

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Spatial Analysis for Identifying Concentrations of Urban Damage

Joseph Wartman, Nicholas E. Malasavage  
*Drexel University Engineering Cities Initiative (DECI)*  
*joseph.wartman@drexel.edu, nem23@drexel.edu*  
*United States of America*

## 1. Introduction

Disasters resulting from earthquakes, hurricanes, fires, floods, and terrorist attacks can result in significant and highly concentrated damage to buildings and infrastructure within urban regions. Following such events, it is common to dispatch investigation teams to catalog and inventory damage locations. In recent years, these data gathering efforts have been aided by developments in high resolution satellite remote sensing technologies (e.g. Matsuoka & Yamazaki, 2005) and by advances in ground-based field data collection (e.g. Deaton & Frost, 2002). Damage inventories are typically presented as maps showing the location and damage state of structures in part or all of an effected region. In some cases information on the post-event condition of major infrastructure systems such as transportation, power, communications, and water networks is also included. Depending on the means used to acquire data, damage inventories may be developed in days (satellite-based data acquisition) or weeks-to-months (ground-based damage surveys) after an event. Once available, these inventories can be used for a range of purposes including guiding emergency rescues (short-term use), identification of neighborhoods requiring post-disaster financial assistance (intermediate-term use), and support of zoning, planning or urban policy studies (long-term use). An important task when analyzing these inventories is to identify and quantify damage concentrations or clusters, as this information is useful for prioritizing post-disaster recovery activities. Additionally, an understanding of damage concentrations can provide insight to the multiscale processes that govern an urban region's performance during an extreme event.

In some cases spatial patterns and clusters can be inferred from damage inventories using simple, qualitative visual assessment techniques. While this may be a satisfactory approach in situations where there is a marked contrast in building performance, its effectiveness is limited when damage contrasts are subtle, and spatial patterns are less obvious. In these instances, more advanced spatial analysis tools such as point pattern analysis can be of benefit.

Point pattern analysis (PPA) techniques are a group of quantitative methods that describe the pattern of point (or *event*) locations and determine if point locations are concentrated (or *clustered*) within a defined region of study. An early and often-cited example of a semi-

qualitative application of the PPA concept is physician John Snow's mid-nineteenth century investigation of a cholera outbreak in London (Johnson, 2006). By mapping the locations of drinking water pumps along with the residences of individuals suffering cholera-related illness, Snow was able to link the epidemic to the local water supply. More recently, a rigorous statistical framework for PPA has largely emerged from work within the plant ecology research community. Since the advent of Geographical Information Systems (GIS), PPA has been used with increasing frequency in a range of applications including identification of crime patterns (e.g. Ratcliffe & McCullagh, 1999) and tracking of disease outbreaks (e.g. Lai et al., 2004).

This chapter will review methods from three classes of PPA within the context of an assessment of a high quality building damage inventory. The mathematical formulation of PPA methods have been discussed in detail elsewhere (e.g. Diggle, 2003; Wong & Lee, 2005; Illian et al. 2008) and therefore will not be repeated here. Instead, this chapter will focus on the *application* of PPA techniques and the *interpretation* of results for an urban damage inventory compiled after the 2001 Southern Peru earthquake. Results of the analyses will be compared and discussed along with other pertinent issues. In fitting with the theme of this volume, this chapter is intended to give readers less familiar with spatial analysis a basic framework for understanding key concepts of PPA. More detailed discussions of the techniques discussed in this chapter can be found in Fotheringham et al. (2000), O'Sullivan & Unwin (2003), Fortin & Dale (2005), Mitchell (2005) and Pfeiffer et al. (2008), among other excellent references. Although the chapter is geared toward urban damage inventories, the concepts presented here are appropriate for a wide range of applications in urban engineering and policy (Table 1). Thus it is hoped that this work will inspire more frequent and innovative use of spatial analyses in urban engineering practice and research.

Discipline	Points/Events	Application
Infrastructure Engineering	Underground service repairs	Plan/prioritize maintenance and future upgrades to system
Infrastructure Engineering	Manufacturing centers	Site specific municipal services facilities such as recycling centers
Transportation Engineering	Automobile accidents/pedestrian incidents	Identify roads and intersections requiring safety enhancements
Transportation Engineering	Persons	Siting of transit hubs and connections
Environmental Engineering	Environmental monitoring locations	Identify and track pollution point sources
Civil Engineering	Landslides	Hazard zonation
Public health	Water-borne disease outbreaks	Drinking water quality evaluation
Environmental Science	Urban wildlife sightings	Assess wildlife nesting or migration habits

Table 1. Example of applications of PPA in Urban Engineering and Policy Making

## 2. Case Study of Damage in San Francisco (Moquegua, Peru) during the 2001 Southern Peru Earthquake

### 2.1 Overview

The 23 June 2001 moment magnitude ( $M_w$ ) 8.4 Southern Peru earthquake affected a widespread area that included several important population centers in southern Peru and northern Chile, including Moquegua, the city that will be the focus of this chapter (Figure 1). The earthquake occurred along the active subduction boundary of the Nazca and the South American plates resulting in widespread damage throughout the region. In general, adobe buildings and older structures were most susceptible to damage, though a significant number of modern engineered structures were also impacted by the earthquake. Rodriguez-Marek and Edwards (2003) present a comprehensive overview of the damage caused by the earthquake. Only a limited number of strong motion instruments recorded the main shock, with the largest peak ground acceleration of 0.33 g being measured in the northern Chilean city of Arica. The only ground motion station in Peru, coincidentally located in the city of Moquegua, registered a moderately high peak ground acceleration of 0.30 g.

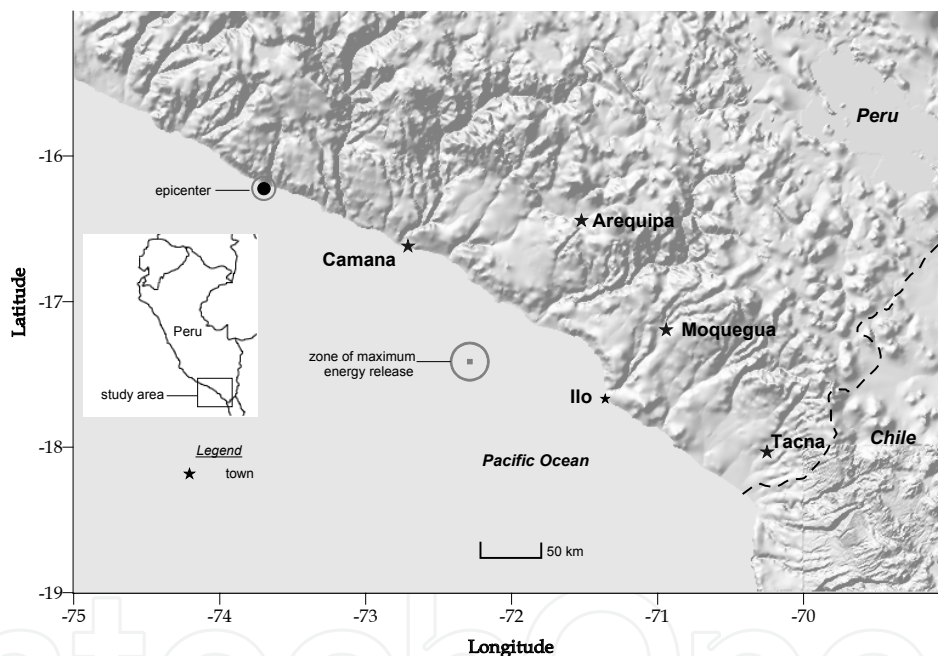


Fig. 1. Regional map showing the location urban centers impacted by the 2001 Southern Peru earthquake

The city of Moquegua (population: 60,000) is situated in an alluvial valley at the base of the Andes Mountains. The city is located approximately 55 km east of the Pacific coast at an elevation of 1400 meters. San Francisco, an approximately 1 km<sup>2</sup> neighborhood located in the southwestern part of Moquegua, was one of the most damaged areas in the city (Figures 2 and 3). In contrast to most of Moquegua, which is relatively flat, San Francisco is distinguished by its variation in topography (Figure 4). San Francisco is situated on a geologic outcrop that includes three ridges rising roughly 100 m above the surrounding portions of the city. This outcrop, which daylights in the upper half of each ridge, consists of stiff conglomerate of the Moquegua geologic formation. This outcrop is also the primary source of alluvium and colluvium that forms a soil mantle that generally thickens with

decreasing elevation. Soil thickness ranges from 0 m on the hillside, to approximately 6 m in valley and flatland areas. San Francisco has grown continuously over the past 40 to 50 years to its 2001 population of 12,000. Buildings in the neighborhood are primarily of masonry or similar construction, with a lesser number of older adobe structures. A summary of land use in San Francisco is presented in Table 2.



Fig. 2. Aerial view of Moquegua showing the San Francisco neighborhood outlined in red. A river is visible at north of the neighborhood. (via Google earth, North is vertical)

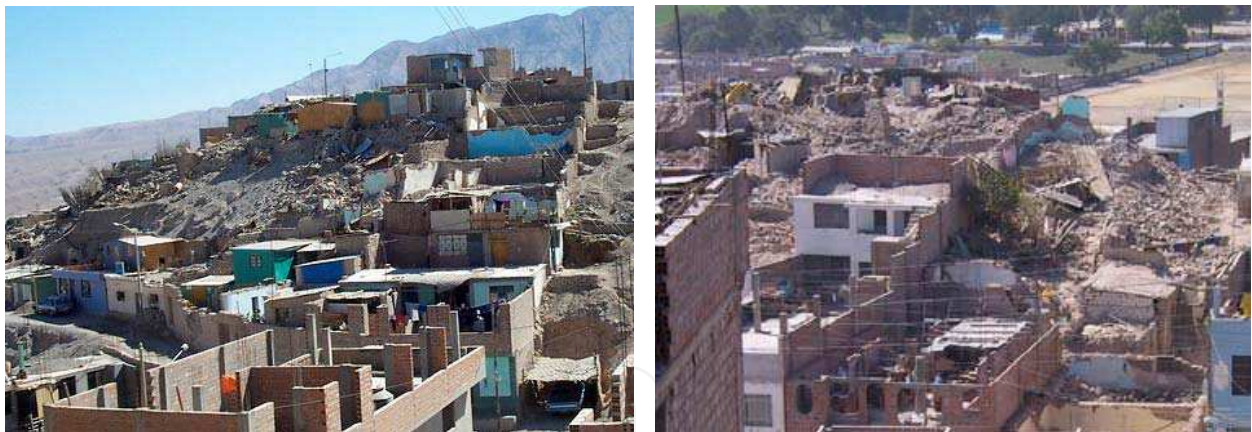


Fig. 3. Building damage in San Francisco after the 2001 Southern Peru earthquake

The absence of earthquake-induced ground failure (i.e., soil liquefaction and landslides) in San Francisco suggested that the high levels of building damage were a result of strong localized shaking. Several preliminary post-earthquake investigation reports (e.g. Kosaka-Masuno et al., 2001; Kusunoki, 2002; Rodriguez-Marek et al., 2003) hypothesized that the high levels of damage were due to topographic amplification (Kramer 1996) of ground motion, resulting in localized strong ground shaking. This phenomenon, where topographic features (e.g. hills and ridges) alters and amplifies local ground shaking, has been observed in past earthquakes and is most pronounced near ridge tops. Given the topography of San Francisco, this was a plausible explanation; however, later published data suggested that damage concentrations were located away from ridge tops, indicating that other factors may

have governed localized ground motion intensity in the neighborhood. A question then remains: what role, if any, did topography play in the damage distribution in San Francisco? This chapter will consider the topographic amplification question further by conducting a series of analyses to determine if building damage in San Francisco was clustered and if so, to see if the cluster locations coincide with areas of relief as would be expected with topographic amplification.

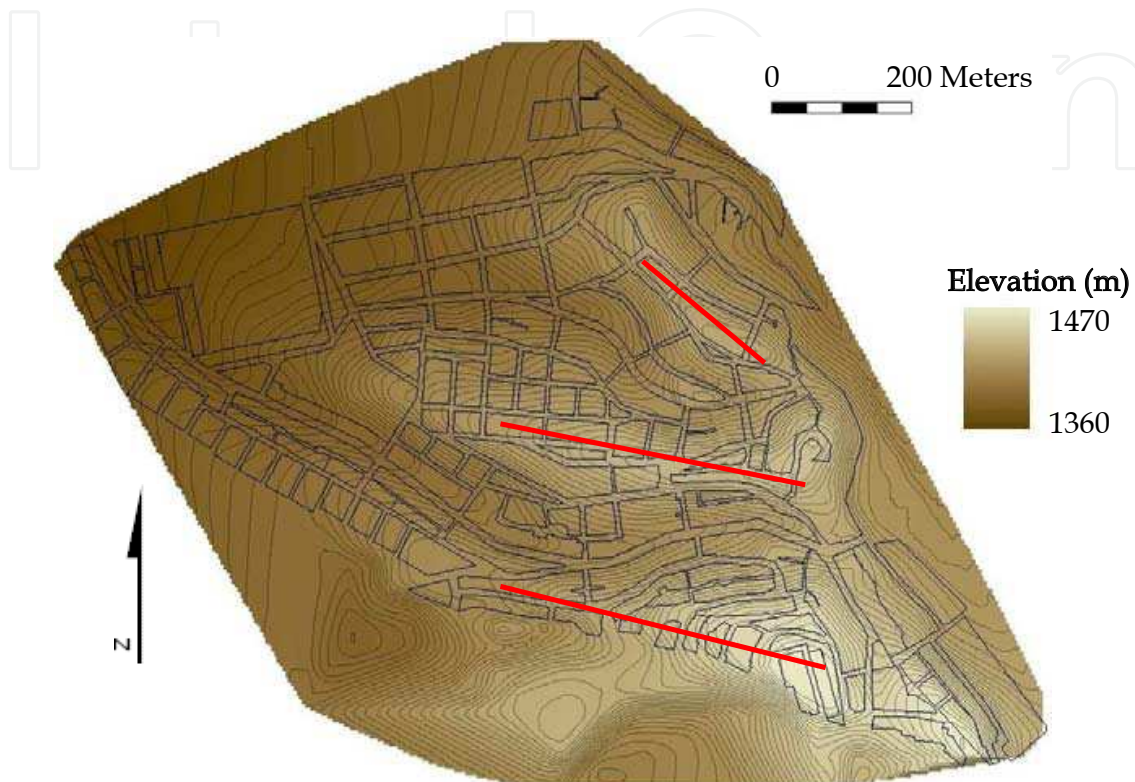


Fig. 4. Street map and topography of San Francisco. The red lines indicate the locations of ridgetops

Land Use	Number of Land Parcels	Percent of Land Parcels
Residential	1611	76.4%
Commercial	89	4.2%
Government	89	4.2%
Vacant	320	15.2%
<b>Total =</b>	<b>2190</b>	<b>100%</b>

Table 2. Summary of land use in San Francisco

## 2.2 PREDES Damage Inventory

Several earthquake damage inventories for the region have been published, including a high quality, comprehensive account produced by Peru's Center for the Study and Prevention of Disasters (PREDES, 2003). This inventory was developed as part of a larger effort by local engineers, architects and social scientists to assess the effectiveness of short term disaster

response in the neighborhood. The PREDES survey was based on individual inspections of close to 1900 buildings whose seismic performance was rated as good, moderate or poor, corresponding to buildings that exhibited no significant damage, significant cracks to loading bearing members, and collapse, respectively. The survey also categorized each building according to its typology (Table 3), as this is known to be an important factor governing the seismic performance of structures. Buildings comprised of masonry and "mixed" construction (i.e., combined masonry and adobe materials) are widely recognized to be of higher construction quality (and seismic resistance) than adobe dwellings.

Building Type	Percentage of Buildings Exhibiting Damage			No. of Buildings (% of Study)
	Low	Moderate	High	
Masonry	81.7%	16.5%	1.8%	968 (52.0%)
Mixed	31.2%	56.6%	12.2%	548 (29.5%)
Adobe	2.6%	21.8%	75.6%	344 (18.5%)

Table 3. Building typology and damage statistics for San Francisco

Figure 5 shows the locations of buildings in San Francisco and their respective typologies. The open areas in the northwest corner of the study area are the locations of two undeveloped land parcels. The study zone was defined to include only the areas where complete data was available for all buildings. This was required because a full data inventory (rather than a sampling) is needed to properly conduct a PPA. Figure 6 shows the post-earthquake damage state of each building in the study area.

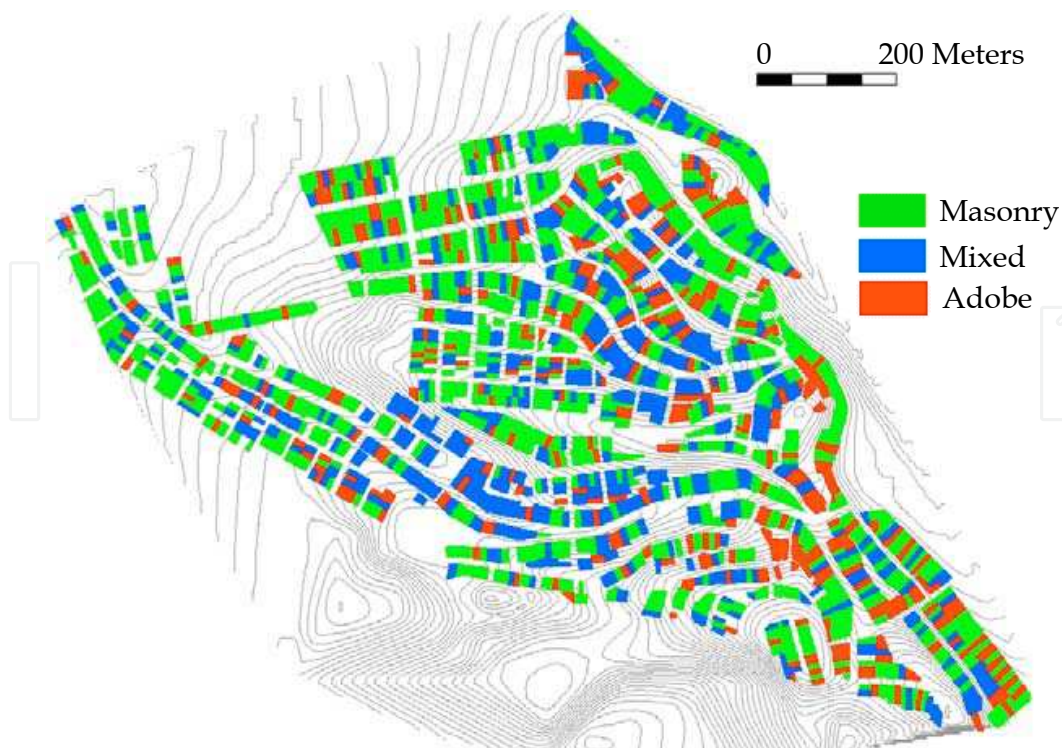


Fig. 5. Building locations and typologies in San Francisco (after PREDES, 2003)

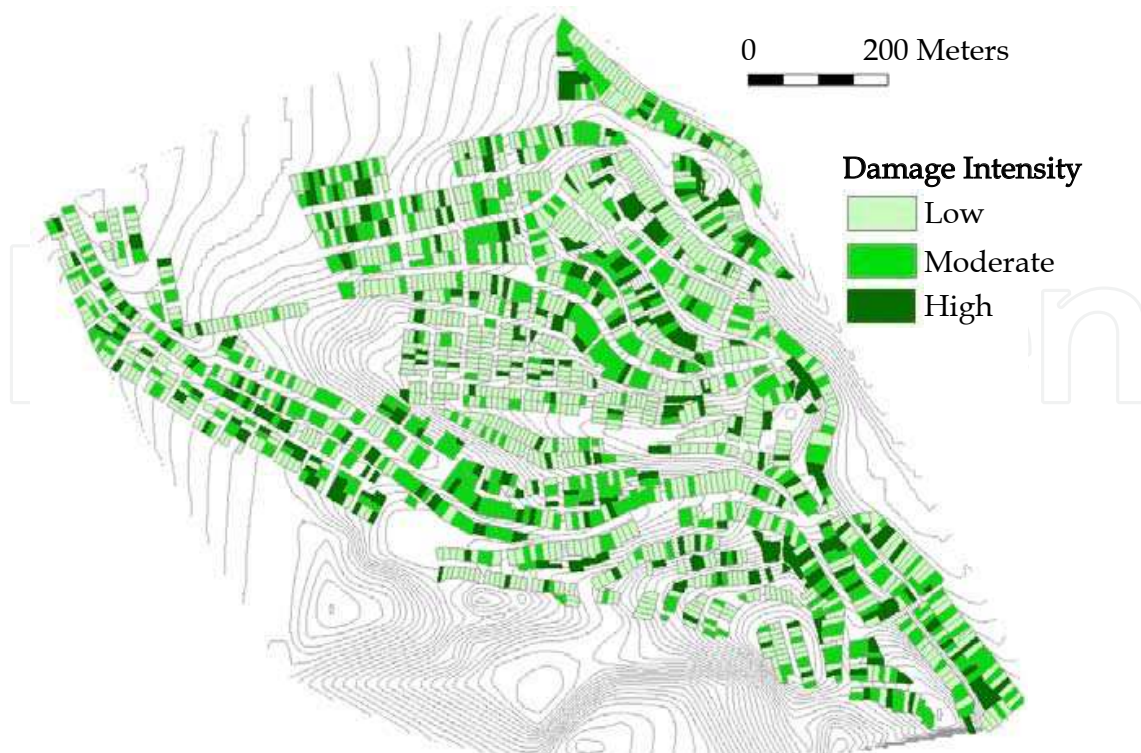


Fig. 6. Post-earthquake damage condition for all structures located in San Francisco (after PREDES, 2003)

Review of Figures 5 and 6 along with the data included in Table 3 indicates the following:

- With some minor exceptions related to a concentration of mixed dwellings located in the south-central portion of San Francisco, the building types appear to be generally well distributed throughout the neighborhood. Most streets are characterized by interspersed masonry, mixed, and adobe dwellings.
- A majority of the buildings in San Francisco are of masonry or mixed construction.
- Overall, a majority of the higher quality (i.e., masonry and mixed structures) buildings in the neighborhood performed well (“low” damage intensity) in the earthquake. In contrast, the seismic performance of nearly all of the adobe structures was poor (“high” damage intensity).
- It is difficult to identify any clear patterns on concentrations of damage from the overall damage survey using visual assessment alone.

The combination of different building types and performance levels make the damage inventory quite complex when considered in aggregate. Although advanced multivariate statistical techniques could be used to analyze the data, PPA provides a more direct approach tailored to the principal objective of the study (i.e., identification of damage clusters). However, this technique requires consistency in the database and thus the PREDES data inventory was modified to ensure uniform building construction quality. Specifically, adobe buildings, which have significantly lower seismic resistance than masonry and mixed structures, were not considered. As these comprised only a small fraction of the total building inventory, this did not significantly reduce the total number of buildings in the inventory. The performance of the remaining 1,513 buildings in the inventory (i.e., masonry and mixed structures) was re-categorized in a binary manner as “no



damage/moderate damage” and “collapse”. As collapsed buildings (rather than partially damaged) were responsible for most of the injuries and fatalities in the earthquake, the PPA was focused specifically on this damage category. Based on these refinements to the original PREDES database, the point events for the PPA are thus defined as masonry and mixed construction buildings (i.e., higher quality structures) that collapsed in the earthquake. The locations of these damaged buildings, which comprise 5% of the structures in the refined database, are shown in Figure 7. Visual inspection of this figure fails to reveal any marked or otherwise obvious concentrations of collapsed buildings. This inventory of collapsed buildings will be considered in a more quantitative manner in the following sections.

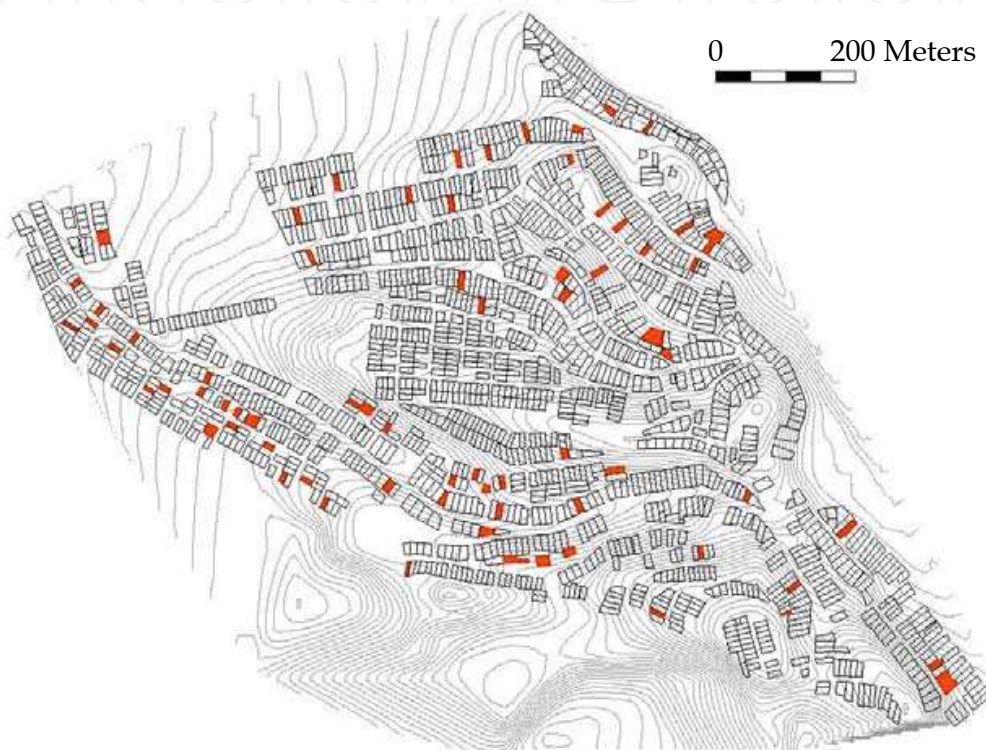


Fig. 7. Collapsed masonry and mixed buildings (shown in red) in San Francisco

### 3. Point Pattern Analysis (PPA) of Collapsed Building Locations in San Francisco

#### 3.1 Data Requirements for PPA

Point patterns consist of a series of spatially distributed points, or *events*. To constitute a point pattern, a set of events must meet the five criteria highlighted below (O'Sullivan & Unwin, 2003) and discussed in the context of the PREDES damage inventory:

1. *The patterns should be mapped on a plane.* Owing to the topographic relief in San Francisco, the events occur over a three-dimensional ground surface. Nevertheless, variations in elevation (1360 m to 1470 m) are small relative to the plan dimensions of the neighborhood (approximately 1000 m by 1300 m) and thus the study area can be reasonably approximated as a plane.

2. *The study area should be determined objectively, rather than arbitrarily.* The boundaries of the San Francisco study area were delineated based on the availability of data; however,

this area fully encompasses the major topographic features of the neighborhood and also includes some of the surrounding flatlands. As such, the PPA can be used to determine if clusters are associated with the neighborhood's topographic features.

3. *The events must be based on a census of the study area, rather than a sampling.* The PREDES inventory was developed based on detailed building-by-building (i.e. census-type) inspections, thus satisfying this criterion.

4. *Objects in the study area must directly correspond to events in the pattern.* The events were defined so as to directly correspond to masonry and mixed construction buildings that collapsed in the earthquake, thereby satisfying this criterion.

5. *Event locations must be proper, rather than representative of a larger object.* The damage event locations were taken as the centroid of a building lot; however, as these lots are relatively small (approximately 180 m<sup>2</sup>) relative to the size of the study area (approximately 750,000 m<sup>2</sup>), this is judged to be a reasonable approach for estimating the event locations.

### 3.2 PPA techniques

There are three basic approaches for conducting a PPA. The first involves determining the number of events within a given area and therefore are referred to as *density-based* methods. The simplest of this class of methods are quadrat counts, whereby quadrats are drafted over the study area and the number of events in each is counted to determine quadrat densities. A more sophisticated but conceptually similar density-based method involves kernel density functions, whereby events are summed within a series of circular regions centered at a given location within the study area. The second approach for PPA involves the determination of distances between events, and is thus referred to as a *distance-function*. The third approach to PPA considers *spatial associations that vary locally* from the larger global trends across a region of study. As will be discussed below, most of these approaches can be combined with statistical analysis to obtain more rigorous quantitative description of point patterns and clustering.

#### 3.2.1 Quadrat Count Methods

Quadrat counts are perhaps the simplest and easiest of the PPA techniques to understand and use. With this method the intensity of a point pattern ( $\lambda$ ) is computed as:

$$\lambda = \frac{n}{\alpha} \quad (1)$$

Where  $n$  is the number of events within a quadrat and  $\alpha$  is the area of individual equal dimension quadrats imposed over the study area.

Figure 8 shows grids of 50 m, 100 m, and 150 m square quadrats over the study area. Each grid was oriented in a north-south/east-west orientation, with the origin located at the south-west limit of the neighborhood. It should be apparent that grid orientation can affect the resulting quadrat counts, especially when larger grid sizes are used. The color intensity on the figures is proportional to the number of points (events) in each quadrat. Note that the value of the color scale varies between each of the three diagrams.

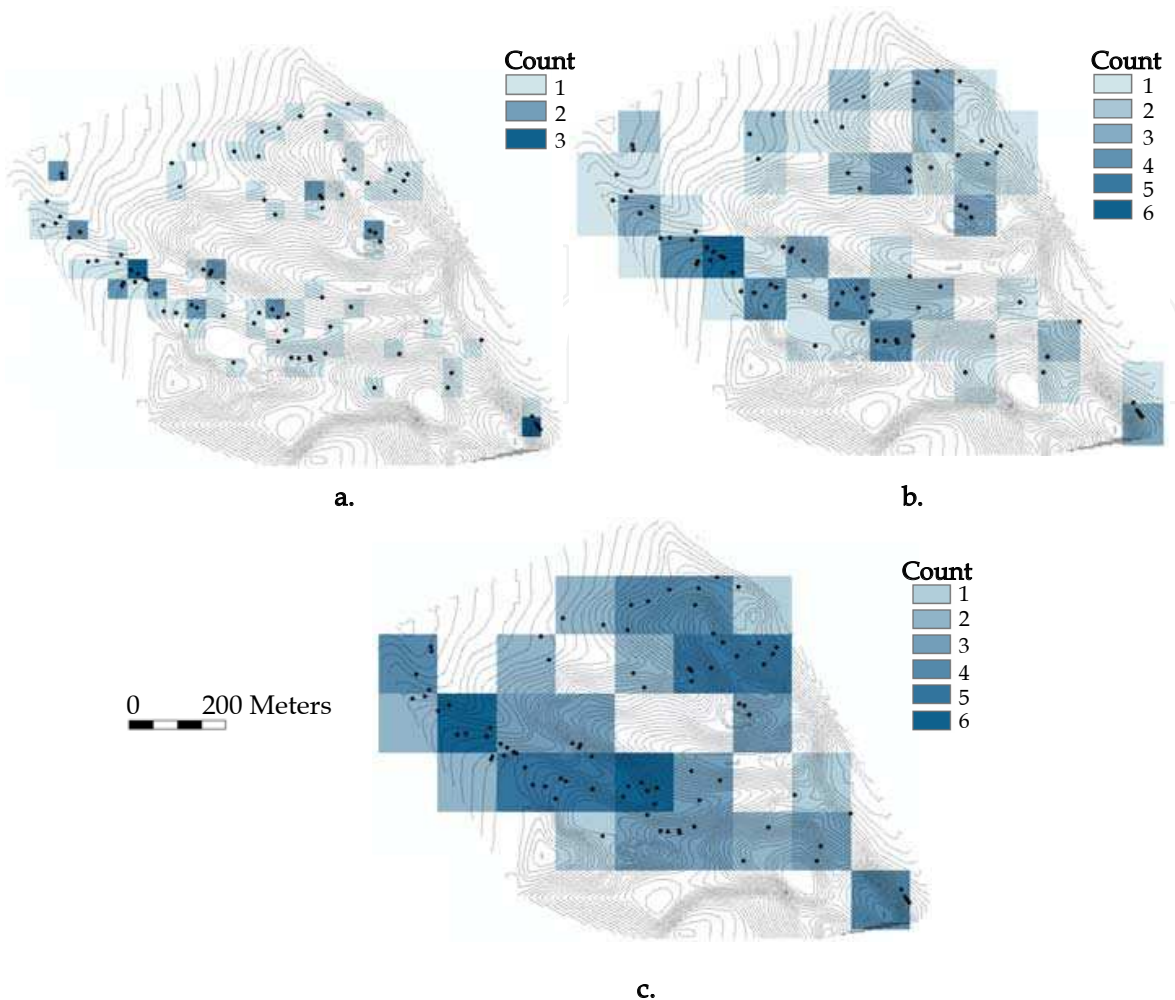


Fig. 8. Results of the quadrat count method for (a) 50 m, (b) 100 m and (c) 150 m square grids

Observe that different quadrat sizes capture different apparent trends in the point patterns. For example, the highest intensity quadrats for closest-spaced grid (50 m) correspond to locations where immediately adjacent or nearly adjacent buildings were damaged, while the largest grid (150 m) captures trends over a significantly larger grouping of dwellings (i.e., street block scale). This observation is not surprising and highlights the scale-dependence of the quadrat count approach. Ideally, the size of the quadrats should reflect the scale of process or mechanism governing the response of the larger spatial system. The question under consideration in this case pertains to the role that relief played in the distribution of collapsed buildings in San Francisco, and thus it is worth considering the scale of the key topographic features in the study area. The local topography is dominated by several ridges and valleys approximately 50 m to 150 m in width, and therefore a reference dimension in this range is appropriate for the analysis. Alternatively, a more general guideline proposed by Greig-Smith (1952) can be used to determine the optimal length ( $d$ ) of a square quadrat:

$$d = \sqrt{\frac{2A}{n}} \quad (2)$$

Where  $n$  is the number of points within the entire study area  $A$ .

Applying the Greig-Smith (1952) guideline to the San Francisco damage inventory yields an optimal quadrant dimension of 132 m, which is in the range of the grid spacing considered here. The range of grid sizes considered in Figure 8 is not large, and so it is not surprising that generally similar observations are made from each of the diagrams. Namely, the highest intensity of damage occurs in the (i) west, (ii) south-west, and (iii) north-east portions of the neighborhood. Still, the square quadrats appear awkward when overlaid onto the irregularly-shaped study area making it difficult to visually relate high intensity quadrats to the neighborhood's rounded topographic features (ridges and valleys).

Quadrat counts can be combined with relatively simple statistical analysis to formulate more conclusive statements about the data inventory (O'Sullivan & Unwin, 2003). For example, if one assumes that the events follow a Poisson distribution, the computed and expected intensities can be compared to see if a null hypothesis of a uniform distribution is obtained. Table 4 computes the variance for the 150 m grid quadrat. The mean quadrat count ( $\mu$ ) is 81 events/42 quadrats, or 1.93 events/quadrat. The observed variance can be computed as the sum-of-squared difference (158.21; see Table 4) divided by the number of quadrats (42) to obtain 3.76. This observed variance is a factor of two greater than the mean quadrat count ( $3.76 > 1$ ), thus indicating a high coefficient of variation among the quadrat counts, which is indicative of clustering. Repeating this exercise for the 100 m (variance/mean = 1.59 > 1) and 50 m (variance/mean = 0.33 < 1) grids, suggesting less pronounced clustering and likely no clustering, respectively. This finding is commensurate with the data presented in Figure 8, which shows greater variability in the intensities (i.e., wider range scale) for the larger grids.

Number of Events (k)	Number of Quadrats (x)	k - $\mu$	(k - $\mu$ ) <sup>2</sup>	x (k - $\mu$ ) <sup>2</sup>
0	16	-1.93	3.72	59.51
1	3	-0.93	0.86	2.59
2	6	0.07	0.01	0.03
3	7	1.07	1.15	8.04
4	4	2.07	4.29	17.16
5	4	3.07	9.43	37.73
6	2	4.07	16.58	33.15
<b>Total</b>	42			158.21
			$\mu =$	1.93
			$\chi^2 =$	82.04

Table 4. Quadrat count and variance calculation for 150m grid quadrat

A chi-squared ( $\chi^2$ ) statistical test can also be made to compute the probabilities that the data distribution is clustered. This approach also has the added advantage to associate a level of confidence with the results of the analysis. A chi-squared distribution is expressed as:

$$\chi^2 = \frac{\sum(k - \mu_k)^2}{\mu_k} \quad (3)$$

Where  $\mu_k$  is the mean quadrat count and  $\sum(k - \mu_k)^2$  is the summation of count difference from the mean quadrat count squared.

Returning to Table 4, the chi-squared value can be compared to standard values for various confidence levels. For the 150 m grid, there are 32 degrees of freedom (i.e. number of quadrats minus 1), and thus the chi-squared values for 99.9% and 99.0% confidence levels, are 62.5 and 53.5, respectively. Comparing these values, it is observed that the actual chi-squared value (79.1) exceeds the both of the critical values, and thus it can be said that the observed quadrat count would be expected to occur in less than 0.1% of the time in random simulations. This high confidence level indicates that the point patterns are very likely to be clustered rather than uniformly distributed. This same statement can also be made for quadrat counts with the 100 m grid ( $\chi^2 = 109.8 < 124.3$ ) and the 50 m grid ( $\chi^2 = 313.4 < 368.3$ ). This finding is commensurate with results of the Poisson-based null hypothesis described earlier.

Before describing the other density method, it is worth noting that there are a number of simple variations on the basic quadrat count theme described here. For example, Wong & Lee (2005) describe the use of hexagon rather than square grids, while O'Sullivan & Unwin (2003) discuss the application of randomly placed sampling quadrats in lieu of the non-overlapping grided census-type described above.

### 3.2.2 Kernel Density Functions

Kernel density functions comprise a second category of density-based techniques for evaluating point patterns. In the context of spatial analysis, kernel densities are "moving" three-dimensional functions that weight events within a sphere of influence based on their distance to a point where density is being estimated (Gatrell et al., 1996). The basic form of a kernel estimator is:

$$\hat{\lambda}_p = \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{s - s_i}{\tau}\right) \quad (4)$$

Where  $\hat{\lambda}_p$  is the density of the spatial point pattern measured at location  $s$ ,  $s_i$  is the observed  $i$ th event,  $k$  is the kernel weighting function, and  $\tau$  is the bandwidth (Borruso, 2008). The simplest of the kernel density functions is the "naive" function, which computes point pattern intensity ( $\hat{\lambda}_p$ ) based on a circle centered at the location where a point density is to be estimated (O'Sullivan & Unwin, 2003). Mathematically, this is expressed as:

$$\hat{\lambda}_p = \frac{no.[S \in C(\mathbf{p}, r)]}{\pi r^2} \quad (5)$$

Where the numerator is the number of events of pattern  $S$  within  $C(\mathbf{p}, r)$ , a circle of radius  $r$  centered at the location of interest  $\mathbf{p}$ . The denominator in this equation ( $\pi r^2$ ) is the area of the circle.

The kernel density estimation differs from the quadrat count approach in two key respects: (i) it includes points at but also beyond the location where the intensity is computed, and (ii) kernel density values can be computed at locations throughout a study region to develop contours of intensity values. A benefit of this latter aspect is that it allows discrete point data to be represented as a continuous surface over a study area. These continuous surface values

can then be used in direct comparisons with other continuous data fields (e.g. ground surface elevation). The surfaces can also be used to modify or normalize density functions to account for local variations in sample density. For example, in their study of spatial patterns of a disease outbreak, Lai et al. (2004) account for variations in population over the study region by normalizing kernel densities by population density to compute incidences on a per capita basis.

Figure 9 presents kernel density (points/km<sup>2</sup>) contours of collapsed buildings for radius values (sample windows) of 50 m, 100 m, and 150 m. Because point densities can be represented as smooth contours, it is easier to visually assess patterns in the study area. The contours of kernel densities based on a 50 m radius (Figure 9) highlight the location where damage occurred, but owing to the relatively small radius, do little to reveal patterns or larger-scale associations between the points. However, kernel density contours based on the 100 m radius begin to show some suggestion of localized clustering along and within a north-west/south-east corridor located in a small valley in the west portion of the study area. This effect is even more pronounced for the data from the 150 m radius kernel density, where the intensity of the points is shown to be quite high along the center portion of the apparent damage corridor.

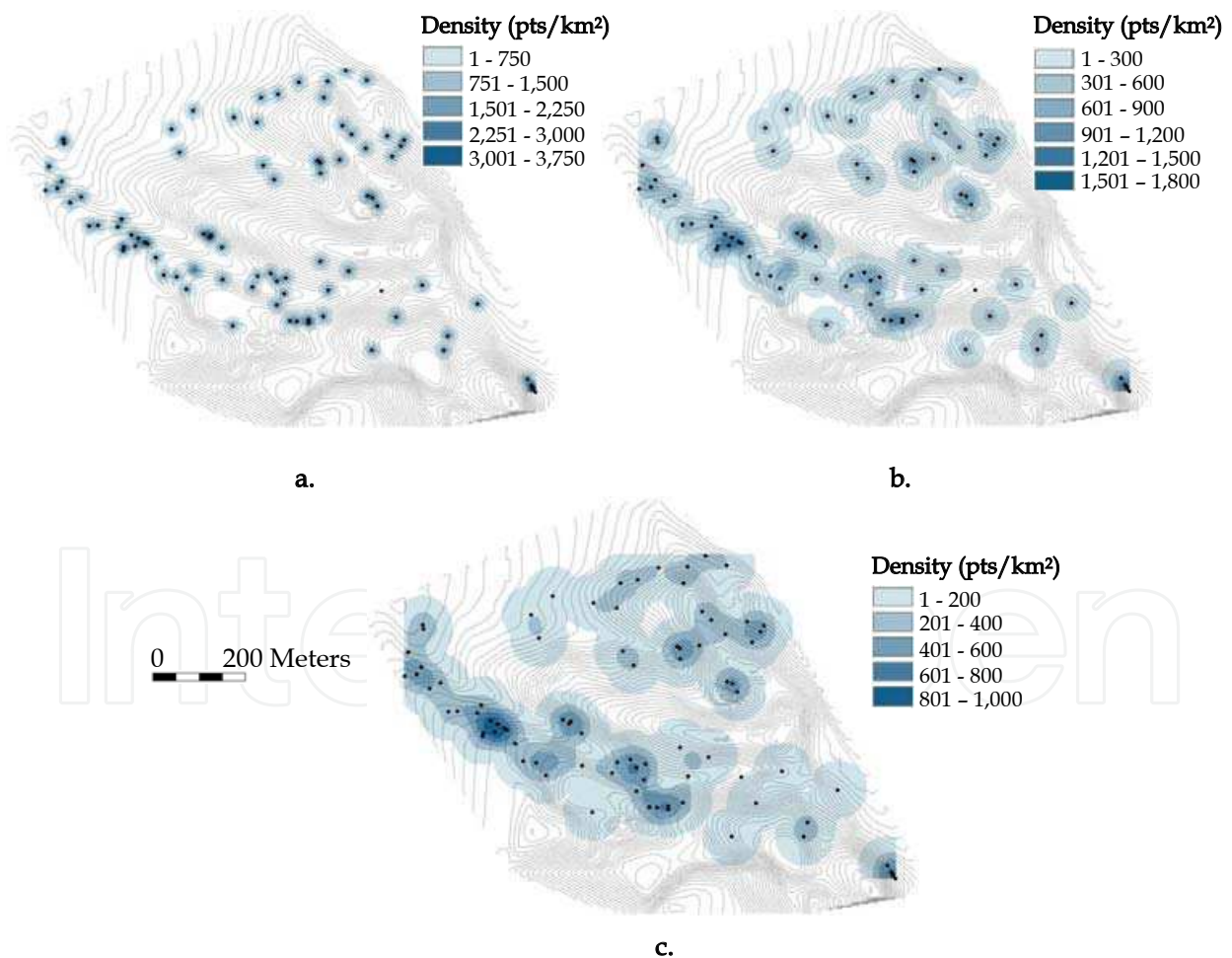


Fig. 9. Results of the kernel density method for (a) 50 m, (b) 100 m and (c) 150 m diameter kernels

Because kernel densities are computed throughout the entire study area, including areas near the periphery where the sampling window may extend into adjacent areas (where point data is not available), artificially low values may be computed in some locations. This phenomenon, referred to as edge effects, has only a limited impact on the San Francisco analysis because very few collapses occur near and beyond the periphery of the study area. Nevertheless, in situations where point clusters exist near the edges of a study area, boundary corrections, such as those discussed by Illain et al. (2008), may be required.

### 3.3.3 Distance-Based Point Pattern Analyses

Another approach for analyzing point patterns considers the *distance* between points rather than their density. For a single point, this is typically done by measuring the distance to the closest adjacent point to determine the "nearest neighbor" distance. By repeating this for all points, the cumulative frequency distribution (typically referred to as a "G-function") of these nearest neighbor distances can be computed as (O'Sullivan & Unwin, 2003):

$$G(d) = \frac{no.[d_{\min}(p) < d]}{n} \quad (6)$$

Where  $no.[d_{\min}(p) < d]$  is the number of events which have a nearest neighbor less than  $d$ , and  $n$  is the population of the dataset. The value of  $G$  for a distance  $d$  is, in effect, the fraction of the dataset population having a nearest neighbor less than  $d$ .

Plotting the G-function distribution can provide insight to the potential clustering of points, and as with the density-based techniques, the utility of the G-function is enhanced when considered relative to a uniformly distributed or background point pattern. Figure 10 shows the G-function distribution for collapsed buildings located in the study area. The function rises sharply between approximately 10 m and 80 m, indicating that most of the nearest neighbor distances are in this range. At longer distances the distribution continues rising, though at a notably diminished rate before flattening and terminating. An easy qualitative check of the computed G-function distribution may be made by again reviewing Figure 7, which shows that a significant majority of the points have nearest neighbor distances up to about 100 m.

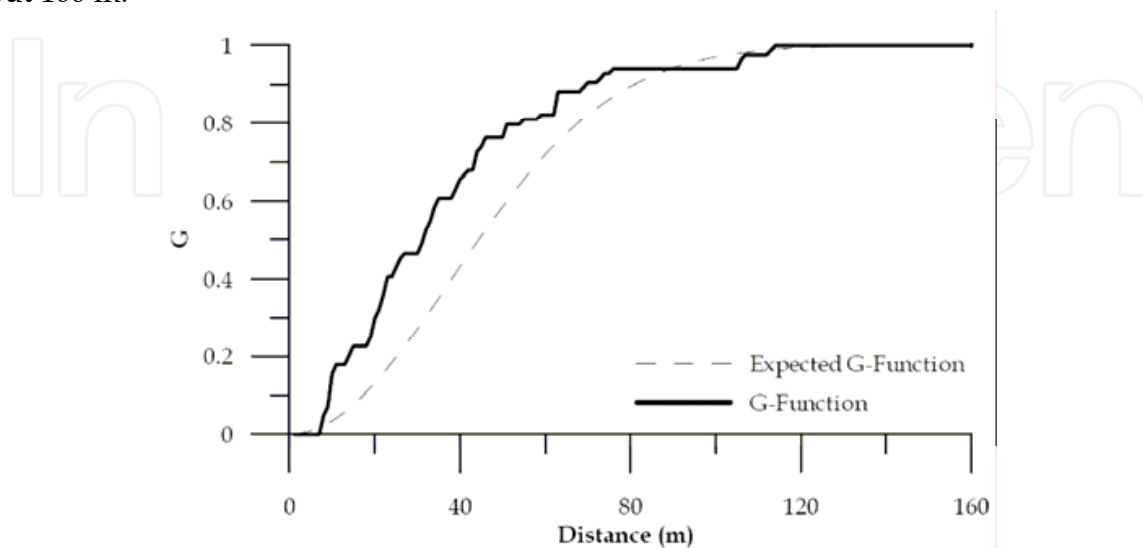


Fig. 10. Observed and expected G-distance function

The dashed line in Figure 10 is the expected G-function distribution for uniformly distributed points. In contrast to the observed data, the expected G-function rises smoothly over the range of distances. Since this function assumes uniform distribution as a null hypothesis, the expected value of  $G$  is equal to one when  $d$  is equal to the study area divided by  $n$  (i.e. 10 points distributed uniformly in an area 10 m<sup>2</sup> would each have an observed nearest neighbor at 1 m). A key observation from Figure 10 is that the observed G-function distribution is positioned to the left of the expected distribution for nearest neighbor distances up to approximately 65 m. This implies a greater than expected number of nearest neighbor distances over this range, and thus indicates clustering of points.

The K-function (Ripley, 1976) is a popular alternative to the G-function that uses the distance between *all* points (rather than only the nearest neighbor distance) to assess clustering. As such, the K-function effectively considers a distribution of events in proximity to a point being considered. A principal advantage of the K-function is that because it uses all distances it better captures clusters that occur at relatively short distances, which tend to be obscured by G-functions (Ripley, 1976). The K-function is calculated by placing a circle of radius  $r$  around a point and summing the total number of points located inside the circle. This is repeated for the all points so that the mean count of all events is computed. The mean count value is then divided by the point density for entire study area to obtain  $K(r)$ :

$$K(r) = \frac{\sum_{i=1}^n no.[S \in C(\mathbf{p}, r)]}{n\lambda} \quad (7)$$

Where  $\sum_{i=1}^n no.[S \in C(\mathbf{p}, r)]$  is the mean of points counted within a circle of radius  $r$  centered at each point in the dataset,  $n$  is the number of points in the entire dataset, and  $\lambda$  is the density of points per unit area in the entire dataset.

This calculation is repeated for different radii to obtain a cumulative frequency distribution such as that shown in Figure 11 for collapsed buildings in San Francisco. Interpretation of a K-function is best made in comparison with an expected distribution for uniformly distributed points, or alternatively, against a "background" or control distribution consisting of an entire population (both events and non-events) of points. For San Francisco, the latter entails the location of the entire population of masonry and mixed-construction buildings, regardless of their damage condition, while the former assumes a uniform distribution of points corresponding to collapsed buildings. An advantage of the background over an assumed uniform distribution is that it reflects the actual spatial distribution of vulnerable buildings. Figure 11 includes the K-functions for the collapsed buildings, the control population, and a uniform distribution of buildings. As is typically the case, the arithmetic differences between the observed and expected functions may be small relative to the overall scale of the K-function thereby making it difficult to discern trends from the data (Figure 11). In these instances the data can be presented in other formats to make the trends more apparent. When a control (or background) population is available, a difference function ( $K_{diff}$ ) can be computed as (Pfeiffer et al. 2008):

$$K_{diff}(r) = K(r)_{damage} - K(r)_{control} \quad (8)$$

Where  $K(r)_{damage}$  is the value  $K(r)$  for structures with high levels of damage occurred, and  $K(r)_{control}$  is the  $K(r)$  with respect to all masonry and mixed buildings in the neighborhood.



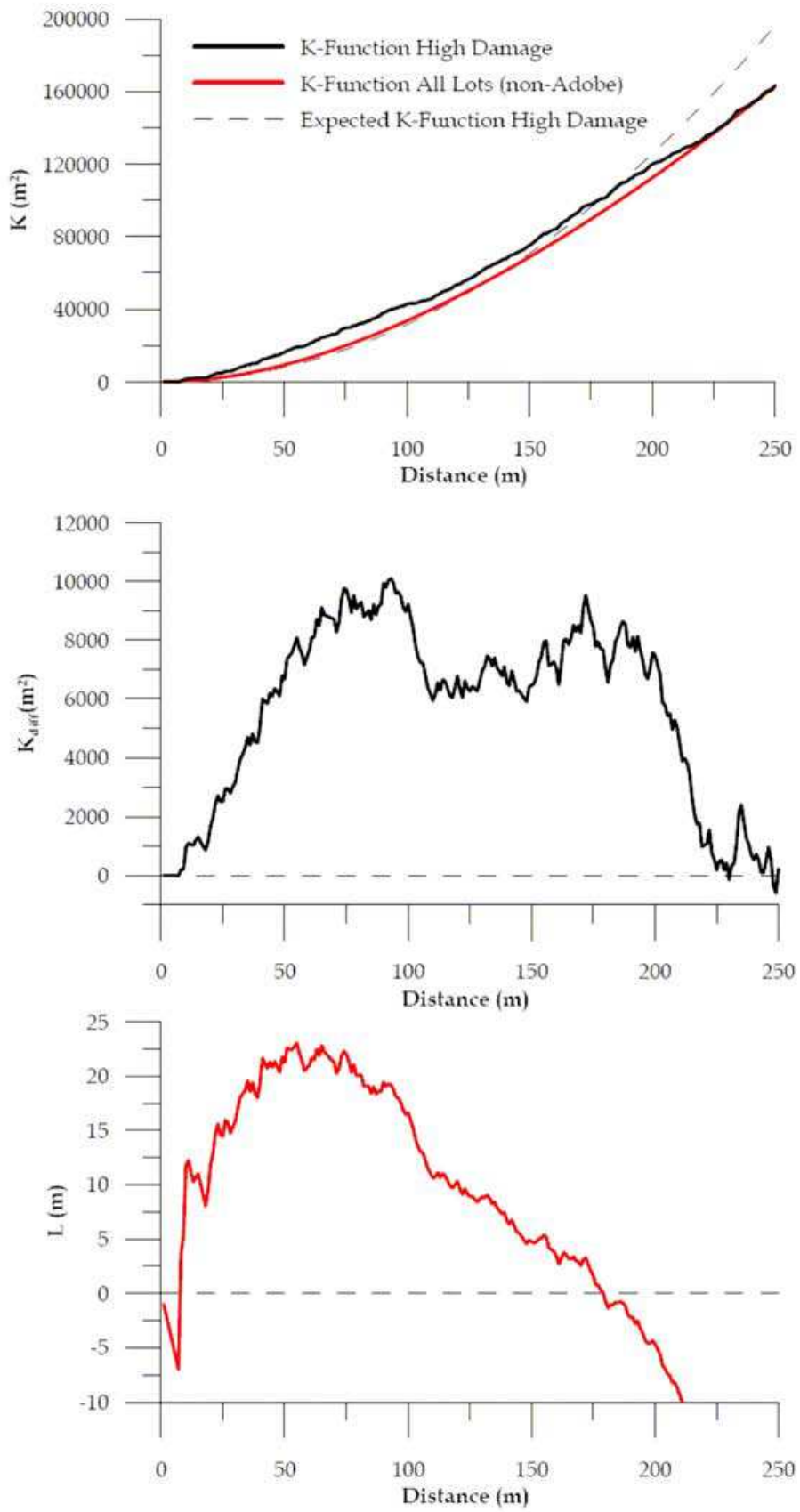


Fig. 11. Observed and expected values for K-, K<sub>diff</sub>- and L-functions

Another comparison can be made with a uniform distribution of points by computing an L-function (O'Sullivan & Unwin, 2003):

$$L(r) = \sqrt{\frac{K(r)}{\pi}} - r \quad (9)$$

The L-function effectively re-scales the K-function so that observed values plot as zero when they match the expected K-function (i.e., when the observed and uniform K-functions are equal, indicative of no clustering). Figure 11 includes both the L-function and  $K_{\text{diff}}$  function and for the San Francisco data. The L-function values are greater than the expected value (zero) over the distance range of 10 m to 130 m, with the largest differences occurring between 25 m and 100 m. This is indicative of a high degree of clustering at these distances. At larger distances the values begin to deviate widely in an opposite manner, indicating that the observed points are more widely distributed than expected for a uniform distribution of points. Rather than representing an actual trend in the data, this is a remnant of edge effects, and highlights why it is necessary to consider boundaries when interpreting K-function distributions. In this instance, the deviation occurs because although there are no points beyond the edge of the study zone (thus widely spaced point distributions cannot exist), these areas are nevertheless included in the K- and L-function computations. The  $K_{\text{diff}}$  function provides additional insight to the clustering of collapsed buildings. Similar to the L-function, the  $K_{\text{diff}}$  function shows peak values in the range of about 50 to 120 m; however, it also reveals a secondary clustering in the range of about 160 to 200 m. This secondary clustering distance corresponds to the distance between the valleys where much of the damage was in fact located.

### 3.3.4 Local Indicators of Spatial Association

A more recent approach for identifying and quantifying clusters involves use of local indicators of spatial association, or LISA (Anselin, 1995). The approach is based on the concept of *autocorrelation*, meaning that correlations or relationships exist among spatially distributed data. Most often, data are positively autocorrelated, that is, closely spaced data are more similar than more distantly spaced data. The most popular of the LISA approaches was developed by Getis and Ord (1992), and later modified by Ord and Getis (1995) to what has is commonly known as the  $G_i^*$  ("G-i-star") statistic. This approach computes the spatial autocorrelation structure over defined sampling windows and compares this to the global autocorrelation structure for the study region. Significant differences between local and global autocorrelation structures imply localized deviations from the global norm, implying localized clustering of events. The  $G_i^*$  statistic is computed as:

$$G_i^* = \frac{\sum_j w_d^*(i, j)z(j)}{\sum_j z(j)} \quad (10)$$

Where  $z(j)$  are known attribute values at a known point (i.e. low/moderate damage = 1, collapse = 2) and  $w_d^*(i, j)$  is a matrix of spatial weights describing the influence of other known points up to a defined threshold distance (i.e. points within  $d$  are assigned a weight of 1, while those beyond  $d$  are assigned a weight of 0).

Similar to other approaches discussed in this chapter, the  $G_i^*$  statistic is most meaningfully interpreted within a larger statistical framework. In this case, a  $Z$ -score [ $Z(G_i^*)$ ] can be computed for each point by subtracting the expected  $G_i^*$  value from the observed value, and dividing this by the square root of the variance for all points in the defined region of study:

$$Z(G_i^*) = \frac{G_i^* - E(G_i^*)}{\sqrt{\text{Var}(G_i^*)}} \quad (11)$$

A  $Z$ -score of 0 implies no clustering, while increasingly positive or negative values suggest point concentrations. The statistical significances of clustering can be determined by comparing the computed  $Z$ -score with a range of values for a given confidence level (Mitchell, 2005). Figure 12 shows the results from this type of analysis for three sampling windows. The filled contours correspond to collapsed building clusters ("hot" spots) associated with statistical confidence levels of 86 % (light red) and 95 % (dark red). The blue contours show the locations of "cold" spots for the same confidence levels.

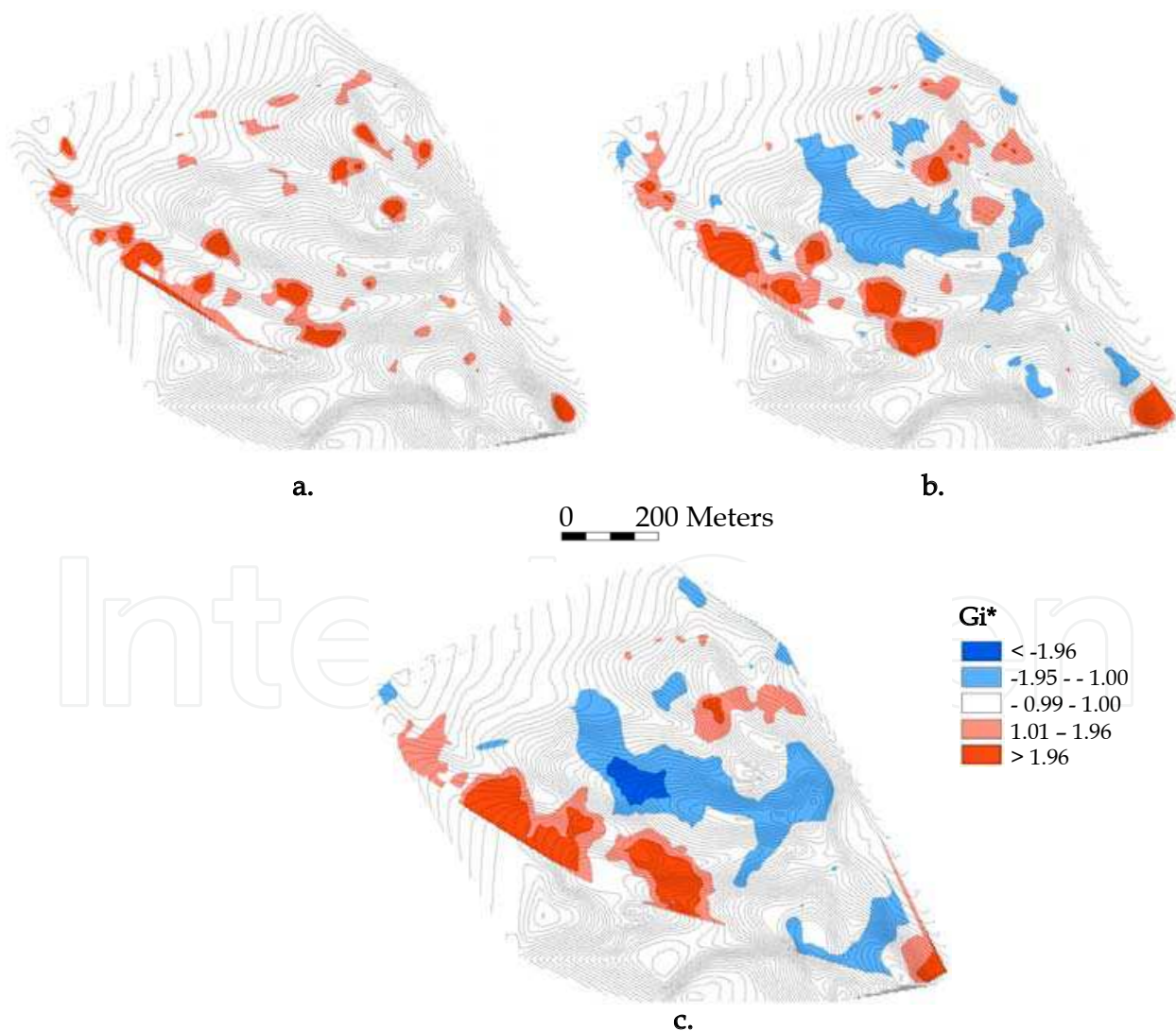


Fig. 12.  $G_i^*$  value contours for (a) 50 m, (b) 100 m and (c) 150 m diameter sample windows.

Note that the damage clusters shown in Figure 12 are quite similar to those identified from the kernel analysis (Figure 9). A unique and especially useful aspect of a LISA analysis is that extremes at both ends of parameter spectrum can be identified (i.e., "hot" and "cold" spots). For example, included on Figure 12 are blue-toned contours corresponding to local clusters of buildings with a low level of damage (i.e., "cold spots" of low damage caused to mixed and masonry buildings). As with the other PPA approaches discussed in this chapter, LISA-based methods are sensitive to the size of the sample window. As shown in Figure 12, larger sampling windows produce large clusters that become increasingly localized as this dimension is reduced. This localization can become so extreme that some clusters no longer appear; for example, note that the cold spot is not present for the smallest of the sampling windows (Figure 12a).

#### 4. Discussion

Although different in concept, the general categories of PPA discussed in this chapter provide generally similar results for the building collapse analysis in San Francisco (Fig. 13).

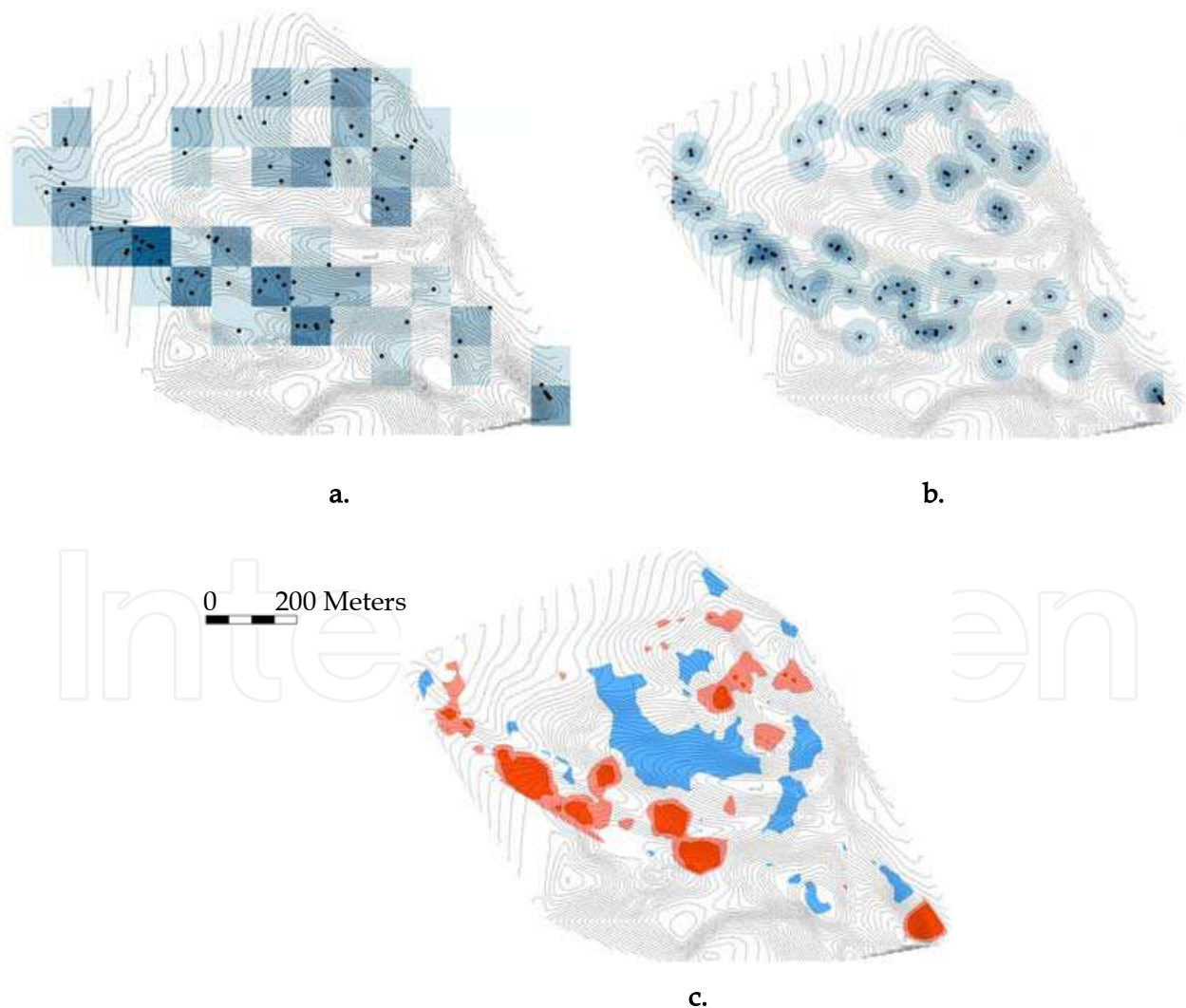


Fig. 13. Comparison of results from (a) quadrat count, (b) kernel density and (c)  $G_i^*$  method for a 100 m sampling dimension. Results correspond to those of Figures 8, 9, and 12

These results are:

1. The locations of the collapsed buildings in San Francisco are locally clustered, with the highest degree of clustering occurring for point (i.e., collapsed building) spacings in the range of approximately 25 to 100 m.
2. Referencing Figure 13, collapsed buildings are clustered along a northwest-southeast trending valley. This cluster extends further into an adjacent flatland area in the western portion of the neighborhood. A second and more highly localized cluster of collapsed buildings is found along a hillslope located in the northeast portion of the neighborhood.

These two findings were not fully apparent from the building damage inventory alone. An additional general observation was that for each of the methods, the size of the sampling window used in the analyses can have an effect on the results of the PPA. When small sample windows are specified, highly (perhaps overly) localized clusters are produced, while larger sample windows yield larger and more diffuse clusters.

There are, of course, differences between the approaches, the most important of which pertain to (i) computational effort, (ii) resolution/visual interpretation, and (iii) the nature of the quantitative interpretations that can be made from the analyses. Quadrant counts, the simplest of the approaches, can be determined manually without the need for computer software. Moreover, statistical analysis of quadrant count data is straightforward. In contrast, calculation of the more sophisticated  $G_i^*$  statistic and related Z-scores is significantly more involved, although modern, commercially available software packages can significantly ease this computational burden.

PPA methods that produce continuous surfaces of contours of clusters intensity (i.e., kernel density and  $G_i^*$  analyses) are relatively easy to interpret visually. In contrast to the quadrant counts, these can be used to identify localized clusters with a high degree of resolution. However, if only the characteristics of the point spacing, rather than the location of clusters are of interest, then the distance based functions ( $G$ -,  $K$ -,  $K_{diff}$  and  $L$ -functions) are most useful.

Perhaps one of the more important differences between the PPA methods pertains to the nature of the quantitative interpretations that can be made from the data. The most comprehensive of these interpretations are made based on the  $G_i^*$  statistic, which provides direct information on both the location of clusters and their statistical significance. While the statistical significance of the quadrant counts and distance based functions can also be determined, these are not directly coupled to high resolution visual presentation of the cluster analysis results.

For the San Francisco case study presented here, the  $G_i^*$  is perhaps the most powerful of the PPA tools for identifying and quantifying point patterns, though as indicated earlier, the other approaches also provided generally similar results, with less computational effort. Perhaps the most useful application of simpler PPA methods is to provide a check on results from other more computationally intensive approaches. Table 5 summarizes the relative merits of the approaches.

Regardless of the method used for analyses, the selection of the size of the sampling window for a PPA analysis is an important consideration that can have a significant effect on the results. For this reason, when performing a PPA it is prudent to consider a range of

scales to better understand the sensitivity of the results to this factor. Ideally, the scale of sampling window should be made relative to a key reference dimension of a process or mechanism governing the response of the spatial system under consideration; however, this is not always known in advance of an investigation. In these instances, criteria developed by Greig-Smith (1952) can provide guidance on an optimal sample window dimension. Edge effects can also have an important influence on the results of a PPA, and measures to account for these may be required for cases where such effects are significant (O'Sullivan & Unwin, 2003).

Finally, note that the collapsed building inventory analyzed in this chapter is static in that it represents damage at a fixed point in time (i.e., damage at the time of the earthquake); it is worth mentioning that PPA can also be used to monitor time-dependant changes in databases having a temporal element.

Category	Method	Benefits	Limitations
Density	Quadrat Count	<ul style="list-style-type: none"> <li>• Easy to perform analysis using simple hand computations</li> <li>• Can be readably combined with statistical analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Square quadrates can obscure visual association of point intensity with underlying process</li> <li>• Orientation and position of grid can affect results</li> </ul>
	Kernel Density	<ul style="list-style-type: none"> <li>• Provides density surface that can be compared or normalized against other continuous variables</li> </ul>	<ul style="list-style-type: none"> <li>• Does not provide a measure of the statistical significance of results</li> </ul>
Distance	G-function	<ul style="list-style-type: none"> <li>• Quantifies point spacing characteristics</li> </ul>	<ul style="list-style-type: none"> <li>• No information on cluster location</li> </ul>
	K-function	<ul style="list-style-type: none"> <li>• Quantifies point spacing characteristics</li> <li>• Allows for direct comparison of events with background points</li> </ul>	<ul style="list-style-type: none"> <li>• No information on cluster location</li> <li>• Can be highly sensitive to edge effects</li> </ul>
LISA	$G_i^*$ statistic	<ul style="list-style-type: none"> <li>• Cluster location directly associated with confidence interval</li> </ul>	<ul style="list-style-type: none"> <li>• Computationally intensive</li> </ul>

Table 5. Benefits and limitations unique to PPA methods featured in this chapter

Returning to a central question of this case study: what role, if any, did topography play in the damage distribution in San Francisco? The PPA results clearly indicate that the largest cluster of collapsed buildings is correlated with flatter ground over the study area, which contrasts with preliminary studies that suggested concentrations of highly damaged buildings were localized at ridge tops. This suggests that topographic implication largely

governed the performance of buildings. Moreover, damage cold spots identified from the  $G_i^*$  analysis (Figure 12) generally overlies areas of topographic relief, which further discredits the topographic amplification hypothesis. The clusters in this study are associated with a high degree of statistical confidence, and thus it is unlikely that the observed concentrations of collapsed buildings were a result of natural background variation. These findings by themselves do not fully disqualify the hypothesis that topographic amplification played a role in damage distribution; however, they suggest that at a minimum these effects were less significant than originally thought. It is possible that the damage distribution was more closely related to "site effects" (Kramer, 1996) whereby earthquake ground motion is amplified by surficial soils and sediments such as those that underlie the small valleys and flatlands of San Francisco. Note that the scale of clustering of collapsed buildings (i.e., 25 to 100 m) corresponds closely with the dimensions of the valleys in the study area, lending further credibility to the site effects hypothesis. This brief discussion highlights both a strength and a limitation of PPA: while this can be used to explore spatial data and test hypotheses about larger governing mechanisms, often they cannot by themselves be used to conclusively identify these larger mechanisms and processes. Aside from investigating the topographic amplification hypothesis, the results of the PPA analyses can serve the important practical function of supporting building code zonation in San Francisco. For example, as the neighborhood is rebuilt and repopulated, it would be prudent to use more seismically resistant construction methods for buildings located in the identified collapse cluster zones as these are likely to again experience high levels of shaking in future earthquakes.

## 5. Conclusions

This chapter introduced three general categories of techniques for identifying and quantifying point patterns in urban engineering applications. These PPA techniques included (i) density-based methods of quadrat counts and kernel density functions, (ii) distance-based  $G$ - $L$ -,  $K_{diff}$ , and  $K$ -functions, and (iii) the local indicator of spatial association (LISA)-based  $G_i^*$  statistic. Each of these techniques was applied to a comprehensive urban damage inventory developed after the 2001 Southern Peru earthquake. While the application of these techniques in a GIS environment is relative straightforward, careful assessment is needed to meaningfully interpret PPA results. This is especially true when considering factors such as the size of the sampling window and the effects of study area boundaries.

Overall, the three procedures all provided generally similar results when applied to the building damage inventory, though differences were noted between each of the methods. The  $G_i^*$  statistic, which provides both the location of clusters and a statistical measure of confidence, proved to be the most powerful PPA technique for analyzing the San Francisco database. Each technique has its relative merits and ultimately the selection of a PPA methodology should be made with consideration of the larger goals of the investigation. For example, distance-based approaches are best for situations when the characteristics of a point pattern spacing are desired, while kernel density function may be best when point pattern data requires normalized against another data field (e.g. population density). Regardless of the technique chosen for analyses, results should be considered in a statistical framework as this can greatly enhance the quantitative interpretation of the results.

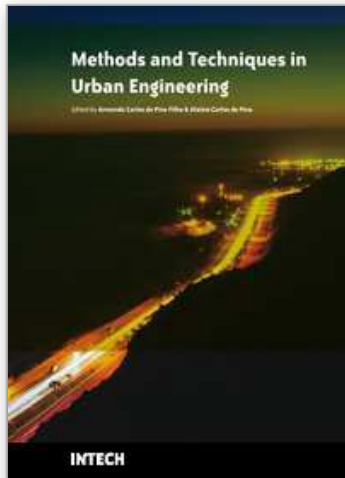
## 6. References

- Ancelin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, Vol. 27, 93-115
- Center for the Study and Prevention of Disasters (PREDES)(2003). Study of the physical environment, social economics and evaluation of the risk in the neighborhood of San Francisco with reference to the District of Moquegua and Mariscal Nieto Province. *Center for the Study and Prevention of Disasters*, 261p.
- Deaton, S. L. & Frost, J. D. (2002). *Integrated Digital Earthquake Reconnaissance*, 7th U.S. National Conference on Earthquake Engineering, CD-ROM
- Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*, Hodder Arnold, 0340740701, London, England
- Fortin, M. & Dale, M. (2005). *Spatial Analysis: A Guide for Ecologists*, Cambridge University Press, 0521804345, New York, New York
- Fotheringham, A. S., Brunsdon, C. & Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*, SAGE Publications Ltd., 0761959475, London, England
- Gatrell, A. C., Bailey, T. C., Diggle, P. J. & Rowlingson, B. S. (1996). Spatial point pattern analysis in geographical epidemiology. *Transactions of the Institute of British Geographers*, Vol. 21, 256-274
- Getis, A. & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, Vol. 24, 189-206
- Greig-Smith, P. (1952). The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany*, Vol. 16, 293-316
- Illian, J., Penttinen, A., Stovan, H., & Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, John Wiley & Sons Ltd., 0470014911, Hoboken, New Jersey.
- Johnson, S. (2006). *The Ghost Map: The Story of London's Most Terrifying Epidemic and How It Changed Science, Cities, and the Modern World*, Penguin Books Ltd., 1-59448-925-4, London, England
- Kosaka-Masuno, M., Gonzales-Zenteno, E., Arias-Barahona, H., Minaya-Lizarraga, A., Faran-Bazan, E., Ticona-Paucara, J. (2001). Seismic Hazard Evaluation in the city of Moquegua. *Convenio UNSA-INDECI. Proyecto PER 98/018 PNUDINDECI, San Agustin de Arequipa National University*, Arequipa, Peru
- Kramer, S. L. (1996). *Geotechnical Earthquake Engineering*, Prentice Hall, 0133749436, Princeton, New Jersey
- Kusunoki, K. (2002). *Report on the damage investigation of the 2001 Atico Earthquake in Peru*. Building Research Institute of Japan. 27 p.
- Lai, P. C., Wong, C. M., Hedley, A. J., Lo S. V., Leung, P. Y., Kong, J. & Leung, G. M. (2004). Understanding the spatial clustering of severe acute respiratory syndrome (SARS) in Hong Kong. *Environmental Health Perspectives*, Vol. 112, No. 15, (November) 1550-1556
- Matsuoka, M. & Yamazaki, F. (2005). Building damage mapping of the 2003 Bam, Iran earthquake using Evnisat/ASAR intensity imagery, *Earthquake Spectra*, Vol. 21, No. S1, 285-294
- Mitchell, A. (2005). *The ESRI Guide to GIS Analysis, Volume 2: Spatial Measurements & Statistics*, ESRI Press, 158948116X, Redlands, California



- O'Sullivan, D. & Unwin, D. J. (2003). *Geographic Information Analysis*, John Wiley & Sons Inc., 0-471-21176-1, Hoboken, New Jersey
- Ord, J. K. & Getis, A. (1995). Local spatial autocorrelation statistics : distributional issues and an application. *Geographical Analysis*, Vol. 27, 286-306
- Pfeiffer, D., Robinson, T., Stevenson, M., Stevens, K., Rogers, D. & Clements, A. (2008). *Spatial Analysis in Epidemeology*, Oxford University Press, 9780198509882, New York, New York
- Ratcliffe, J. H. & McCullagh, M. J. (1999). Hotbeds of crime and the search for spatial accuracy. *Journal of Geographical Systems*, Vol. 1, No. 4, (December) 385-398.
- Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, Vol. 13, 255-266
- Rodriguez-Marek, A & Edwards, C. (2003). Southen Peru earthquake of 23 June 2001 reconnaissance report, *Earthquake Spectra*, Vol. 19, No. S1, pp. 165
- Rodriguez-Marek, A., Williams, J. L., Wartman, J., Repetto, P. C. (2003), Ground motion and site response. *Earthquake Spectra*, Vol. 19, No. S1, 11-34
- Wong, D. W. S. & Lee, J. *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*, John Wiley & Sons, Inc., 0471468991, Hoboken, New Jersey

IntechOpen



## **Methods and Techniques in Urban Engineering**

Edited by Armando Carlos de Pina Filho and Aloisio Carlos de Pina

ISBN 978-953-307-096-4

Hard cover, 262 pages

**Publisher** InTech

**Published online** 01, May, 2010

**Published in print edition** May, 2010

A series of urban problems such as dwelling deficit, infrastructure problems, inefficient services, environmental pollution, etc. can be observed in many countries. Urban Engineering searches solutions for these problems using a conjoined system of planning, management and technology. A great deal of research is devoted to application of instruments, methodologies and tools for monitoring and acquisition of data, based on the factual experience and computational modeling. The objective of the book was to present works related to urban automation, geographic information systems (GIS), analysis, monitoring and management of urban noise, floods and transports, information technology applied to the cities, tools for urban simulation, social monitoring and control of urban policies, sustainability, etc., demonstrating methods and techniques applied in Urban Engineering. Considering all the interesting information presented, the book can offer some aid in creating new research, as well as incite the interest of people for this area of study, since Urban Engineering is fundamental for city development.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Joseph Wartman and Nicholas E. Malasavage (2010). Spatial Analysis for Identifying Concentrations of Urban Damage, *Methods and Techniques in Urban Engineering*, Armando Carlos de Pina Filho and Aloisio Carlos de Pina (Ed.), ISBN: 978-953-307-096-4, InTech, Available from: <http://www.intechopen.com/books/methods-and-techniques-in-urban-engineering/spatial-analysis-for-identifying-concentrations-of-urban-damage>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen