5-2017

The Predictive Relationship between Earthquake Intensity and Tweets Rate for Real-Time Ground Motion Estimation

Yelena Kropivnitskaya
Kristy F. Tiampo
Jinhui Qin
Michael A. Bauer

Follow this and additional works at: https://scholar.colorado.edu/cires_facpapers
The Predictive Relationship between Earthquake Intensity and Tweets Rate for Real-Time Ground Motion Estimation

Yelena Kropivnitskaya¹, Kristy F. Tiampo¹,², Jinhui Qin³ and Michael A. Bauer⁴

¹Department of Earth Sciences, Western University, London, Canada
²CIRES and Department of Geological Sciences, University of Colorado, Boulder, CO, USA
³SHARCNET, Western University, London, Canada
⁴Department of Computer Science, Western University, London, Canada

ykropivn@uwo.ca, kristy.tiampo@colorado.edu, jqin5@uwo.ca, bauer@uwo.ca

Corresponding author: Yelena Kropivnitskaya

Department of Earth Sciences,
Western University,
Biological & Geological Sciences Building, Room 1020
1151 Richmond Street N.
London, Ontario, Canada, N6A 5B7
Phone: +1 226 238 7070
Email: ykropivn@uwo.ca

Abstract
The standard measure for evaluation of the immediate effects of an earthquake on people and man-made structures is intensity. Intensity estimates are widely used for emergency response, loss estimation and distribution of public information after earthquake occurrence. Intensity measures are designed to standardize the measurements of seismic effect and their subsequent evaluation and response (Brazee, 1976; Wood and Neumann, 1931). Modern intensity
assessment procedures process a variety of information sources. Those sources are primarily from two main categories: physical sensors (seismographs and accelerometers) and social sensors (witnesses reports). Acquiring new data sources in the second category can help to speed up the existing procedures for intensity calculations and improve the accuracy of those assessments in a more timely fashion. One potentially important data source in this category is the widespread micro-blogging platform Twitter, ranked ninth worldwide as of January 2016 by number of active users, ~320 million (Twitter, 2016). In our previous studies, empirical relationships between tweet rate and observed Modified Mercalli Intensity (MMI) were developed using data from the M6.0 South Napa, CA earthquake (Napa earthquake) that occurred on August 24, 2014 (Kropivnitskaya et al., 2016). These relationships allow us to stream data from social sensors, supplementing data from other sensors to produce more accurate real-time intensity maps. The streaming application implementation is based on IBM InfoSphere Streams, a cloud platform for real-time analytics on big data (IBM, 2014). These relationships could potentially decrease latency in the intensity calculations for future earthquakes in California and other places around the world. However, there is a strong need for their validation and calibration in regions other than California. In this study, we validate empirical relationships between tweet rate and observed MMI using new datasets from earthquakes that occurred in California, Japan and Chile during the period March-April 2014. The statistical complexity of the validation test and calibration process is complicated by the fact that the Twitter data stream is limited for open public access, reducing the number of available tweets. In addition, in this analysis only spatially limited positive tweets (marked as a tweet about earthquake) are incorporated into the analysis, further limiting the data set and restricting our study to a historical data set. In this work, the predictive relationship for California is recalibrated slightly and a new set of relationships is estimated for the Japan and Chile.
Earthquake intensity is a location-specific characteristic of the extent and amount of seismic damage that depends not only on the magnitude of the earthquake but also on the distance from the earthquake epicenter to the site of interest and the geological features of the surrounding area. The first earthquake intensity classification was designed and used by Italian scientist Schiantarelli in the 1780s. However, the first modern intensity scale was created in 1828 by P.N.G. Egen. The first widely adopted intensity scale, the Rossi–Forel scale, was introduced to the scientific community in the late 19th century. Since then, numerous intensity scales have been developed and are used in different parts of the world (Hough, 2007).

The traditional understanding of earthquake intensity is based on estimates of infrastructure damage at the specific site of interest estimated from the subjective perception of professional observers and/or volunteers who witnessed the earthquake and its consequences. For this purpose, specially designed questionnaires are designed and at particular locations. A successful example of the modern implementation of this approach is the Did You Feel It? program created by United States Geological Survey (USGS) that collects information from people who experience an earthquake and volunteer to share their observations online to create Community Internet Intensity Maps (CIIM) with observations and extent of damage (Wald et al., 1999).

In addition to the intensity evaluation method based on human observations, another approach is to determine intensity from the peak ground motion, either velocity or acceleration, at a station near the site of interest. Empirical relationships then are employed to calculate intensity level (Worden et al., 2012). Results can be obtained much faster than results obtained from the observation-based analysis. A successful realization of this method is the ShakeMap application, also developed by the USGS (Wald et al., 2006).
The intensity evaluation approaches can be divided into two main categories according to the input information source type, above. The first evaluation category employs social sensor data from people who witness the consequences of the earthquake. The second utilizes data from physical sensors, such as seismometers and accelerometers to estimate intensity. Although the first method historically estimated intensity at a much slower rate than that of physical sensors, today electronic questionnaires and observers’ reports can be supplemented with auxiliary online data sources from social networks, where people also share their observations after an earthquake. Here we access data from the online social networking service Twitter. Twitter enables users to send and read short 140-character messages called "tweets". Registered users can read and post tweets while unregistered users can only read those tweets. Users can access Twitter through the website interface, SMS or a mobile device application that is the true source of the real-time nature of tweets (Twitter, 2016). Twitter data is immediately available in a data stream, which can be mined using stream mining techniques (Schonfeld, 2009). In this study we work with data following the stream model. In this model, data arrive at high speed and data mining algorithms must be able to predict intensity level in real-time and under strict constraints of space and time (Bifet et al., 2010).

Previous studies have shown the potential of Twitter data for earthquake detection (Earle et al., 2010, 2011; Sakaki, 2010; Crooks, 2012; Burks et al., 2014) and intensity estimation (Kropivnitskaya et al., 2016). Kropivnitskaya et al. (2016) created four empirical predictive relationships (linear, two-segment linear, three-segment linear, exponential) that link the positive tweet rates in the first ten minutes following the earthquake with the instrumental intensity level in MMI scale units from a regression analysis of data from physical and social sensors during the Napa earthquake. Figure 1 shows the ratio between the combined data (instrumental and Twitter)
and the MMI intensity values recorded from the USGS in the Napa region. The ratio between estimated and actual intensity is relatively low for this particular earthquake. The proposed joint processing technique using social and physical data demonstrates a significant potential for near real-time predictive streaming applications. However, the developed empirical relationships between earthquake intensity and tweet rates still need to be validated for all of California and other seismically active regions of the world and, if necessary, they need to be spatially calibrated.

The statistical complexity of validating and calibrating the model is complicated by the fact that Twitter data stream is limited for open public access. The basic levels allow only up to 1% of the total tweets volume to be streamed (Twitter, 2016). For our purposes only spatially limited positive tweets (tweets about earthquake) are used, so the rate limitations are critical and historical data are used in our application instead of a real data stream. The relationships are validated and calibrated for three regions: California, Chile and Japan. The degree of social media engagement across these countries is relatively high. The proportion of each country’s population that had a Twitter account as of February 2012 is 36% in the USA, 24% in Japan and 33% in Chile (Dawson, 2012). Therefore, potential for the success of Twitter streaming applications in these regions is high. In the next section we discuss the social and physical sensor data used in this study. The following section contains the details of the validation procedure followed by an explanation of and results for the calibration process for California, Japan and Chile regions. Finally, concluding remarks are provided.

2 Data preparation
In order to validate whether the numerical results quantifying hypothesized relationships between logarithmic tweet rates following ten minutes after an earthquake and earthquake
intensity on MMI scale, obtained from Kropivnitskaya et al. (2016), can be used for other earthquakes in California and in other regions around the world, we selected independent events for testing purposes listed in Table 1. The magnitude scale used in Table 1 is moment magnitude. The moment magnitude is a static measure of earthquake size that strongly depends on stress drop. It does not directly quantify the radiated energy and, as a result, there are limits to its ability to reliably estimate earthquake size (Hanks and Johnston, 1992). However, these considerations do not impact the current analysis as earthquake size is given here only for reference purposes and not used as predictive attribute in the model. The events listed in the Table 1 are shown in Figure 2a (California), Figure 2b (Chile) and Figure 2c (Japan). None of these events was included in the original analysis of Kropivnitskaya et al. (2016).

The selection of seismic events used in this validation is dictated by two constraints. The first is to ensure that there is a large amount of freely available Twitter data. Gathering information from social media feeds is, in essence, a web-mining process (Kosala and Blockeel, 2000; Sakaki et al., 2010). It entails three operations: Extracting data from the data providers (in this case Twitter) via application programming interfaces (APIs); parsing, integrating, and storing these data in a resident database; and then analyzing these data to extract information of interest. However, currently available Twitter tools offer limited capabilities for information gathering procedures (Twitter, 2016). As a result, we used an archived historical dataset of Twitter records from March-April 2014 (see Data and Resources Section). The second limitation is related to the lack of geotagged data within available Twitter data sets. We selected seismic events that both showed a spike in the amount of positive tweets within ten minutes following the earthquake and where at least fourteen of those tweets were geotagged (the minimum number of tweets in the models (Kropivnitskaya et al., 2016)). The geotagged tweets contain the current user location indicator at the time of tweeting and can be directly used in the location-specific intensity
estimation algorithms. However, Graham et al. (2014) showed that only 0.7% tweets are geotagged among 19.6 million tweets. In this case, other tweets that do not have a direct specific location reference, but contain a link to specific cities can be used. The percentage of geotagging in this case is between 2% and 5% (Severo et al., 2015). A geotag can also be obtained from a field in the user account description (7.5% of profiles contain latitude and longitude values, 57% include a named location, 20.4% referenced information that can be used to identify a country, while 15.1% provided humorous or non-spatial information (Takhteyev et al., 2012)). In the approach used by Kropivnitskaya et al. (2016), all three types of geotagged technics were used and the same logic has been implemented here. A text-based geolocation algorithm has been improved by employing a location database extension for California, Japan and Chile. In updating the location database, the complete, shortened or abbreviated names are included for any settlements with a population of more than five thousand people in a 200 km radius of the epicenter of the earthquakes listed in Table 1 (see Figure 2).

After building a data set for each event that is limited in time and space, we generated a tweet-frequency time series with one second time bins and normalized to number of tweets per minute. Note that in the current study we did not use strong motion records and did not calculate intensity level using empirical relationships. Instead, for each point from the Twitter data set we assigned an MMI intensity value from the instrumental intensity database of the USGS National Earthquake Information Center (see Data and Resources Section) corresponding to the same or closest latitude-longitude location (for closest location selection the smallest distance between the points is estimated). For California we selected the M6.8 Ferndale, CA earthquake (March 10, 2014) and the M5.1 La Habra, CA earthquake (March 29, 2014) (Figure 2a) (Kropivnitskaya et al., 2016).
3 Relationship Validation

Earlier relationships between positive tweet rates and observed MMI were developed only using data from the Napa earthquake (Kropivnitskaya et al., 2016). In order to employ them to supplement data from physical sensors for more accurate real-time intensity maps production, the relationships have to be validated for other areas of California and other tectonic regions worldwide. The validation process analyzes the goodness of fit of the regression for Napa earthquake, determining whether the regression residuals are random, and checking whether the model's predictive performance deteriorates substantially when applied to data for the Ferndale and La Habra earthquakes, events in California not used in the original model derivation. Both events are not large enough to validate empirical relationships at intensity levels higher than VI. As a result, the three-segment model is excluded from the validation test.

Both earthquakes’ datasets show that the positive tweets rate increases slower than that predicted by the empirical relationships for the Napa earthquake derived in Kropivnitskaya et al. (2016).

The residuals for both events are shown with whisker diagrams for different models in Figure 3, and display the distribution of the residuals based on a five number summary: minimum, first quartile, median, third quartile, and maximum. The quartiles of the present data values are the three points that divide the data set into four equal groups, each group comprising a quarter of the data. The first quartile is defined as the middle number between the smallest number and the median of the data set. The median of the data is a second quartile. The third quartile is the middle value between the median and the highest value of the data set. The Interquartile Range (IQR) is used here to characterize outliers that skew the data.

Despite the fact that the La Habra data (Figure 3b) contains significantly more outliers in the residuals than the Ferndale data (Figure 3a), the mean residuals for every model for the La Habra
event are much closer to zero (the linear model mean residual is 0.37 MMI units; the exponential model mean residual is -0.25 MMI units; the two-segment model mean residual is 0.28 MMI units) than for Ferndale event (the linear model mean residual is 1.28 MMI units; the exponential model mean residual is 0.98 MMI units; the two-segment model mean residual is 1.57 MMI units).

A visual examination of the residuals has an advantage over numerical model validation methods because it illustrates the complex aspects of the relationship between the model and the new data. Figure 4 shows the residuals from the fitted models. The residuals are not randomly distributed around zero, indicating that the linear assumption may be not reasonable (Figure 4b). However, even for the exponential model case (Figure 4a), the residuals are not distributed normally around zero. In this case, the variances of the error terms at each MMI level are not equal due to the different number of twitter responses at each level. Moreover, some of the residuals stand out from the basic pattern, confirming there are outliers in the data as indicated by whisker diagrams (Figure ). These data points have to be excluded from the calibration procedure detailed below.

Intensity attenuates with distance from the epicenter of an earthquake and intensity prediction equations usually rely on some distance metric. For example, some models use epicentral distance (Bakun and Wentworth, 1997), or closest distance to rupture, or some variant that considers extended fault sources, such as the Joyner-Boore distance (Joyner and Boore, 1993). For small magnitude earthquakes, where an earthquake can be approximated by a point source, the difference between point and extended source distance metrics can be minimal. However, at larger magnitudes where source finiteness can be significant, prediction equations which use point source distance metrics may not be applicable (Cua et al., 2010). In this evaluation, we consider the epicentral distance, a point-source distance metric, as a potential predictor for the
equations of Kropivnitskaya et al. (2016). The anticipated influence of the distance metric on the predictive equations is based on the assumption that number of positive tweets and tweet rates are expected to attenuate with distance from the epicenter. As a result, the behaviour of the ground motion components may vary with epicentral distance and result in a variation in the type and/or magnitude of structural damage. To assess this possibility, we plotted residuals versus epicentral distance (Figure 4a, c, e). Note that these plots do not exhibit a systematic structure and, as a result, the form of the function cannot be improved using that predictor.

To check for temporal variation in the data, the residuals also are plotted versus time (Figure b, d, e). At each minute, the intensity map is updated according to new the values received during the last minute. However, at the same time all values obtained earlier remain on the map and are incorporated into the overall intensity representation. Results show that the residuals plotted versus time do not show any drift in the errors and appear to behave randomly.

Both sets of residual plots do record the delay from zero distance and zero time for the Ferndale earthquake, an offshore event that occurred 78 kilometers off the coast. The cumulative error over the ten minute interval is shown on Figure. The maximum error for the Ferndale earthquake registered at the third minute, the minimum error occurred at the tenth minute. For the La Habra event, the maximum error registered at the first and seventh minutes while the minimum error occurred at the fifth and tenth minute. Results suggest that the original model of Kropivnitskaya et al. (2016) does not fit the validation dataset as well as the data from the Napa Valley earthquake it was developed from and a more complete calibration process for the California region is warranted as more data becomes available.
Figure shows the data for Chile and Japan earthquakes from Table 1 plotted with the prediction models of Kropivnitskaya et al. (2016). The data distribution follows the same pattern as the models, but it is clear that the models have to be calibrated and shifted to the left to represent the data relationships better. One of the Chile earthquakes has a significantly higher amount of tweets after an earthquake (Figure a, upward triangles). Noting that all earthquakes generally show common characteristics, one possible reason could be related to the fact that this earthquake happened the day after the M6.7 event that occurred 64 km WNW of Iquique on March 16, 2014. People were likely alert and their reaction was stronger. The data from Japan does not allow for validation of the empirical relationships at an intensity level higher than V. As a result, the three-segment and two-segment models are excluded from the calibration. Observed data for the Chile region represent the intensity levels up to VI, therefore three-segment model is not taken into consideration.

4 Calibration of Existing Relationships
In validating the earlier relationship in the previous step, discrepancies were noted in the spatial variation of intensity, resulting in the need for spatial model calibration. This calibration task involves systematic adjustment of model parameter estimates so that the model outputs more accurately reflect external benchmarks (Thompson, 2012). The forward problem that describes the relation between the logarithmic tweets rate and intensity predictions may be represented with an equation in the following form:

\[ MMI = f(\theta) + e, \]

Where,

\[ MMI \] is the aggregated instrumental intensity level observed,
\( \theta \) is the vector of the tweet rates observed,

\( f \) is the forward equation representing each mathematical model (linear, exponential, 2-segment, 3-segment),

\( e \) is the residuals vector which describes the deviation between the measured and predicted values of intensity (2). Here

\[ e = e_o + e_m, \quad (2) \]

where \( e_o \) accounts for measurement error in the observations, and \( e_m \) is the model error.

The calibration here is an inverse procedure. An inverse estimator \( C \) that connects the observations MMI to “good” estimates \( g \) of the parameters of interest:

\[ g = C[f(\theta) + e] = \min \sum [MMI - f(\theta)]^2 \quad (3) \]

We used known data in the observed relationship between the dependent variable MMI and independent variable logarithmic tweets rate to estimate values for regions other than Napa Valley using new observations from the earthquakes listed in the Table 1. Linear regression is used to produce new regression coefficients for each region. We regress instrumental MMI against the logarithmic mean of the number of tweets per minute to obtain new predictive equations (Figure ) within a legitimate range of values for each model (see Table 2). We also regress the average ground-motion values for specified MMI levels to approximately follow the appropriate trend instead of producing a relationship that is overly influenced by the greater statistical volume of data at lower intensities. We applied a least squares solution with 95% confidence bounds for each model.
For the California region the lowest root-mean-square (RMS) of 0.0029 MMI units is observed for the two-segment model. For the Napa earthquake (Kropivnitskaya et al., 2016) the lowest error was achieved with the three segment model, excluded from validation here. The exponential model demonstrates the RMS error of 0.0035 MMI units. For the Japanese data the difference between the RMS error for the linear and exponential models also is not significant (0.0207 vs. 0.0186). This may be explained by the use of the mid-valued intensity data in the regression (III - V). For lower intensity levels (I-III), the difference will be more significant and for intensity levels higher than V it is less significant (see Figure b). For the Chile earthquakes, the lowest RMS error of 0.0143 MMI units is observed for the linear model, which is 0.0002 units lower than for exponential model.

Because the RMS error cannot determine whether the model estimates and predictions are biased, we also assessed the residual plots (Figure ). The residuals between the predicted and observed data are shown in Figure a for each model in the California region and demonstrate normality in every case. The residual distribution for Japan also is nearly normal for both calibrated models. The distribution of the Chile data residuals is skewed to the left for the two-segment model and skewed to the right for the linear and exponential models. The condition that the error terms are normally distributed is not met in that case.

5 Conclusions
Incorporation of social sensor data with traditional data sources using advanced computational processing methods can provide more complete and accurate coverage of damage and loss for rapid earthquake response. The prediction equations obtained in this work could be used for real-time seismic hazard mapping and emergency management purposes in California, Japan and
Chile using real-time data streaming concept of Twitter data. However, because our previous work showed a higher level of uncertainty resulting from the use of Twitter data alone when compared to results streamed jointly with instrumental intensity, we propose to use the empirical equation for Twitter data together with the data from physical sensors as presented in Kropivnitskaya et al. (2016).

The twitter stream processing here uses data recorded within the ten minutes following an earthquake. We confirm our earlier hypothesis that the logarithmic number of tweets can be used as a proxy for shaking intensity not just in the California region, but also in other regions with large numbers of Twitter users, such as Japan and Chile. However, for actual real-time implementations, there is a need to remember that the calibrated relationships presented in this paper are constrained by the number of tweets in the historical dataset. Unlimited access to the Twitter data stream would allow for new, improved calibration of the regional relationships from around the world. In many areas, the importance of this additional data source could be very significant, due to the complete or partial lack of traditional, instrumental data sources as a result of the high cost of their installation and ongoing operation.
Data and Resources
Instrumental intensity data used in this study is obtained from USGS National Earthquake
Information Center at http://earthquake.usgs.gov/earthquakes/search/ (last accessed on June 20,
2016). Archived historical dataset of Twitter records from March-April 2014 (downloaded from
https://archive.org/details/twitterstream) were used in this study. Figures were created using
GMT (Wessel and Smith, 1991) and Matlab plotting software.

Acknowledgments
The research of KFT and YK was made possible by a MITACS Accelerate grant and an NSERC
Discovery Grant and is the result of collaboration between the Western University
Computational Laboratory for Fault System Modeling, Analysis, and Data Assimilation and
SHARCNET, a high performance computing consortium of Canadian academic institutions..
References


Dawson, R., 2012 Which countries have the most Twitter users per capita? [Online] Available at: http://rossdawsonblog.com/weblog/archives/2012/02/which-countries-have-the-most-twitter-users-per-capita.html/[Accessed 2016].


Joyner, W.B. and D.M. Boore, 1993. Methods for regression analysis of strong-motion data, 


Estimation Using Streaming Data Analysis of Social and Physical Sensors. _Pure and 
Applied Geophysics_, accepted with minor revisions.

Event Detection by Social Sensors. _Raleigh, NC, World Wide Web Conference (WWW)._ 

Schonfeld E., 2009. Mining the thought stream. TechCrunch Weblog Article, 

on four European cities. In C. Levallois (Ed.), _Handbook of Twitter for Research._

Takhteyev Y., Wellman B., Gruzd A., 2012. Geography of Twitter Networks, _Social Networks_, 
Volume 34, Issue 1, January 2012, pages 73–81

Twitter, 2016. The Twitter Platform Documentation. [Online] Available at: 
https://dev.twitter.com/overview/documentation [Accessed 2016].


Wald, D.J., Quitariano, V., Heaton, T.H., Kanamori, H., Scrivner C.W., and Worden, c.B. 

between ground-motion parameters and Modified Mercalli intensity in California. _Bull. 
Figure 1. Ratio between combined intensity level (from physical and social sensors) and instrumental intensity level (triangles – seismic stations) after the Napa earthquake (red star - epicenter): a) linear model, b) exponential model, c) two-segment model and d) three-segment model.
Figure 2. Population density in the regions with earthquake epicenters used in the validation and calibration process and areas covered for analysis (circles):

a - California (1 – Ferndale earthquake, 2 – Napa Valley earthquake, 3 – La Habra earthquake);

b - Chile (1 - 64km WNW of Iquique, 2 - 80km WNW of Iquique, 3 - 91km WNW of Iquique, 4 - 94km NW of Iquique);

c - Japan (1 – Nago earthquake, 2 - Kunisaki-shi).

(Population density data from GPWv3 (CIESIN, 2005))
Figure 3. Whisker diagram for the residuals (red – median, blue square – mean, error bars – standard deviation): a - Ferndale earthquake; b - La Habra earthquake.
Figure 4. Residuals vs. epicentral distance (a,c,e) and time (b,d,f) for the exponential (a, b), linear (c, d) two segment (e,f) models (grey squares – La Habra earthquake, black circles – Ferndale earthquake).
Figure 5. Cumulative average error over ten minutes (grey – Ferndale earthquake, black – Lahabra earthquake, solid line – linear model, dashed line – exponential model, dotted line – two-segment linear model).
Figure 6. Observed earthquake data:

a) Chile (circles - 64km WNW of Iquique, upward triangles - 80km WNW of Iquique, downward triangles - 91km WNW of Iquique, dots - 94km NW of Iquique) with prediction models (solid thin line – linear, dot dashed line - exponential, thick solid line – two-segment linear, dashed line – three segment);

b) Japan (circles – Nago earthquake, dots - Kunisaki-shi earthquake) with prediction models (solid thin line – linear, dot dashed line - exponential, thick solid line – two-segment linear, dashed line – three segment).
Figure 7. Calibrated models (black line – linear, dark grey line - exponential, light grey line – two-segment linear) for California (solid lines), Japan (dotted lines), Chile (dashed lines).
Figure 8. Residuals between calibrated models and observed data:

a). California
b). Chile
c) Japan
Table 1. List of earthquakes used in validation and calibration processes

<table>
<thead>
<tr>
<th>Date and Time</th>
<th>Magnitude</th>
<th>Depth (km)</th>
<th>Epicenter Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-03-02 20:11</td>
<td>6.5</td>
<td>119</td>
<td>111km NNW of Nago, Japan</td>
</tr>
<tr>
<td>2014-03-10 5:18</td>
<td>6.8</td>
<td>16.6</td>
<td>78km WNW of Ferndale, California</td>
</tr>
<tr>
<td>2014-03-13 17:06</td>
<td>6.3</td>
<td>79</td>
<td>15km NNE of Kunisaki-shi, Japan</td>
</tr>
<tr>
<td>2014-03-16 21:16</td>
<td>6.7</td>
<td>20</td>
<td>64km WNW of Iquique, Chile</td>
</tr>
<tr>
<td>2014-03-17 5:11</td>
<td>6.4</td>
<td>21</td>
<td>80km WNW of Iquique, Chile</td>
</tr>
<tr>
<td>2014-03-22 12:59</td>
<td>6.2</td>
<td>20</td>
<td>91km WNW of Iquique, Chile</td>
</tr>
<tr>
<td>2014-03-29 4:09</td>
<td>5.1</td>
<td>5.1</td>
<td>2km E of La Habra, California</td>
</tr>
<tr>
<td>2014-04-01 23:46</td>
<td>8.2</td>
<td>25</td>
<td>94km NW of Iquique, Chile</td>
</tr>
</tbody>
</table>
### Table 2. Calibrated Predictive Relationships

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Equation</th>
<th>RMS error, MMI units</th>
<th>Valid range of values (Ntweets/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>California</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>MMI = 19.11*log(Ntweets/min) – 23.72</td>
<td>0.0033</td>
<td>[17.43; 58.15]</td>
</tr>
<tr>
<td>Exponential</td>
<td>MMI = 0.04636<em>exp(3.086</em> log(Ntweets/min))</td>
<td>0.0035</td>
<td>(0; 55.1]</td>
</tr>
<tr>
<td>Two-segment linear</td>
<td>MMI = 15.41<em>log(Ntweets/min) – 18.63, log(Ntweets/min)&lt;1.52, MMI = 35.47</em>log(Ntweets/min) – 49.19, 1.52&lt;log(Ntweets/min)&lt;1.61</td>
<td>0.0029</td>
<td>[24.4; 46.63]</td>
</tr>
<tr>
<td><strong>Japan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>MMI = 1.56*log(Ntweets/min) + 2.2</td>
<td>0.0207</td>
<td>[0.04; 10^5]</td>
</tr>
<tr>
<td>Exponential</td>
<td>MMI = 2.584<em>exp(0.3653</em> log(Ntweets/min))</td>
<td>0.0186</td>
<td>(0; 5064]</td>
</tr>
<tr>
<td><strong>Chile</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>MMI = 1.58*log(Ntweets/min) + 3.241</td>
<td>0.0143</td>
<td>[0.009; 18 960]</td>
</tr>
<tr>
<td>Exponential</td>
<td>MMI = 3.4<em>exp(0.339</em> log(Ntweets/min))</td>
<td>0.0145</td>
<td>(0; 1517.6]</td>
</tr>
<tr>
<td>Two-segment linear</td>
<td>MMI = 16.06<em>log(Ntweets/min) – 3.5338, log(Ntweets/min)&lt;0.47, MMI = 3.59</em>log(Ntweets/min) + 2.32, 1.52&lt;log(Ntweets/min)&lt;0.86</td>
<td>0.0358</td>
<td>[1.66; 137.8]</td>
</tr>
</tbody>
</table>