Optimization of legacy lidar data sets for measuring near-field earthquake displacements

Craig L. Glennie¹, Alejandro Hinojosa-Corona², Edwin Nissen³, Arpan Kusari¹, Michael E. Oskin⁴, J. Ramon Arrowsmith⁵, and Adrian Borsa⁶

¹Department of Civil and Environmental Engineering, University of Houston, Houston, Texas, USA, ²Department of Geology, CICESE Ensenada, Baja, California, Mexico, ³Department of Geophysics, Colorado School of Mines, Golden, Colorado, USA, ⁴Department of Earth and Planetary Sciences, University of California, Davis, California, USA, ⁵School of Earth and Space Exploration, Arizona State University, Tempe, Arizona, USA, ⁶Scripps Institution of Oceanography, University of California, San Diego, California, USA

Abstract

Airborne lidar (light detection and ranging) topography, acquired before and after an earthquake, can provide an estimate of the coseismic surface displacement field by differencing the preevent and postevent lidar point clouds. However, estimated displacements can be contaminated by the presence of large systematic errors in either of the point clouds. We present three-dimensional displacements obtained by differencing airborne lidar point clouds collected before and after the El Mayor–Cucapah earthquake, a $M_w$ 7.2 earthquake that occurred in 2010. The original surface displacement estimates contained large, periodic artifacts caused by systematic errors in the preevent lidar data. Reprocessing the preevent data, detailed herein, removed a majority of these systematic errors that were largely due to misalignment between the scanning mirror and the outgoing laser beam. The methodology presented can be applied to other legacy airborne laser scanning data sets in order to improve change estimates from temporally spaced lidar acquisitions.

1. Introduction

Airborne laser scanning (ALS) or airborne light detection and ranging (lidar) has become a common instrument for use in the Earth Sciences. There are numerous examples of the use of ALS to help answer fundamental science questions (see Glennie et al. [2013] and Meigs [2013] overviews). The use of lidar has also been prevalent in the documentation of surface ruptures of recent and historic earthquakes, see for example, [Hudnut et al., 2002; Salisbury et al., 2012; Zielke et al., 2010].

While postevent lidar provides discrete measurements of slip along the fault (via geomorphically informed reconstruction of offset features along the rupture), it does not directly provide an absolute model of the spatially varying, three-dimensional (3-D) surface displacement in the near field (hundreds of meters), as this requires a model of the preearthquake topography. As recently demonstrated with experiments on point clouds deformed with simulated earthquake displacements [Borsa and Minster, 2012; Nissen et al., 2012], lidar differencing has a rich potential for complementing far-field surface deformation estimated from interferometric synthetic aperture radar (InSAR), GPS [Segall and Davis, 1997] and pixel-tracking techniques [Leprince et al., 2007; Sandwell et al., 2000; Wei et al., 2011], and on-fault estimates of slip made in the field. The desire to have a precise preevent topography model was a significant motivation behind the comprehensive ALS surveys of the San Andreas fault system undertaken by the NSF-funded B4 [Bevis et al., 2005] and GeoEarthScope [Prentice et al., 2009] projects from 2005 through 2007, although a major earthquake has yet to occur at any location covered by these data sets.

Other unexpected opportunities have arisen, where preevent lidar data were available: for the 2010 El Mayor–Cucapah earthquake in Mexico [Oskin et al., 2012] and the 2010 Darfield New Zealand earthquake [Duffy et al., 2013]. While studies of these events have afforded the first detailed looks at near-field earthquake displacements through the differencing of ALS data sets, they have also highlighted several shortcomings of older “legacy” (or preevent) lidar data sets when compared to more recent data collected with state-of-the-art instrumentation and processed with current lidar system calibration procedures. The erroneous calibration and improper error modeling that is prevalent in these older data sets appear as significant systematic errors in the surface displacement field, manifested as apparent slip, and spatially...
distributed surface warping that can adversely impact models of seismic slip at depth and of the magnitude of strain in the surrounding rock volume [e.g., Oskin et al., 2012]. It is therefore important to minimize or remove these systematic errors before differencing.

To this end, in this paper we describe the reprocessing of the El Mayor–Cucapah legacy lidar data set to remove artifacts that became apparent during differencing with a high-quality postevent data set acquired in 2010. The additional processing steps are potentially relevant to existing high-resolution data sets because in the event of an earthquake along the San Andreas Fault, or other areas with existing lidar surveys (as is becoming more and more common globally), it is likely that the already high-resolution data acquired could be improved using similar methods to more accurately model the displacement field.

2. Preevent and Postevent Data Sets

On 4 April 2010, there was a $M_w$ 7.2 earthquake in northern Baja California, Mexico, herein referred to as the El Mayor–Cucapah (EMC) earthquake. The earthquake principally involved right-lateral strike slip along a 120 km long, NW-SE trending fault zone in the Sierra Cucapah Mountains and the Colorado River delta [Hauksson et al., 2011; Wei et al., 2011]. Within the northwestern half of the rupture zone—in the interior of the Sierra Cucapah—several distinct faults were activated in the earthquake, many of which included a significant normal component of slip [Fletcher et al., 2010] (Figure 1b). The entire area had previously been mapped in a regional, low-resolution, high-altitude lidar survey undertaken in 2006 by the Instituto Nacional de Estadística y Geografía (INEGI). Postearthquake lidar data were acquired 4 months after the earthquake. The two data sets are briefly detailed below.

1. Preevent:

From 31 July to 7 August 2006, a team from INEGI collected lidar observations along the future EMC earthquake surface rupture as part of a larger regional survey of Baja California. The ALS observations were acquired using a Leica ALS40 laser scanner from an aircraft flying at an altitude of 6000 m and a speed of 100 m/s. The ALS sensor was operated at a 16 kHz pulse rate and a scanning angle of ±32°. There are 16 north-south flight lines over the area of interest, with nominally 30% side overlap. The resultant data set has a point density of approximately 0.013 points/m². The areal extent of the preevent data set is given by the blue polygon in Figure 1a.

2. Postevent:

Four months after the EMC earthquake, a team from the National Center for Airborne Laser Mapping (NCALM) resurveyed the topography within several kilometers of the identified surface rupture. ALS observations were collected on 16–19 August 2010 from an aircraft flying at an altitude of 600 m at approximately 80 m/s using an Optech Gemini ALS sensor operating at a 100 kHz pulse rate, a scanning angle of ±14°, and a scanning rate of 60 Hz. Twenty four lines (with 50% side overlap), comprising a corridor 3 km wide and 106 km in length, were mapped, with a nominal point density of ~10 points/m². The flight lines were flown approximately parallel to the NW-SE trace of the main surface rupture, and the approximate coverage of the postevent data set is shown in Figure 1a (purple polygon).

3. Need for Reprocessing the Preevent Data Set

Initial differencing results from the preevent and postevent lidar scans from the EMC earthquake was presented by Oskin et al. [2012], who used vertical differencing of digital elevation models (DEMs). This approach does not account for horizontal displacements, which are likely to dominate the deformation field of what is a largely strike-slip earthquake. The best results in Oskin et al. [2012] were in locations with large vertical displacements and smaller horizontal displacements. Significant horizontal displacements were in fact observed via subpixel correlation of optical and synthetic aperture radar (SAR) imagery [Wei et al., 2011]. Differencing methods such as the ICP (iterative closest point) algorithm, as first reported by Besl and McKay [1992] and recently tested on lidar data deformed with simulated earthquake displacements [Nissen et al., 2012], afford the possibility of determining the full 3-D displacement field from a difference of lidar point clouds only. Here we concentrate on the mountainous northwestern half of the fault zone (Figure 1b), which lends itself better to topographic differencing. In our implementation, the two data sets are each split into 100 m square cells, and ICP is run on each cell pair to determine the 3-D displacement and rotation that best
aligns the preearthquake and postearthquake unfiltered point clouds. For our initial analysis, we examined each component of the ICP-derived displacement field independently. While the vertical and east-west (horizontal $x$ component) displacements appeared reasonable, the north-south (horizontal $y$ component) showed significant systematic trends not attributable to earthquake-related deformation or other natural processes (Figure 2a).

An examination of Figure 2a shows that the vertical and horizontal $x$ components of the estimated displacement field from ICP are smoothly varying, with a discontinuity delineating the EMC surface rupture. The $y$ component however did not delineate the surface rupture and shows large systematic deviations across the study area. For the 2006 lidar acquisition, flight lines were flown with 0° (north) or 180° (south) azimuths, as indicated by the rectangular outlines on the $y$ component displacement field in Figure 2a. The error in the $y$ component is correlated with both the flight line location and flight line direction in the 2006 survey. This strongly suggested that there were processing errors in the 2006 INEGI lidar data set. Fortunately, we were able to obtain a copy of the raw lidar survey data from INEGI. We undertook an analysis and reprocessing of these raw data (described in the following section) in order to identify, model, and remove (or diminish) these systematic errors from the preevent point cloud.
4. Description of Reprocessing Undertaken

4.1. GPS Reprocessing

The original kinematic differential global positioning system (DGPS) processing for the INEGI 2006 survey was done using only one base station (labeled as INEGI 1 in Figure 1), located at the Tijuana airport, which is approximately 115 km west of the northwest boundary and 200 km west of southeast corner of the EMC rupture zone, respectively. Over these long baseline distances, it is difficult or impossible to correctly resolve GPS ambiguities to provide consistent subdecimeter level kinematic positioning [Borsa et al., 2007]. However, the 2006 raw data set also contained observations from an additional GPS base station (labeled as INEGI 2 in Figure 1) located significantly closer to the fault line (25–75 km from NW and SE extents, located just north of the data acquisition area) that was not used in the initial DGPS processing. The data from both of these base stations, along with archived 15 s GPS data from Plate Boundary Observatory stations P066, P494, P496, and P500 [Phillips et al., 2013] and Continuous Operating Reference Station IID2 [Snay and Soler, 2008], were combined using the simultaneous kinematic DGPS baseline processing module of Novatel’s Grafnav commercial software package to produce an improved estimate of the aircraft trajectory. See Figure 1 for the location of the GPS base stations.

In order to compare the initial INEGI trajectories with the improved multiple base station estimates, we calculated the three-dimensional distance between the two solutions for each GPS epoch. The two solutions vary approximately 0.7 m 3-D root-mean-square error (RMSE) difference, which is significantly greater than the centimeter level error typically obtained for kinematic GPS solutions (Table 1). Note that the differences in

---

Figure 2. (a) Estimated displacement field by ICP between pre-event and postevent lidar point cloud in (left) x, (middle) y, and (right) z direction before reprocessing of pre-event ALS data set; arrows show flightline direction for pre-event data collection. (b) Estimated displacement field by ICP between pre-event and postevent lidar point cloud in (left) x, (middle) y, and (right) z direction after reprocessing of pre-event ALS data set. (c) Four representative profiles (location of profiles shown in Figure 2b) of displacement field, both before (red) and after (blue) reprocessing.
still some residual errors in the displacement estimates that are correlated with location of the surface rupture is now delineated in all three components of displacement. However, there are figures 2b and 2c show a noticeable improvement in the 3-D ICP estimates of surface displacement. The displacement are displayed in figure 2b, and a comparison of four representative before and after pro

the mirror misalignment simultaneously with the boresight adjustment, regenerated the preearthquake in Glennie boresight calibration model to include parameters for a second-order polynomial following the development outgoing laser beam [2012] for estimating interior orientation errors in the Velodyne laser scanner. We then estimated in Glennie, 2005]. We modeled the mirror/laser beam misalignment by extending our field are shown in figure 2c.

4.2. Reboresight Calibration

The determination of the coordinates of a lidar return in a georeferenced global coordinate frame is calculated using the lidar georeferencing equation

\[ \mathbf{r}_G = \mathbf{r}_{GPS} + \mathbf{R}_b (\mathbf{R}_s \mathbf{r}_s + \mathbf{l}_b) \]  

(1)

where \( \mathbf{r}_G \) is the lidar point coordinates in the global reference system, \( \mathbf{r}_{GPS} \) and \( \mathbf{R}_b \) are the navigation system-derived position and attitude of the laser scanner at the measurement time, \( \mathbf{r}_s \) is the laser measurement vector, and \( \mathbf{R}_s \) and \( \mathbf{l}_b \) are the boresight and lever arm values, respectively [Glennie, 2007]. These final two terms are system calibration values that arise because the measurement origin of the laser scanner and navigation system cannot be physically colocated and aligned. The lever arm offset can normally be measured externally with sufficient accuracy, but unfortunately, the determination of boresight angles is not direct and requires estimation through an analysis of the resultant airborne lidar point cloud. There are a variety of methods to determine these calibration parameters [e.g. Habib et al., 2011; Skaloud and Lichti, 2006]. The improper determination of boresight angles can be a root cause of systematic errors in a lidar point cloud such as those apparent in Figure 2a. We therefore reestimated the boresight angles using the approach detailed by Glennie [2012], where flightline differences on common surfaces in the overlap between flightlines are minimized by a least squares adjustment of the boresight angles. While the new estimates of the boresight angles partially eliminated the systematic errors in the y direction (reducing the range of deviations from ±12 m to ±5 m), there were still significant residual trends in the data set. To illustrate the remaining systematic trend, we plot the y displacement between the preevent and postevent lidar versus the laser scan angle in Figure 3. To remove any possible influence of the surface rupture on the apparent displacement, the data in Figure 3 considers only lidar differences on the northeast side of the surface rupture.

The y displacement curve in Figure 3, which should be flat across the full range of recorded scan angles, shows a clear correlation with scan angle and a large 10 m difference between observations near nadir and those at the 30° extent of the scan. This error pattern, which resembles a second-order polynomial for shots on either side of nadir, is indicative of a misalignment between the lidar system scanning mirror and the outgoing laser beam [Latypov, 2005]. We modeled the mirror/laser beam misalignment by extending our boresight calibration model to include parameters for a second-order polynomial following the development in Glennie [2012] for estimating interior orientation errors in the Velodyne laser scanner. We then estimated the mirror misalignment simultaneously with the boresight adjustment, regenerated the preearthquake point cloud, and recomputed the ICP differences between the preevent and postevent data sets. The results are displayed in Figure 2b, and a comparison of four representative before and after profiles of the displacement field are shown in Figure 2c.

Figures 2b and 2c show a noticeable improvement in the 3-D ICP estimates of surface displacement. The location of the surface rupture is now delineated in all three components of displacement. However, there are still some residual errors in the displacement estimates that are correlated with flight line location and

<table>
<thead>
<tr>
<th>Table 1. Comparison of 2006 INEGI Aircraft Trajectories Before and After DGPS Reprocessing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition Date</td>
</tr>
<tr>
<td>(Number of Flight Lines)</td>
</tr>
<tr>
<td>1 August (6 lines)</td>
</tr>
<tr>
<td>6 August (5 lines)</td>
</tr>
<tr>
<td>7 August (5 lines)</td>
</tr>
</tbody>
</table>

Table 1 do not correspond to errors in the INEGI trajectory; they are differences only. However, since the reprocessed solution used additional base stations with significantly shorter baseline lengths, it is probable that a majority of the differences present in Table 1 represent errors in the original aircraft trajectories. It should also be noted that the GPS reference datum for the 2006 INEGI flights and the 2010 NCALM flights was also different. The 2006 data were referenced to ITRF92 with an epoch date of 1998.800, whereas the NCALM flights were processed using ITRF2000 with an epoch date of 2010.627. The National Geodetic Survey tool Horizontal Time-Dependent Positioning [Pearson and Snay, 2013] was utilized to update the INEGI survey datum to the NCALM flight datum. This datum change resulted in a coordinate shift of approximately −0.900, 0.429, and 0.004 m (east, north, and up) that we applied to the original INEGI data.
orientation. These errors are mainly at the edge of the flight line swaths (i.e., at larger scan angles), where any remaining laser misalignment or boresight errors would be magnified. These mismatches reach a maximum of approximately 40 cm in some locations for the horizontal displacement components, a factor of about 25X smaller than the original observed errors. It should be noted that to minimize these errors at the swath edges, we cropped the preevent point cloud data set to remove points at higher scan angles in areas of overlap between flightlines.

5. Estimated Accuracy of Preevent Lidar Point Cloud

The results in Figure 2b still exhibit systematic deviations. However, to understand the significance of these apparent residual errors, it is important to recognize the expected accuracy of the individual point clouds used to estimate the displacement field. For airborne lidar surveys, the accuracy of the final point cloud is strongly correlated with the airborne acquisition’s height above ground, as angular errors in both the laser scanner and the INS are inversely proportional to height [Baltsavias, 1999]. For the EMC earthquake lidar differences, the underlying accuracy will be dominated by the estimated precision of the preevent lidar data set because it was flown at a lower point density and at a significantly higher flight elevation (6000 m versus 600 m).

We determine the expected accuracy of the preevent lidar point cloud using standard methods to estimate the total propagated error of an ALS point cloud, [e.g., Glennie, 2007; Schaer et al., 2007]. Using those methods, we estimated the uncertainty for (x, y, and z) components in the preevent lidar point cloud (Figure 4).

The estimated horizontal accuracy of the preevent point cloud is greater than 1 m for the y direction and between 0.4 and 1.0 m for the x direction (Figure 4). The estimated vertical accuracy is approximately 1.0 to 1.2 m. The expected accuracy shows a distinct pattern, with the highest confidence at instrument nadir (0° scan angle), where the outgoing circular laser beam would be expected to have nearly perpendicular incidence on the terrain. As the angle of incidence increases (with increasing scan angle or as a result of topography), the estimated error of the lidar point increases as the intersection of the circular laser beam and the terrain becomes more and more elliptical. This effect can become exaggerated in topography, as the insert images in Figure 4 clearly show. The estimated accuracy values for the preevent point cloud are significantly larger than the remaining systematic errors presented in Figure 3. Therefore, while it may still be possible to eliminate additional residual systematic artifacts, the magnitude of the remaining errors is below the expected overall accuracy of the preevent lidar point cloud.

Fault slip and distributed earthquake-driven deformation in the surrounding volume are now well imaged in the resulting improved ICP displacement fields. The earthquake displacements are less than a few meters—in agreement with both field and satellite measurements [Fletcher et al., 2010; Wei et al., 2011]—thus demonstrating the value of the higher accuracy preearthquake data we report on here. Vertical displacements dominate the important Paso Inferior accommodation zone in the northwestern portion of the rupture, but a significant right-lateral motion is also evident from the combination of horizontal components, in close agreement with detailed field mapping undertaken in this area [Fletcher et al., 2010]. This right-lateral component was not characterized by the vertical differencing of the lidar DEMs [Oskin et al., 2012]. Well localized (over the 100 m differencing pixels) displacement discontinuities in all components are also evident in the Puerta accommodation zone, where only very disconnected surface breaks were mapped in the field.

![Figure 3](image-url)  
**Figure 3.** The ICP residuals in y displacement between preearthquake and postearthquake data sets (red and green, respectively), plotted with respect to laser scan angle and perpendicular to flight direction of preevent data collection. Only data on the northeast side of surface rupture were considered.
Fletcher et al., 2010] (Figures 1b and 2b). This indicates localized, shallow (less than hundreds of meters) fault slip occurred in an area, where primary surface ruptures are either absent or were missed by field reconnaissance. More generally, these displacement fields characterize deformation in the nearest 1 km or so of the surface rupture and are thus related to the slip distribution and mechanical properties of the shallow part of the fault zone. In this respect, the differential lidar complements very well the narrow aperture field measurements as well as the broader coverage from InSAR and GPS.

6. Conclusions
Airborne laser scanning, combined with new methodologies for estimating earthquake surface displacements by differencing of pre-event and post-event lidar point clouds, is becoming a powerful tool for investigating the spatial distribution of near-field earthquake surface deformation. With the increasing number of legacy lidar data sets becoming publicly available, including many large-scale, regional surveys similar to the one collected by INEGI, the likelihood of being able to apply lidar differencing methods on future earthquakes is growing quickly. However, accurate estimation of the displacement field relies on lidar point clouds that are of high relative and absolute accuracy. Since many legacy lidar data sets exhibit residual systematic errors, careful analysis and reprocessing of these data sets are recommended if the original raw observations can be obtained. Reprocessing also allows users to perform rigorous error propagation to estimate the spatial accuracy of the lidar point cloud. Future work will focus on the inclusion of these accuracy estimates in point cloud differencing techniques, with accuracy estimates propagated to the resultant displacement fields. This should lead in turn to more refined slip models.

Figure 4. Total propagated uncertainty map for all three directions, (left) x, (middle) y, and (right) z considering estimated accuracy of raw data. Inset images show a zoomed in area and digital elevation model to highlight the correlation of topography with estimated lidar accuracy.

[Fletcher et al., 2010] (Figures 1b and 2b). This indicates localized, shallow (less than hundreds of meters) fault slip occurred in an area, where primary surface ruptures are either absent or were missed by field reconnaissance. More generally, these displacement fields characterize deformation in the nearest 1 km or so of the surface rupture and are thus related to the slip distribution and mechanical properties of the shallow part of the fault zone. In this respect, the differential lidar complements very well the narrow aperture field measurements as well as the broader coverage from InSAR and GPS.

6. Conclusions
Airborne laser scanning, combined with new methodologies for estimating earthquake surface displacements by differencing of pre-event and post-event lidar point clouds, is becoming a powerful tool for investigating the spatial distribution of near-field earthquake surface deformation. With the increasing number of legacy lidar data sets becoming publicly available, including many large-scale, regional surveys similar to the one collected by INEGI, the likelihood of being able to apply lidar differencing methods on future earthquakes is growing quickly. However, accurate estimation of the displacement field relies on lidar point clouds that are of high relative and absolute accuracy. Since many legacy lidar data sets exhibit residual systematic errors, careful analysis and reprocessing of these data sets are recommended if the original raw observations can be obtained. Reprocessing also allows users to perform rigorous error propagation to estimate the spatial accuracy of the lidar point cloud. Future work will focus on the inclusion of these accuracy estimates in point cloud differencing techniques, with accuracy estimates propagated to the resultant displacement fields. This should lead in turn to more refined slip models.
Acknowledgments
This work was supported by the U.S. National Science Foundation (NSF) through a collaborative research grant (EAR 1148319). The NSF supported Open Topography facility (award 1226353 and 1225810) and National Center for Airborne Laser Mapping (EAR 1339015) are also gratefully acknowledged. The authors would also like to thank INEGI for granting them access to the raw lidar point clouds for the EMC earthquake and collecting the data. The NSF supported the U.S. Geological Survey National Geodetic Survey (Award EAR 1148319). The NSF supported this work. The Editor thanks Benjamin Brooks and an anonymous reviewer for their assistance in evaluating this paper.

References