Available online at www.sciencedirect.com
SCIENCE DIRECT®
Applied Geography 24 (2004) 97–117
APPLIED GEOGRAPHY
www.elsevier.com/locate/apgeog

Defining area at risk and its effect in catastrophe loss estimation: a dasymetric mapping approach

Keping Chen *, John McAneney, Russell Blong, Roy Leigh, Laraine Hunter, Christina Magill

Risk Frontiers—Natural Hazards Research Centre, Macquarie University, Sydney, NSW 2109, Australia

Abstract

Catastrophe loss estimation for natural hazards combines both hazard and exposure data. While hazard attributes such as intensity distributions are usually represented at a spatially explicit raster (or pixel) level, exposure data such as population, dwellings and insurance portfolios are usually only available at spatially lumped census tracts. In current loss estimation studies, this spatial incompatibility is often inadequately addressed and a uniform distribution of exposure data within an areal unit assumed. As a result, loss estimation models overlook a great deal of spatial disparity. This paper defines *occupied residential area* as the area at risk and uses a dasymetric mapping approach to obtain this from areal census tracts.

Using Sydney (Australia) as an example, residential areas at risk were produced through street buffers. The effect of incorporating area at risk in loss estimation models in this manner was then tested for two hazards (earthquakes and hailstorms) that impose very different-sized damage footprints. Total numbers of separate houses (as exposure data) were represented at two hierarchically nested areal unit forms (postcode and census collection district—CCD) and their corresponding *residential area* forms. The spatial distribution of calculated losses for these different forms was then pairwise compared. For earthquakes, estimated losses were insensitive to the manner of delineating area at risk and the use of finer-resolution exposure data. This follows because the affected zone is much larger than the areal unit to which exposure data are attached. Hailstorms, on the other hand, have relatively small affected zones, and loss estimates at a coarse postcode level were considerably different to those at postcode-based residential, CCD and CCD-based residential levels. Differences between CCD and CCD-based residential area levels were relatively small in both cases because the distribution of fine CCD units closely reflects the underlying residential

* Corresponding author. Tel.: +61-2-9850-9473; fax: +61-2-9850-9394.
E-mail address: kchen@els.mq.edu.au (K. Chen).

0143-6228/$ - see front matter
doi:10.1016/j.apgeog.2004.03.005

areas. Our empirical findings suggest that improved delineation of the area at risk and employing exposure data based on finer areal units are important for improving loss estimation from catastrophic events, particularly those that affect only a small proportion of the area under consideration. The results also have significance for other multidisciplinary studies concerned with the integration of spatially explicit environmental data and spatially lumped socioeconomic data.

Keywords: Natural hazards; Catastrophe loss estimation; Catastrophe modelling; Dasymetric mapping; GIS; Area at risk; Sydney

Introduction

The past few decades have seen dramatic worldwide increases in economic losses from catastrophic natural hazards. For example, the 1992 Hurricane Andrew (Florida) and the 1994 Northridge (California) earthquakes caused insurance losses of US$ 15.5 billion and 12.5 billion, respectively (NRC, 1999); the 1995 Great Hanshin-Awaji earthquake in Kobe, Japan, resulted in over 6400 casualties and economic losses to the city of US$ 60 billion; and the 1998 Hurricane Mitch damaged up to 70% of the infrastructure in Honduras and Nicaragua and severely devastated the economies of all Central American countries (ISDR, 2003). In the wake of these disasters, catastrophe loss modelling that simulates potential losses (e.g. economic, insured, properties and human fatalities) from major hazard events has advanced over the past decade, and now serve as an instrument for effective risk management by both the insurance industry and government agencies (e.g. Bendimerad, 2001; Clark, 2002; FEMA, 2002; Kunreuther & Roth, 1998; Leigh & Kuhnel, 2001; Walker, 1995).

Catastrophe loss modelling for natural hazards comprises four major steps (Fig. 1):

- Hazard analysis: quantifies the physical characteristics of a hazard, including probability of occurrence, magnitude, intensity, location, influence of geological or meteorological factors.
- Exposure analysis: identifies and maps underlying elements at risk or exposures, including the built environment and socioeconomic factors such as population and economic activity.
- Vulnerability analysis: assesses the degree of susceptibility to which elements at risk are exposed to the hazard. A community with strong capacities would decrease susceptibility. A common form of vulnerability analysis uses historical damage records to prescribe relationships between damage to dwellings and hazard intensity. Different building construction classes and occupancy types will have distinct vulnerability curves. Vulnerability curves can also be established for socioeconomic exposures, such as population age groups, although for the time being such relationships are not well developed.
- Risk analysis: synthesises the above three components and determines the resulting losses as a function of return period or as an exceedance probability.

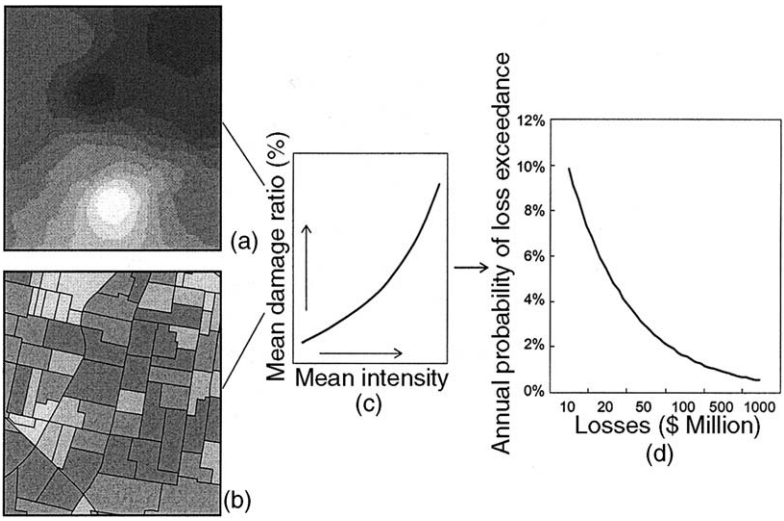


Fig. 1. Four major components of catastrophe loss modelling: (a) hazard analysis, e.g. intensity distribution; (b) Exposure analysis, e.g. exposure data distribution; (c) Vulnerability analysis showing a vulnerability curve; (d) Risk analysis showing a probabilistic loss exceedance curve.

In this loss modelling framework, risk is seen a function of hazard, exposure and exposure-related vulnerability; that is, risk $= f$ (hazard, exposure, vulnerability). Analysis on the hazard versus exposure dichotomy lays the foundation for integrative vulnerability and risk analyses. Since hazard and exposure data are associated with spatial and temporal non-stationarities or variations, catastrophe loss estimation models are often developed within a geospatial analysis environment.

Loss estimation models can be improved by incorporating new findings about hazard mechanisms and occurrence, improved classification of elements at risk and/or locally oriented vulnerability curves. In this paper, our interest is the spatial mismatch between hazard and exposure data: the hazard component (e.g. ground shaking intensity in the case of earthquakes) is commonly modelled at a spatially explicit raster level, while the elements at risk are often only available at spatially lumped and coarse areal unit levels, such as postcodes and census tracts (FEMA, 2002; Leigh & Kuhnel, 2001). (Exposure data at individual address level are not usually provided for obvious confidentiality reasons.) The latter areal units were created for administration purposes and their boundaries lack any physical meaning for catastrophe loss estimation. This spatial mismatch of data can be a source of error since the elements at risk are often assumed to be uniformly distributed across a coarse areal unit and any spatial disparity within the unit ignored. As a consequence, the entire areal unit is improperly treated as being at risk. Since the same set of exposure data may exist at hierarchical areal unit levels having different spatial resolutions, corresponding loss estimates using those data could differ markedly. This paper explores the likely magnitude of such differences.

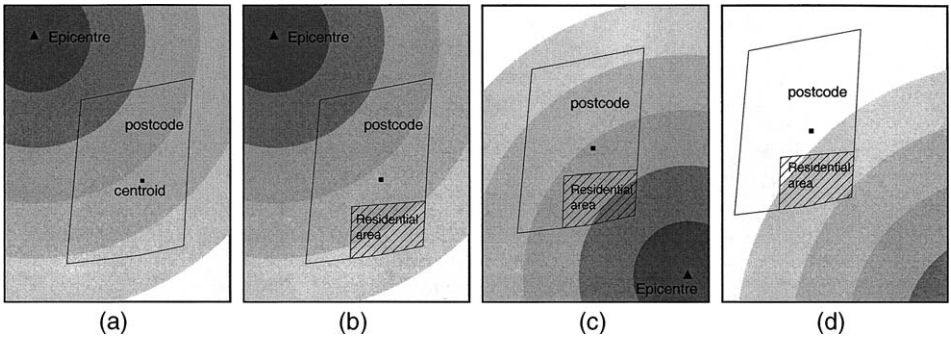


Fig. 2. A schematic representation of hazard intensity distributions, a postcode and identified residential areas.

Fig. 2a shows the spatial distribution of hazard intensity within a postcode. Traditional practice attributes the intensity within the postcode to the centroid. This oversimplification has two problems. First, because the centroid is determined by postcode boundaries that can be arbitrary, the centroid for a polygon of irregular shape may not necessarily even be located within the postcode! In this case, using the hazard intensity at the centroid could be quite misleading. Second, by assuming that the entire area of the postcode is at risk, any spatial disparity of the elements at risk within this areal unit is ignored. In the case of residential dwellings, if the *occupied* residential areas could be more accurately identified, a better estimation of risks is guaranteed by simply disregarding non-occupied areas within the postcode. In Fig. 2b, since the residential area is further away from the epicentre than the centroid, a lower intensity should be assigned to the elements at risk or losses will be overestimated. By comparison, losses will be underestimated for the situations shown in Fig. 2c and d.

The primary objectives of this study are: (1) to draw attention to the importance of appropriately defining the area at risk for catastrophe loss estimation; (2) to define *occupied residential areas* as the area at risk with dasymetric mapping; and (3) to assess the effect on estimated losses of defining the area at risk and comparing these with the exposure data re-modelled and represented in four forms—postcode, postcode-based residential area, census collection district (CCD) and CCD-based residential area. The Sydney region of New South Wales, Australia, is used for two case studies. Both earthquakes and hailstorms will be considered.

Defining the area at risk with a dasymetric mapping approach

Choropleth mapping is the most popular technique for representing area-based exposure data. An important assumption is that value at risk is spread uniformly across each census tract. This is often far from true. For instance, the distribution of dwellings may be unevenly scattered and/or concentrated in only a small part of the zone, while the rest of the zone is empty. What is needed is finer-grained aver-

aging; that is, to dis-aggregate area-based exposure data at a finer spatial resolution. One approach for this dis-aggregation is dasymetric mapping (e.g. Dent, 1996; Langford & Unwin, 1994; Longley, Goodchild, Maguire, & Rhind, 2001), an approach that transforms data from arbitrary areal units to physical settlement areas. An example is illustrated in Fig. 3 where a hypothetical areal map and associated residential areas are combined to produce a new map of the residential areas. After a GIS intersection operation, the final representation of the attribute is divorced from the original boundaries of the census zones and properly allocated to the residential area.

Since exposure data are often related to residential areas, denoting residential areas as the area at risk should represent a major improvement over conventional practice. By so doing, spatial contextures important to risk quantification are conceptually embodied. In theory, residential areas could be further differentiated and classified into land-covers and land-uses, dwelling densities, and even individual dwellings if this were warranted for any particular application (Chen, Blong, & Jacobson, 2003).

This paper presents an example of identifying occupied residential areas through the use of street buffers. Occupied residential areas are assumed to be physically linked by street networks. MapInfo-compatible StreetWorks™ (MapInfo Australia, 2001) is a detailed street database, covering all capital and major regional cities in Australia. In urban areas, each street consists of a series of line segments, with the start and end street addresses on either side of the segment indicated. To derive the residential area, a distance of 100 m was used as a buffer zone on either side of the street segment. While this distance is subjective, a conservative view was taken to ensure the inclusion of sufficient occupied residential areas. If the separation of two parallel neighbouring streets is less than 200 m, their buffered zone indicates the entire area between the two streets as residential.

Spatial layers of local and national parks, lakes and rivers were also used to refine the boundaries of residential areas. The final residential area was then con-

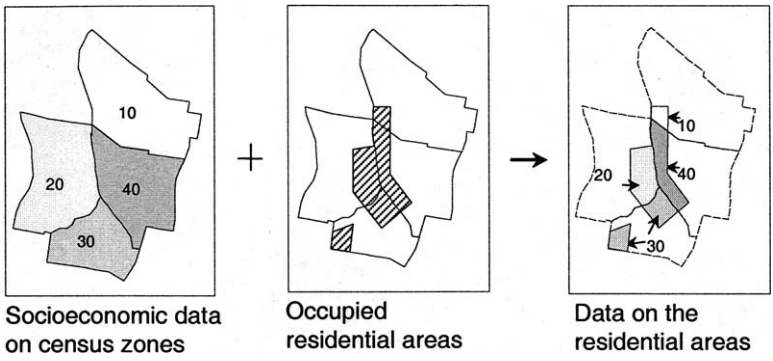


Fig. 3. Transformation of socioeconomic data from census tracts to residential areas with dasymetric mapping. The numbers in the figure (i.e. 10, 20, 30 and 40) refer to the numbers of dwellings in each of the four zones.

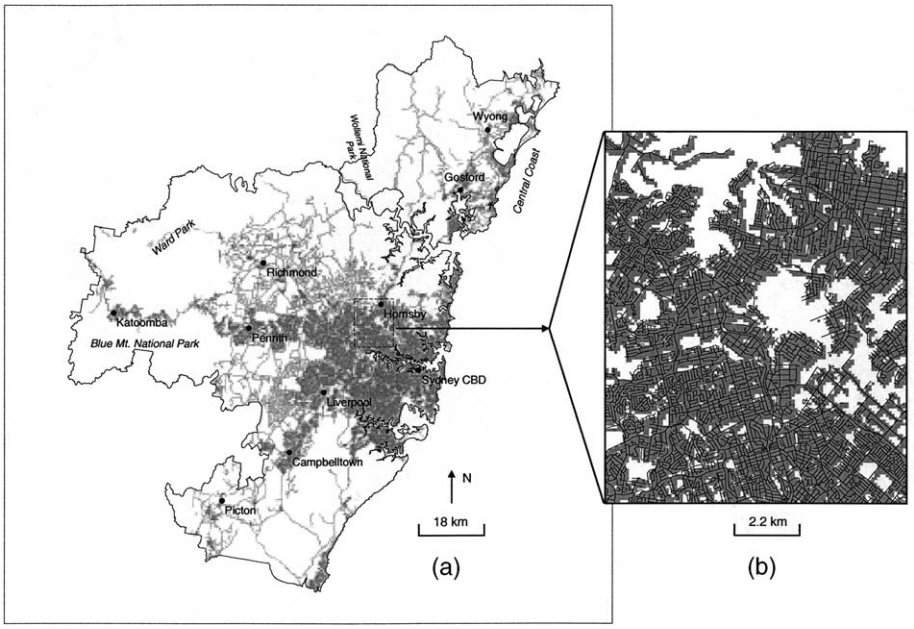


Fig. 4. (a) Street buffer-based residential areas for the Sydney region; (b) A zoom-in of the dashed rectangle area in (a). Residential areas are shown in gray, and local streets are superimposed.

verted into a raster format, with a spatial resolution of 100 m. Fig. 4a shows the derived occupied residential areas for the Sydney region. More than 150 000 street segments from StreetWorks™ were used. A zoom-in of a small area south of the suburb of Hornsby is shown in Fig. 4b, with local street segments superimposed.

The above approach of street buffering to derive approximate residential areas can be easily implemented in GIS. While the approach is not immune to error, and the buffer distance of 100 m may overestimate the true residential areas in some areas, the transformation of areal exposure data from arbitrary areal units to the residential area is a major improvement in representing the area at risk. In Fig. 4a, the Sydney region covers 8879 km^2, whereas the derived residential area is only 1956 km^2—22.0% of the study area. Loss estimation should focus on this occupied sub-set.

Case study 1: earthquake loss estimation

Hazard simulation and vulnerability curve

The Sydney region is susceptible to earthquakes, and the 1989 Newcastle Earthquake (epicentre 120 km north of Sydney) produced insurance losses of around AUS$ 1 billion (Blong, 1997). Our main concern here is to demonstrate how loss estimates vary with exposure data represented in different forms, given identical hazard simulation characteristics. A modified Mercalli intensity (MMI) scale

(Dowrick, 1996) was employed to indicate the intensity of ground shaking resulting from an earthquake of local magnitude 6.0 (ML = 6.0) (cf. ML = 5.6 for the 1989 Newcastle Earthquake) and with a focal depth of 15 km ($D = 15$. cf. $D = 11$ for the 1989 Newcastle Earthquake). Complications due to local soil amplification and geology are ignored. The attenuation function proposed by Gaull, Michael-Leiba, and Rynn (1990) for southeastern Australian continent was adopted.

$$\mathrm{MMI} = 1.5\mathrm{ML} - 3.9\log_{10}((D^2 + R^2)^{0.5}) + 3.9$$

where R is the horizontal distance from the epicentre. The MMI distribution is mapped in Fig. 5a. It was assumed that scenario earthquakes can occur at any location within the study area with the same symmetric seismic attenuation function. Deterministic hazard simulations were implemented in a raster GIS environment. To reduce computing time, a total of 22 400 scenario earthquakes were placed at a 140 km × 160 km grid extending well beyond the study area. If the epicentre is located beneath the built-up area, the risk (estimated losses) will be very large; on the other hand, if the epicentre is located far from residential areas, the risk will be minimised.

The likelihood of buildings being damaged is related to factors such as type, materials, age and maintenance. In this case study, a hypothetical exponentially shaped vulnerability curve (Cochrane & Schaad, 1992) linking cost of damage to ground shaking intensity in MMI units was used (Fig. 5b). Losses are expressed as the percentage of total sum insured. Vulnerability curves are usually established through analysis of insurance losses from historical hazard events or on the basis of expert engineering experience and judgment. By combining the intensity distribution in Fig. 5a with the vulnerability curve, the distribution of potential losses surrounding the epicentre is mapped in Fig. 5c; the shape is highly peaked and has a radius of impact (MMI $\geq$ 5) of around 100 km.

Separate houses as elements at risk

In this case study, separate houses were chosen as the insured elements (assumed portfolio data). Other residential building-types such as separate detached houses, flats, industrial and commercial buildings were not considered. The numbers of separate houses are available from CDATA 2001 (Australian Bureau of Statistics, 2002) at two areal unit levels: postcodes and CCDs. CCDs are the smallest areal unit used for reporting census data. The boundaries of postcodes and CCDs are spatially compatible as postcodes are derived from aggregating CCDs. There are a total of 255 postcodes (Fig. 6a) and 6674 CCDs (Fig. 6c) within the study area. The characteristics of two areal units and their associated portfolio data are shown in Table 1. The total number of separate houses is 919 898, and an average sum insured of AUS$ 200 000 was assumed. Table 1 indicates slight differences in the total number of separate houses at the two areal unit levels; our results are insensitive to this small difference of 40 houses or 0.004% of the total number.

Fig. 6 shows the portfolio data in the following four forms: (1) uniform over each postcode (Fig. 6a); (2) postcode-based residential area (Fig. 6b); (3) uniform

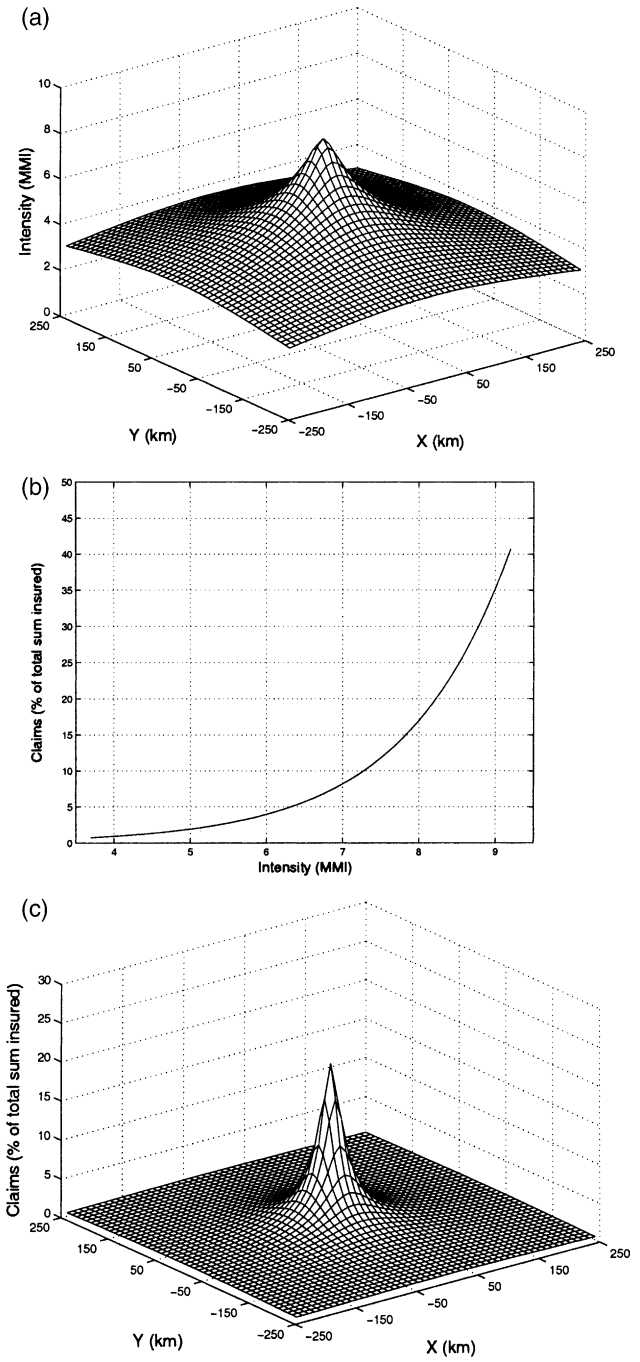


Fig. 5. (a) Prescribed intensity distribution about the earthquake epicentre at ($x = 0$, $y = 0$); (b) A hypothetical vulnerability curve linking insured losses as a percentage of total sum insured; (c) Combination of (a) and (b) showing the distribution of potential losses with distance from the epicentre.

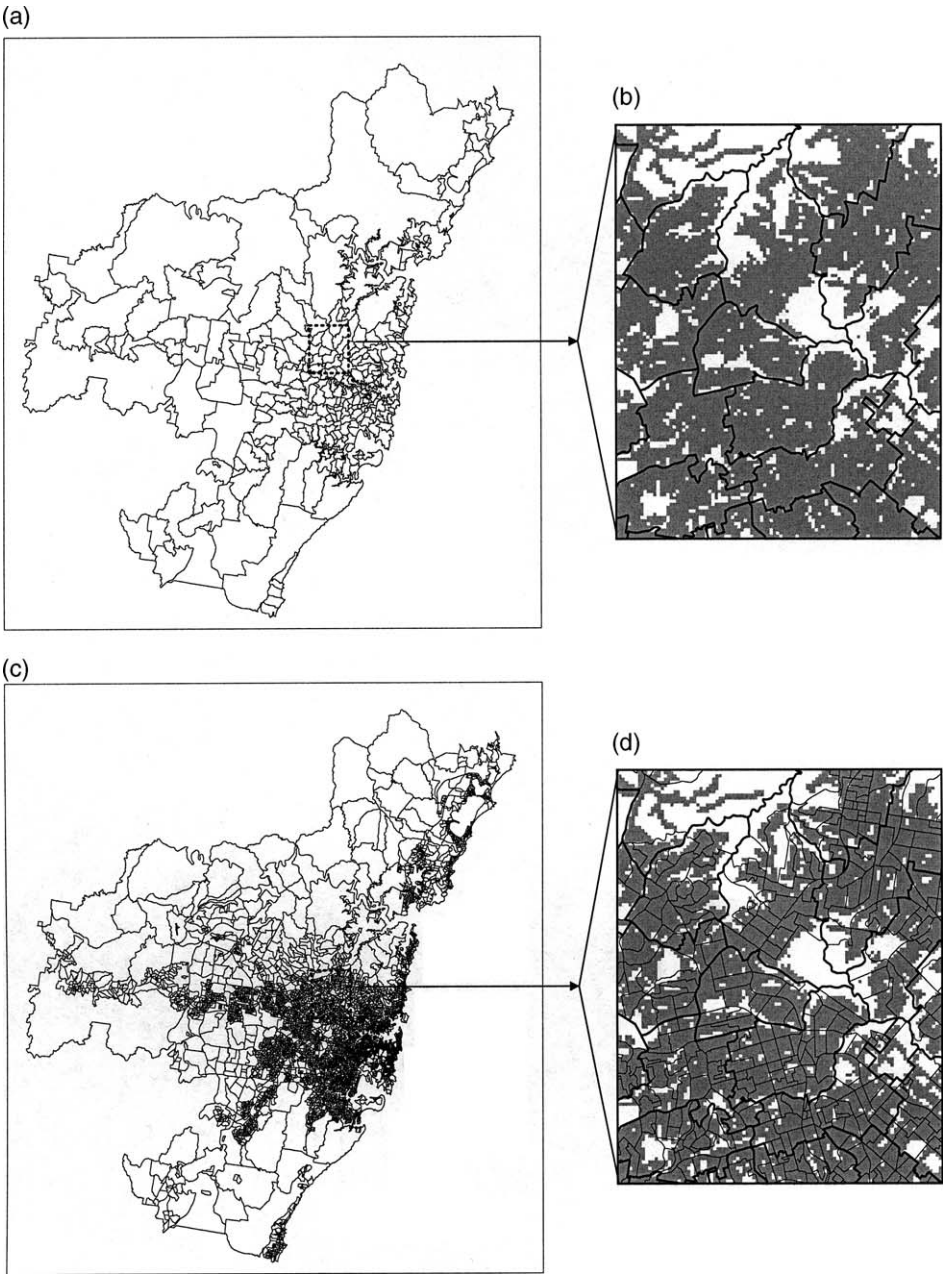


Fig. 6. At-risk exposure data (separate houses) are expressed as four forms. (a) 255 postcodes for the Sydney region; (b) A zoom-in of the dashed rectangle area in (a). Residential areas are superimposed with postcode boundaries, and the residential area within each postcode indicates the number of houses in the postcode; (c) 6674 CCDs; (d) A zoom-in of the dashed rectangle area in (c). Residential areas are superimposed with CCD boundaries, and the residential area within each CCD indicates the number of houses in the CCD. The boundaries of postcodes and CCDs are hierarchically nested.

Table 1
Statistical characteristics of areal units and portfolio data (i.e. separate houses)

Areal units	Number	Area (km^2)					Number of separate houses				
		Total	Min.	Max.	Mean	Std. dev.	Total	Min.	Max.	Mean	Std. dev.
Postcode	255	8879.0	0.581	758.3	34.8	96.0	919 898	3	19 118	3607	3364
CCD	6674	8879.0	0.002	605.6	1.3	11.5	919 858	0	663	138	93

over each CCD (Fig. 6c); and (4) CCD-based residential area (Fig. 6d). Each form was converted to a raster format with a spatial resolution of 100 m. Among these four forms, it is clear that the postcode form represents the coarsest resolution while the CCD-based occupied residential area form represents the ideal and most detailed portfolio data format for loss estimation. Because some CCDs are very small and cannot be represented at a spatial resolution of 100 m, a total of 23 and 63 houses could not be included in the CCD form and CCD-based residential area form, respectively.

Catastrophe losses were calculated for each scenario earthquake over all postcode and CCD units and the values assigned to the earthquake epicentre. To examine the spatial difference in the estimated losses, the following six comparisons were conducted: (1) at a postcode level and a postcode-based residential area level; (2) at a CCD level and a CCD-based residential area level; (3) at a postcode level and a CCD level; (4) at a postcode-based residential area level and a CCD-based residential area level; (5) at a postcode level and a CCD-based residential area level; and (6) at a postcode-based residential area level and a CCD level. For each pairwise comparison, the loss estimation difference (LED) for each earthquake location was calculated using the following equation:

$$\mathrm{LED}_i = 100 \times (L_{1i} - L_{2i})/L_{1i}$$

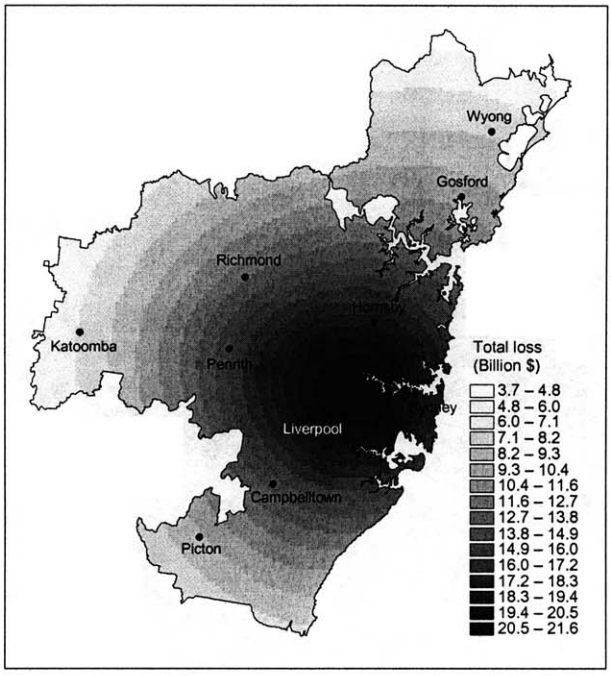
where L_{1i} and L_{2i} are losses calculated at location (epicentre) i ($i = 1, 2, \ldots, 22\,400$) using the first and the second representation forms of the exposure data, respectively. The loss estimation difference (%) is relative to the loss estimated in the first form, and negative differences imply that the loss estimate in the second representation was higher. Negative difference is referred to as underestimation.

Loss estimation differences

Fig. 7a shows the distribution of estimated losses using the portfolio data at the postcode level. Estimated losses per scenario event ranged from AUS$ 3.7 billion to 21.6 billion. The largest losses are naturally for the earthquakes that occur

Fig. 7. (a) Distribution of losses calculated at the postcode level for scenario earthquakes occurring at different locations; (b) Distribution of grouped loss estimation differences between a postcode level and a postcode-based residential area level.

(a)


Wyong
Gosford
Richmond
Katoomba
Penrith
Liverpool
Campbelltown
Picton
Total loss
(Billion $)
3.7 – 4.8
4.8 – 6.0
6.0 – 7.1
7.1 – 8.2
8.2 – 9.3
9.3 – 10.4
10.4 – 11.6
11.6 – 12.7
12.7 – 13.8
13.8 – 14.9
14.9 – 16.0
16.0 – 17.2
17.2 – 18.3
18.3 – 19.4
19.4 – 20.5
20.5 – 21.6

(b)

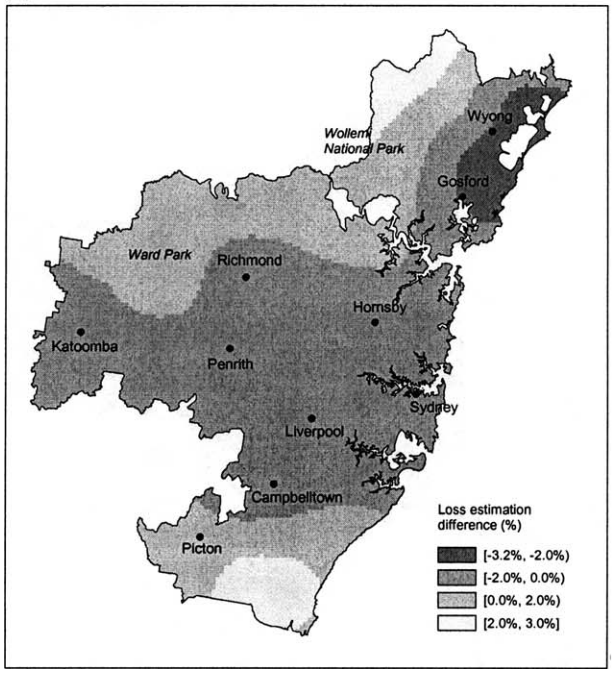

Wyong
Wollemi
National Park
Gosford
Ward Park
Richmond
Hornsby
Katoomba
Penrith
Sydney
Liverpool
Campbelltown
Picton
Loss estimation
difference (%)
[-3.2%, -2.0%)
[-2.0%, 0.0%)
[0.0%, 2.0%)
[2.0%, 3.0%]

Table 2
Difference ranges of six comparisons in earthquake loss estimation

Comparisons	Difference ranges
(a) Postcode level and postcode-based residential area level	[−3.2%, 3.0%]
(b) CCD level and CCD-based residential area level	[−0.1%, 0.1%]
(c) Postcode level and CCD level	[−4.6%, 4.2%]
(d) Postcode-based residential area level and CCD-based residential area level	[−1.4%, 1.7%]
(e) Postcode level and CCD-based residential areal level	[−4.6%, 4.3%]
(f) Postcode-based residential area level and CCD level	[−1.4%, 1.6%]

beneath the densely populated parts of city. Fig. 7b shows the spatial distribution of the loss estimation differences between a postcode level and a postcode-based residential area level. Differences are small and range from −3.2% to 3.0%. The difference ranges of the other five comparisons in Table 2 also indicate only minor underestimations and overestimations.

Underestimations and overestimations are possible depending on the location of the earthquake in relation to residential areas. If the earthquake epicentre lies within or near a densely residential area, then the finer representation will produce higher losses than if these same number of houses were spread uniformly over a coarse postcode (see Fig. 2). Overestimations are also possible and occur in non-residential areas (Fig. 7b). In fact, for non-residential areas, the numbers of dwellings at risk and associated contributing losses should be zero.

Overall, using finer-resolution exposure data in earthquake loss estimation does not appear to dramatically affect the result. We attribute this to the very large footprint of the earthquake. For an earthquake with a radius of impact of 100 km, the corresponding circular areal extent of impact is up to 31 415 km^2. In this case study, scenario earthquakes at any location would always affect the overwhelming majority of the study area, and the improvement of representing exposure data at detailed local levels in producing aggregated and global loss estimates is compromised. The results may be different if the effect of amplification of ground shaking due to local soils and geological changes were to be taken into account.

The above result is application-specific. For hazards such as hailstorms and tornadoes that are associated with smaller, more narrowly defined footprints, differences in loss estimates may be more sensitive to the representation of residential areas, particularly at the urban fringe. We explore this possibility in the next section using a hailstorm case study.

Case study 2: hailstorm loss estimation

Hazard loss potential distribution

Hailstorms in the Sydney urban areas often cause substantial damages. The April 1999 Sydney hailstorm, for example, caused an estimated total insured loss of AUS$ 1.6 billion, making it the most costly natural disaster in Australia (Leigh

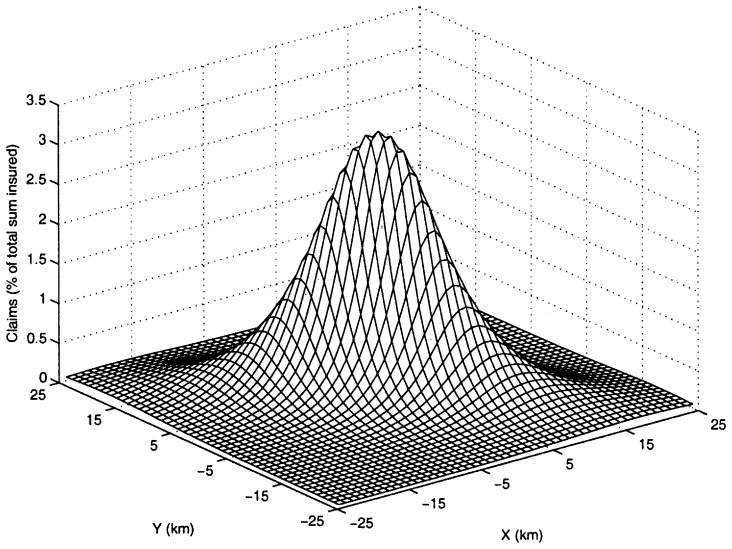


Fig. 8. Distribution of loss potential from a scenario hailstorm, similar to that in Fig. 5c.

& Kuhnel, 2001). Hailstorm loss modelling requires a comprehensive set of physical data that characterise the spatial and temporal distribution of hailstorm attributes such as direction, path, shape and areal extent of the storm, hailstone size, occurrence season, date and time. These variables are usually generated in a stochastic matter with reference to regional hailstorm history and other climatological information (e.g. El Niño—southern oscillation cycle) (Leigh & Kuhnel, 2001). However, to address the key theme of this paper, the detailed simulation of hailstorms was not used. Scenario hailstorms with a hypothetical loss potential shown in Fig. 8 were placed on the 140 km × 160 km grid deterministically in a similar manner as was done with the earthquake study. The distribution of loss was assumed to follow a lognormal function (Vose, 2000), with a SW–NE oriented elliptical footprint (semi-major axis 25 km and semi-minor axis 10 km) and maximum loss percentage for buildings (claim % of the total sum insured) at a discrete 1 km × 1 km grid level of 3.5%. These characteristics are broadly consistent with those of severe Sydney hailstorms such as the March 1990 and April 1999 events (Andrews & Blong, 1997; Leigh & Kuhnel, 2001).

Hailstorm loss estimates were calculated using the same set of exposure data (separate houses) at the four representation forms as with earthquakes, and then cross compared. By comparison with earthquake example discussed previously, the differences are striking (Table 3). Each pairwise comparison is discussed below, staring with the comparison with loss estimates at the coarsest level and the most detailed level.

Table 3
Difference ranges of six comparisons in hail loss estimation

Comparisons	Difference ranges
(a) Postcode level and postcode-based residential area level	[−159.5%, 99.6%]
(b) CCD level and CCD-based residential area level	[−83.0%, 96.5%]
(c) Postcode level and CCD level	[−152.9%, 98.3%]
(d) Postcode-based residential area level and CCD-based residential area level	[−41.7%, 93.0%]
(e) Postcode level and CCD-based residential areal level	[−161.3%, 99.5%]
(f) Postcode-based residential area level and CCD level	[−171.1%, 95.3%]

At a postcode level and a CCD-based residential area level

Fig. 9a shows the spatial distribution of estimated losses using the portfolio data at the CCD-based residential area level—the finest resolution employed in this case study. Depending on the location of the hailstorm, losses range from below AUS$ 1 million to 600 million. Fig. 9b shows the distribution of loss estimation differences on a pixel basis between a postcode level and a CCD-based residential area level; differences range between −161.3% and 99.5%. Fig. 9c maps the same information using the six difference categories defined in Table 4. For the areas of each group, mean absolute losses estimated at a CCD-based residential area level and at a postcode level, and their corresponding differences are shown in Fig. 10. As the percentage of underestimation and overestimation increases, the difference of mean absolute losses shows an escalating trend. For example, the large underestimation group (<−50.0%) has an absolute mean losses difference of AUS$ 29.7 million.

Fig. 9c shows the areal proportions of the most extreme underestimation and overestimation groups to be substantial, occupying almost one-third of the study area (31.8%, see line (e) in Fig. 11). The large underestimation groups (estimates at

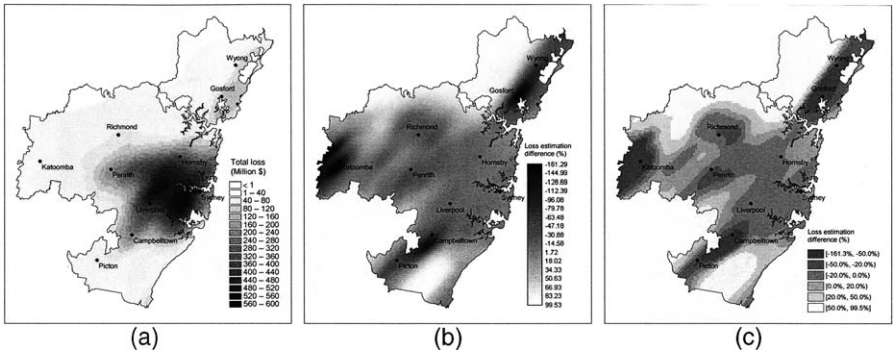


Fig. 9. (a) Distribution of losses calculated at the CCD-based residential area level for scenario hailstorms occurring at different locations; (b) Distribution of continuous loss estimation differences between a postcode level and a CCD-based residential area level; (c) Distribution of grouped loss estimation differences between a postcode level and a CCD-based residential area level.

Table 4
Six groups of loss estimation differences

	Groups	Difference ranges
Underestimation	Large	$<-50.0\%$
	Medium	$[-50.0\%, -20.0\%)$
	Small	$[-20.0\%, 0\%)$
Overestimation	Small	$[0\%, 20.0\%)$
	Medium	$[20.0\%, 50.0\%)$
	Large	$\geq 50.0\%$

a postcode level appreciably less than those at a CCD-based residential area level) are concentrated around Picton, Campbelltown, Katoomba, Gosford and Wyong. These regions are typically away from the city centre and located at the urban fringes between residential and non-residential areas (e.g. national parks). While loss modelling should concentrate on the residential areas only, when portfolio data at the coarse postcode level are used, some proportions of houses within postcodes are incorrectly assigned to non-residential areas.

For inner city suburbs dominated by continuous residential areas, loss estimation differences are much smaller because the chance of mis-allocating proportions of the portfolio to non-residential areas is much smaller than at the urban fringes. This result suggests that when scenario hailstorms are centred over extensive and continuous residential areas such as inner city suburbs, loss estimation

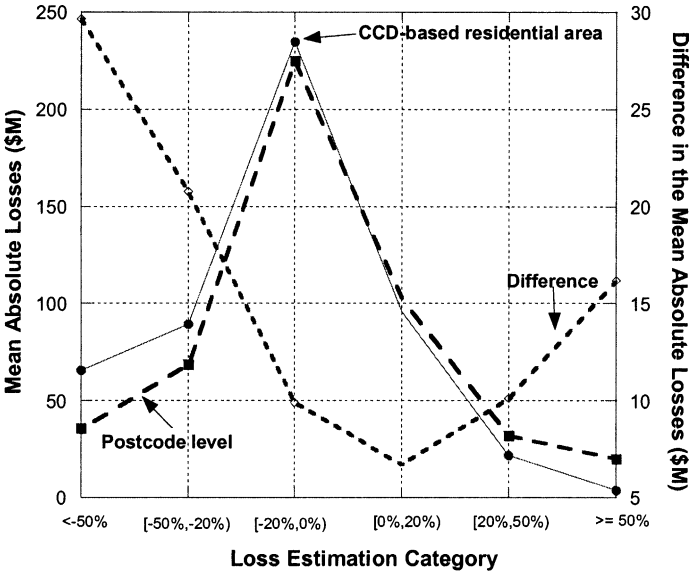


Fig. 10. Absolute mean losses for the areas of six loss estimation difference groups are summarised at the CCD-based residential area level and at the postcode level. The difference of their absolute mean losses is indicated at the right-hand y-axis.

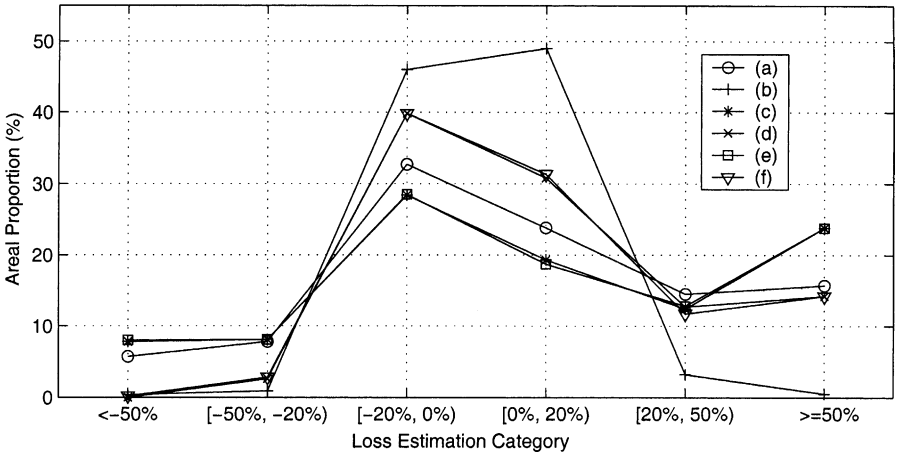


Fig. 11. Areal proportions relating to six loss estimation difference groups for scenario hailstorms. Line (a): comparison between a postcode level and a postcode-based residential area level; Line (b): comparison between a CCD level and a CCD-based residential area level; Line (c): comparison between a postcode level and a CCD level; Line (d): comparison between a postcode-based residential area level and a CCD-based residential area level; Line (e): comparison between a postcode level and a CCD-based residential area level; Line (f): comparison between a postcode-based residential area level and a CCD level.

should be largely indifferent to whether the portfolio data are employed at a postcode level or a CCD-based residential area level.

On the other hand, significant overestimations occur for storms centred within national parks (e.g. the Wollemi National Park in the north, Ward Park in the west) and non-residential areas (e.g. east of Picton). Homogeneity assumed at the coarse postcode level allocates portfolio data to these non-residential areas, incorrectly attributing larger loss estimates for these areas. When portfolio data at the CCD-based residential area level are used, loss estimation disregards the coarse size and the shape of postcodes, and reflects the true location of the dwellings more realistically.

At a postcode level and a postcode-based residential area level

In Fig. 12a, the overall spatial distribution of underestimation and overestimation groups is similar to that illustrated above between a postcode level and a CCD-based residential area level. Similar reasoning can be applied to explain this. When postcode-based portfolio data are used in loss estimation, some separate houses are wrongly assigned to non-residential areas. The result is smaller losses at continuous residential areas and larger losses for non-residential areas and national parks.

At a CCD level and a CCD-based residential area level

Fig. 12b shows the spatial distribution of grouped loss estimation differences between a CCD level and a CCD-based residential area level. The range of loss

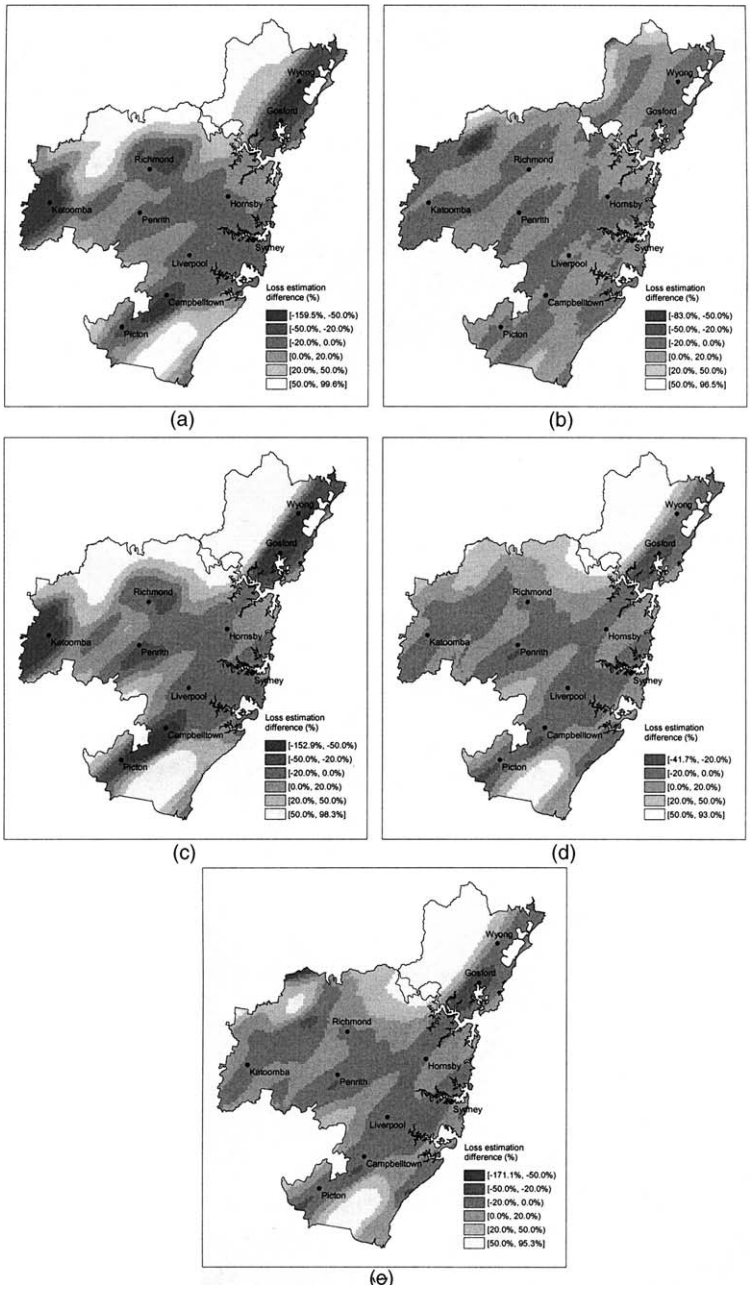


Fig. 12. Spatial distribution of six loss estimation difference groups for scenario hailstorms. (a) Comparison between a postcode level and a postcode-based residential area level; (b) Comparison between a CCD level and a CCD-based residential area level; (c) Comparison between a postcode level and a CCD level; (d) Comparison between a postcode-based residential area level and a CCD-based residential area level; (e) Comparison between a postcode-based residential area level and a CCD level.

estimation differences is relatively small; the mean percentages of loss estimation difference for small underestimation and overestimation groups are only −1.4% and 3.0%, respectively. A total of 95.0% of the study area belongs to these small difference groups (line (b) in Fig. 11). This is attributed to the fact that the fine resolution of CCDs reflects the general distribution of residential areas quite well (compare Figs. 4a and 6c). Recall that the average size of CCDs for the entire study area is 1.3 km^2, and the size of CCDs for the concentrated residential areas is even smaller. For distant national park areas with relatively coarse CCD units, loss estimation differences could still be large, but the area affected is small—only 0.9% of the study area is associated with the large underestimation and overestimation groups (line (b) in Fig. 11). Therefore, for simulated hailstorms centred near residential areas (e.g. inner city suburbs), there appears to be little benefit in using CCD-based residential data over CCD level data.

At a postcode level and a CCD level

Fig. 12c shows the spatial distribution of grouped loss estimation differences calculated from postcode level and CCD level data. The patterns are extremely similar to those between a postcode level and a CCD-based residential area level in Fig. 9c. The areal proportions of each difference group at these two pairwise comparison levels are also closely matched (line (c) in Fig. 11). This is not surprising given the small difference of loss estimation between a CCD level and a CCD-based residential level shown in Fig. 12b. From here, we can conjecture that the finer size of CCD also plays an important role in improving loss estimation.

At a postcode-based residential area level and a CCD-based residential area level

Fig. 12d shows the distribution of grouped loss estimation differences between a postcode-based residential area level and a CCD-based residential area level. A total of 70.6% of the study area is subject to small underestimation and overestimation groups. There are no areas for the large underestimation group (line (d) in Fig. 11). Providing both the postcode- and CCD-based portfolio data are always confined to the residential area, medium and large loss estimation differences are directly ascribed to the difference of spatial resolutions or areal sizes with which the original portfolio data are associated.

At a postcode-based residential area level and a CCD level

The general distribution of grouped loss estimation differences between a postcode-based residential area level and a CCD level shown in Fig. 12e is similar to that in Fig. 12d. The areal proportion of each difference group at this pairwise comparison level is also very similar to the previous case (line (f) in Fig. 11 is closely aligned with line (d) in Fig. 11). This is understandable given the pattern demonstrated above between a postcode-based residential area level and a CCD-based residential area level, and the small loss difference between a CCD level and a CCD-based residential level. Nevertheless, this comparison is important as moving from postcode level exposure data to postcode-based residential area level or

CCD level data are the two most likely options for improving many existing loss models. From a practical perspective, residential areas can be readily derived, but portfolio data from the insurance industry are not often available at a CCD level. If the difference in this comparison is minor, it implies that incorporating the residential area at risk at the postcode level is effective and sufficient. However, the major loss estimation differences shown here again highlight the importance of refining exposure data at a finer areal unit.

Discussion

This study has examined one important aspect of catastrophe loss estimation, the delineation of the area at risk and its effect in loss estimation. GIS-based catastrophe loss estimation models are increasingly employed for tasks such as estimating probable maximum losses and setting risk premiums. Current loss estimation practice largely uses portfolio data at postcode or census tract levels; it is logical to anticipate that a better loss estimation result can be achieved through a more accurate definition of the area at risk. In each case studies, the true residential area represented only 22.0% of the entire study area. Natural hazards often have the same areas at risk (i.e. occupied residential areas) and once these areas have been delineated, it is convenient to use them for estimating losses from a range of natural hazards, such as thunderstorms, bushfires and volcanic eruptions. This idea could be extended to assess the potential losses associated with other types of portfolio data. If, for example, an industrial portfolio data set is used, the delineation of industrial land use is likely to be helpful.

With consistent and simplified hazard simulation conditions for scenario earthquakes and hailstorms, loss estimation differences using portfolio data at different representation forms have been compared. Due to the large footprint of earthquakes, loss estimation at a regional scale is relatively insensitive to the manner in which the area at risk is defined. However, some differences of loss estimation do exist, both spatially and quantitatively. If more hazard factors (e.g. local soil conditions in the case of earthquakes) were to be incorporated, the outcome could be quite different.

Loss estimations are commonly conducted at hierarchical areal unit levels, and the results are not always invariant. Given the large loss estimation discrepancies calculated at different census-based areal unit levels in the hailstorm case study (e.g. Fig. 12c), losses for the coarse areal units (e.g. postcode) would be better derived from loss estimates at finer areal unit levels (e.g. CCD). Such a bottom-up scaling approach can ensure the estimated losses are not heavily affected by coarse-grained averaging. In the US, HAZUS® (FEMA, 2002) uses census tract as a basic areal analysis unit. Because of its relatively coarse size (for example, the State of California has 5858 census tracts with an average size of 70.0 km^2), loss estimates at a census tract level could be derived from and compared with the losses estimated at still finer areal unit levels (e.g. census block groups), if related exposure information is indeed also available at this level.

It is recognised that the delineation of occupied residential areas is subject to uncertainty. In this study, a conservative view was taken to derive approximate residential areas as the area at risk. Both street data and exposure data collected in 2001 provided a consistent temporal representation. For loss estimation at a regional level, commonly available and finer-resolution spatial data, such as the topologically integrated geographic encoding and referencing (TIGER)/Line® data and SPOT-5 imagery (with 2.5 and 5-m spatial resolutions in a panchromatic band, 10-m spatial resolution in multispectral bands, and wide swath), would enable the delineation of residential areas with greater accuracy. Theoretically, as more detailed exposure data are derived and used, more accurate results are possible. Sensitivity analyses allow us to choose the most cost-effective spatial representation level. An in-depth and systematic examination of the scale-dependency of socioeconomic data at different representation areal levels (e.g. Chen et al., 2003; Fisher & Langford, 1996) and its effect in loss estimation modelling using methods such as modern geostatics, remains for future research.

The applications of the approach demonstrated here go well beyond catastrophe modelling. Broadly speaking, this can be important for many multidisciplinary studies investigating "environment–human" interactions, where a successful integration of spatially explicit physical environmental data and spatially lumped socioeconomic data are essential. For example, in the estimation of population at risk from environmental and technological hazards such as industrial chemical releases (Cutter, Holm, & Clark, 1996; Walker, Mooney, & Prattes, 2000), or from the spread of diseases (Lawson et al., 1999), better delineation of areas at risk and effective spatial analysis will increase the objectivity of environmental impact assessments.

Conclusion

As an effort towards spatially integrated risk science, defining the meaningful area at risk and effectively representing the spatial distribution of exposure data is important. Four main findings of this study are summarised as follows:

- To reconcile spatially explicit hazard data and spatially lumped exposure data in a GIS-based loss estimation methodology, the exposure data at arbitrary areal unit levels can be transformed into occupied residential areas using a dasymetric mapping approach. For regional loss estimations, the occupied residential area as the area at risk is both rational and viable.
- The effect of incorporating the area at risk in catastrophe loss estimation is application-specific. For earthquakes with large footprints compared with the areal unit at which the exposure data are represented, the effect of postcode averaging is small.
- For hailstorms with relatively small footprints, differences were significant.
- Loss estimation differences between a CCD level and a CCD-based residential area level were relatively small because the distribution of fine-size CCDs for the study area reflects the underlying residential areas.

References

Andrews, K. E., & Blong, R. J. (1997). March 1990 hailstorm damage in Sydney, Australia. *Natural Hazards*, *16*, 113–125.

Australian Bureau of Statistics. (2002). *CDATA2001 CD-ROMs*. Canberra, Australia.

Bendimerad, F. (2001). Loss estimation: a powerful tool for risk assessment and mitigation. *Soil Dynamics and Earthquake Engineering*, *21*(5), 467–472.

Blong, R. J. (1997). A geography of natural perils. *Australian Geographer*, *28*(1), 7–27.

Chen, K., Blong, R., & Jacobson, C. (2003). Towards an integrated approach to natural hazards risk assessment: with reference to bushfires using GIS. *Environmental Management*, *31*(4), 546–560.

Clark, K. (2002). The use of computer modeling in estimating and managing future catastrophe losses. *The Geneva Papers on Risk and Insurance*, *27*(2), 181–195.

Cochrane, S. W., & Schaad, W. H. (1992). Assessment of earthquake exposure of building. *Proceedings of the 10th World Conference on Earthquake Engineering*, Vol. 1. (pp. 497–502). Rotterdam, Netherlands: Balkema.

Cutter, S. L., Holm, D., & Clark, L. (1996). The role of geographic scale in monitoring environmental justice. *Risk Analysis*, *16*(4), 517–526.

Dent, B. D. (1996). *Cartography: Thematic map design*. (4th ed.). Dubuque, IA: Wm. C. Brown Publishers.

Dowrick, D. J. (1996). The modified Mercalli intensity scale—revisions arising from recent studies of New Zealand earthquakes. *Bulletin of the New Zealand National Society for Earthquake Engineering*, *29*(2), 92–106.

FEMA. (2002). *Technical manual of HAZUS® 99: earthquake loss estimation methodology*. Washington, DC: Federal Emergency Management Agency. Available from: http://www.fema.gov/HAZUS/

Fisher, P. F., & Langford, M. (1996). Modelling sensitivity to accuracy in classified imagery: a study of areal interpolation by dasymetric mapping. *Professional Geographers*, *48*(3), 299–309.

Gaull, B. A., Michael-Leiba, M. O., & Rynn, J. M. W. (1990). Probabilistic earthquake risk maps of Australia. *Australian Journal of Earth Sciences*, *37*, 169–187.

International Strategy for Disaster Reduction (ISDR) Secretariat. (2003). *Living with risk: a global review of disaster reduction initiatives*. Geneva: United Nations ISDR Secretariat. Available from: http://www.unisdr.org/

Kunreuther, H., & Roth, R. J. (1998). *Paying the price: The status and role of insurance against natural disasters in the United States*. Washington, DC: Joseph Henry Press.

Langford, M., & Unwin, D. J. (1994). Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal*, *31*, 21–26.

Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J. F., & Bertollini, R. (1999). *Disease mapping and risk assessment for public health*. Chichester: John Wiley & Sons.

Leigh, R., & Kuhnel, I. (2001). Hailstorm loss modelling and risk assessment in the Sydney region, Australia. *Natural Hazards*, *24*, 171–185.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2001). *Geographic information systems and science*. Chichester: John Wiley & Sons.

MapInfo Australia. (2001). *StreetWorks™ CD-ROMs* (ver. 6.0). Greenwich, NSW, Australia.

National Research Council (NRC). (1999). *The impacts of natural disasters: A framework for loss estimation*. Washington, DC: National Academy Press.

Vose, D. (2000). *Risk analysis: A quantitative guide*. (2nd ed.). Chichester: John Wiley & Sons.

Walker, G. R. (1995). The development of wind and earthquake damage risk models to predict insurance loss. In R. E. Melchers, & M. G. Stewart (Eds.), *Integrated risk assessment: Current practice and new directions* (pp. 1–6). Rotterdam, Netherlands: A. A. Balkema Publishers.

Walker, G., Mooney, J., & Prattes, D. (2000). The people and the hazard: the spatial context of major accident hazard management in Britain. *Applied Geography*, *20*, 119–135.