maps for the Nisqually, WA, earthquake of 28 February 2001, USGS Open-File Report 02-0346, 57 pp.
- Dowrick, D.J. (1996). The modified Mercalli earthquake intensity scale; revisions arising from recent studies of New Zealand earthquakes, *Bulletin of the New Zealand National Society for Earthquake Engineering*, 29, 92-106.
- Gasperini, G., D. Vannucci, D. Tripone and E. Boschi (2010). The location and sizing of historical earthquakes using the attenuation of macroseismic intensity with distance, *B. Seismol. Soc. Am.*, 100 (5A), 2035-2066.
- Geonames (2011): www.geonames.org (accessed July 6, 2011).
- Giardini, D., G. Grünthal, K.M. Shedlock and P. Zhang (1999). The GSHAP Global Seismic Hazard Map, *Annali di Geofisica*, 42 (6), 1225-1230.
- Goltz, J. (2003). Science can save us: Outreach as necessity and strategy, *Seismol. Res. Lett.*, 74, 491-493.
- 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.
- Haase, J. and C.H. Park (2006). *Mapping Amplification of Earthquake Ground Motion in Indiana*, Series “Lessons Learned Over Time”, Vol. VI-A, Earthquake Engineering Research Institute (EERI), Oakland, CA, USA, 36 pp.
- Mazet-Roux, G. and R. Bossu (2010). *EMSC Real Time Earthquake Information Services*, European-Mediterranean Seismological Centre (EMSC) Report, 31 pp.; available at: www.emsc-csem.org/Files/news/EMSC/EMSC_RTEI_services.pdf.
- Musson, R.M.W. and I. Cecić (2002). Macroseismology, In: W.H.K. Lee, H. Kanamori, P.C. Jennings and C. Kisslinger (eds.), *International Handbook of Earthquake and Engineering Seismology*, Academic Press, Amsterdam, 807-822.
- Musson, R. (2010). *Draft Standard Internet Macroseismic*

- Questionnaire, In: Proceedings of the European Seismological Commission (ESC2010), p. 65.
- Musson, R.M.W., G. Grünthal and M. Stucchi (2010). The comparison of macroseismic intensity scales, *J. Seismol.*, 14, 413-428; doi: 10.1007/s10950-009-9172-0.
- Musson, R. and the ESC Working Group "Internet Macroseismology" (2009). Towards a standard Internet macroseismic questionnaire, Report to ESC, 35 pp.; http://seismologist.co.uk/resources/WG_Discussion_v4.doc.
- Petersen, M., A. Frankel, S. Harmsen, C. Mueller, K. Haller, R. Wheeler, R. Wesson, Y. Zeng, O. Boyd, D. Perkins, N. Luco, E. Field, C. Wills and K. Rukstales (2008). Documentation for the 2008 Update of the United States National Seismic Hazard Maps, U.S. Geological Survey Open-File Report 2008-1128, 61 pp.
- Richter, C.F. (1958). *Elementary Seismology*, Freeman, San Francisco, 768 pp.
- Sbarra, P., P. Tosi and V. De Rubeis (2010). Web-based macroseismic survey in Italy: method validation and results, *Nat. Hazards*, 54, 563-581.
- Sieberg, A. (1930). Scala MCS (Mercalli-Cancani-Sieberg), *Geologie der Erdbeben, Handbuch der Geophysik*, 2, 552-555.
- Stover, C.W. and J.L. Coffman (1993). *Seismicity of the United States, 1568-1989 (Revised)*, U.S. Geological Survey Professional Paper 1527, 418 pp.
- Trumbull, D., R. Bonney, D. Bascom and A. Cabral (2000). Thinking scientifically during participation in a citizen-science project, *Science Education*, 84 (2), 265-275.
- Wald, D.J., V. Quitoriano, L. Dengler and J.W. Dewey (1999a). Utilization of the Internet for Rapid Community Intensity Maps, *Seismol. Res. Letters*, 70 (6), 680-697.
- Wald, D.J., V. Quitoriano, T.H. Heaton and H. Kanamori (1999b). Relationships between peak ground acceleration, peak ground velocity, and modified Mercalli intensity in California, *Earthq. Spectra*, 15, 557-564.
- Wald, D.J., V. Quitoriano, T.H. Heaton, H. Kanamori, C.W. Scrivner and C.B. Worden (1999c). TriNet "ShakeMaps": Rapid generation of peak ground motion and intensity maps for earthquakes in southern California, *Earthq. Spectra*, 15, 537-555.
- Wald, D.J. and J.W. Dewey (2005). Did you feel it? Citizens Contribute to Earthquake Science, U.S. Geological Survey Fact Sheet, 4 pp.
- Wald, D.J., P.S. Earle, K. Lin, V. Quitoriano and B. Worden (2006a). Challenges in Rapid Ground Motion Estimation for the Prompt Assessment of Global Urban Earthquakes, *B. Earthq. Res. I. Tokyo*, 81 (3/4), 273-281.
- Wald, D.J., V. Quitoriano and J.W. Dewey (2006b). USGS "Did You Feel It?" Community Internet Intensity Maps: Macroseismic Data Collection via the Internet, In: *First European Conference on Earthquake Engineering and Seismology (a joint event of the 13th ECEE & 30th General Assembly of the ESC; Geneva, Switzerland, September 3-8, 2006)*, 10 pp.
- Wald, D.J. and D. Bausch (2009). New Research and Tools Lead to Improved Earthquake Alerting Systems, *CUSEC Journal (Central United States Earthquake Consortium)*, 13 (5), 6 pp.
- Wessel, P. and W.H.F. Smith (1995). New Version of the Generic Mapping Tools Released, *Eos Trans. AGU*, 76 (33), p. 329.
- Worden, C.B., D.J. Wald, K. Lin, G. Cua and D. Garcia (2010). A Revised Ground-Motion and Intensity Interpolation Scheme for ShakeMap, *B. Seismol. Soc. Am.*, 100, 3083-3096.
- Worden, C.B., M.C. Gerstenberger, D.A. Rhoades and D.J. Wald (2011). Probabilistic relationships between peak ground motion and Modified Mercalli intensity, *B. Seismol. Soc. Am.*, 101.

*Corresponding author: David J. Wald,
U.S. Geological Survey, National Earthquake Information Center,
Golden, Colorado, USA; email: wald@usgs.gov.