POPULATION DENSITY MODELLING IN SUPPORT OF SEISMIC RISK ASSESSMENT

Patrizia TENERELLI\textsuperscript{1}, Ufuk HANCILAR\textsuperscript{2}, Javier F. GALLEGO\textsuperscript{3} and Daniele EHRLICH\textsuperscript{4}

ABSTRACT

Demographic data is a fundamental component of an earthquake loss model for the estimation of human exposure and social vulnerability. Population distribution information is needed for the assessment of causalities, determination of shelter needs and proper implementation of evacuation plans in pre- and/or post-disaster phases, i.e. earthquake scenario modelling and rapid emergency response.

This paper describes the techniques that are used to map the population distribution, and to integrate the building damage information with demographic data for casualty estimation. Methods to map population density are described focusing on different downscaling techniques and the contribution of ancillary data.

An application for the Larger Urban Zone (LUZ) of Vienna is illustrated. The case study disaggregates the residential population from a local and a country level census at the level of single building blocks. The downscaling is based on a dasymetric approach using an urban land use map as ancillary information. The model was applied after testing two different methods: a limiting variable and a fixed-ratio method. The latter performed slightly better and was applied on the entire study area.

The results of the proposed methodology can be used to perform population vulnerability analysis for night time scenarios which deals mainly with residential building typologies. The enhanced spatial detail influences the accuracy of the information on human exposure when the population map is used in risk assessment models.

INTRODUCTION

Population data are needed for estimating injured and casualties in disastrous events. The data are usually available from censuses, which account for specific entities (i.e. people, household) as statistic aggregations. Those aggregations refer to areal unit that do not reflect the patterns of the built

\textsuperscript{1} Dr., Global Security and Crisis Management (GLOBSEC), Institute for the Protection and Security of the Citizen (IPSC), Joint Research Centre of the European Commission (EC- JRC), Ispra, patrizia@tenerelli.it
\textsuperscript{2} Assoc. Prof. Dr., European Laboratory for Structural Assessment (ELSA), Institute for the Protection and Security of the Citizen (IPSC), Joint Research Centre of the European Commission (EC- JRC), Ispra and Department of Earthquake Engineering, Boğaziçi University, Istanbul, hancilar@boun.edu.tr
\textsuperscript{3} Dr., Monitoring Agricultural Resources (MARS), Institute for Environment and Sustainability (IES), Joint Research Centre of the European Commission (EC- JRC), Ispra, javier.gallego@jrc.ec.europa.eu
\textsuperscript{4} Dr., Global Security and Crisis Management (GLOBSEC), Institute for the Protection and Security of the Citizen (IPSC), Joint Research Centre of the European Commission (EC- JRC), Ispra, daniele.ehrlich@jrc.ec.europa.eu
environment, its use and thus of population distribution. Thematic maps which are based on such predefined areal units are also called choropleth maps.

Socio-economic census data can be disaggregated using spatial techniques (Eicher and Brewer 2001, Mrozinski and Cromley 1999, Chen 2002) which place non-spatially explicit attributes over fine scale geographical units (reporting zones). This is needed when performing analyses which integrate the census data with physical parameters (i.e. environmental data, built-up maps, topography) which follow natural boundaries or regular grid cells. The disaggregating techniques may be based on ancillary data that are normally related to the land use. This paper describes the methods that are used to map population density, focusing on the main downscaling techniques and different types of ancillary parameters, including remote sensing data. The paper also presents the methodologies which are used to combine the building fragility with the population data for estimating casualties.

A case study wherein a population downscaling model is applied for the Larger Urban Zone (LUZ) of Vienna is presented. In this application the residential population density is disaggregated using a model based on an urban land use map as ancillary information. The case study was conducted as part of the EC FP7 Project SYNER-G (Systemic Seismic Vulnerability and Risk Analysis for Buildings, Lifeline Networks and Infrastructures Safety Gain, https://www.syner-g.eu/). Syner-G aims to assess the seismic risk of physical and socio-economic systems which are characterized by different components (i.e. buildings, lifelines, population). The general methodology is based on systemic interactions within the physical and socio-economic systems, and between them (Franchin et al. 2013). Within this methodology specific socio-economic models integrate the physical damage with the socio-economic losses (Khazai et al. 2012). The disaggregated population map meets the input data requirements of the casualties’ assessment model. This model, in fact, combines the building fragility with fine scale population data for estimating the human losses.

**DOWNSCALING MODELS FOR POPULATION DENSITY MAPPING**

Population data at sub-national level are available in Europe by aggregating fine scale population census information conducted every 5 to 10 years. Census units have a spatial dimension which commonly corresponds to building aggregates. For public purpose the figures are released at coarser units often matching administrative units at different scales. In many cases census data are only available at the commune level. In some cases demographic and socio-economic surveys are conducted using grid cell spatial units. Grid-based representations of population offer several advantages when population data must be integrated with a representation of settlements or environmental phenomena (Martin 2009). More precise demographic data are collected in local administrative offices, but are normally not accessible or not available as geo-referenced digital data format.

When studying exposure to natural hazards, the spatial detail of the elements exposed to risk affects the scale of analysis and allows scenarios to be performed with different levels of approximation. The census spatial units therefore affect the detail at which the information on human exposure can be provided. When it is not possible to collect ad hoc demographic data from the field, downscaling or spatial disaggregating techniques can be used to address heterogeneity within census units, displacing the population density to smaller and more homogeneous spatial units.

Downscaling techniques can map three types of population distribution: residential, ambient and time-specific. Time-specific and ambient population maps are based on spatio-temporal models, which take into account the movements of population during different times of the day, through a given area (Martin 2009, Ahola et al. 2007). Ambient population refers to an average distribution of population over 24 hours (Dobson et al. 2000). When mapping ambient or temporal distribution of the population, the input data collection is much more challenging. Two datasets are needed in this case: the map of the activity location, or physical features where they take place, and the census of the mobile population, which include statistics on tourism, all work related travels, temporary accommodations, education, traffic, etc. (Bhaduri 2007, Martin 2009, Ahola et al. 2007, McPherson and Brown 2004, Freire 2010).

The most usual downscaling methods can be classified as geostatistical methods and areal interpolation methods (Wu et al. 2006). In geostatistical models continuous density surfaces (isopleth
Areal interpolation methods are commonly used in population downscaling, they apply a homogeneous zone approach where the census units are disaggregated into smaller enumeration zones. Areal methods, which interpolate census data using ancillary information, are generally referred to as dasymetric, or “intelligent areal interpolations”. The dasymetric models normally apply one downscaling parameter. The ancillary information (proxy) mostly used is the land use/land cover or built-up map. Dobson et al. (2000) applied a multi-dimensional dasymetric model which includes multiple physical parameters to map ambient population in 30 x 30 arcsec grid cells at the global scale. In dasymetric models, the categorical or continuous proxies are related to population through sampling techniques (Wu et al. 2008, Mennis 2003, Mennis and Hultgren 2006), regression analysis (Yuan et al. 1997, Wu et al. 2006, Chen 2002, Silván-Cárdenas et al. 2010, Lu et al. 2010, Briggs et al. 2007) or expert knowledge (Eicher and Brewer 2001). Gallego et al. (2011) described and compared different methods for assigning the ratio of population density to different land use classes, dividing them in two main classes: “fixed ratio methods”, which assign the same ratio of density to all classes, and “limit-based methods”, which apply upper limits to the population classes; the second class of methods provide the best performance.

Another important distinction must be made between volume preserving methods and non-volume preserving methods. In volume preserving methods the pycnophylactic constraint is applied, which means that the sum of population from all zones coincide with the known population (Tobler, 1979, Gallego et al, 2011).

The criteria for the choice of method for population interpolation are the data availability and quality obtained (accuracy, scale), together with the purpose of the investigation. Dasymetric downscaling is particularly suitable for discrete variables with approximately homogeneous intra-zone distribution and inter-zones with actual changes, as is the case of many socioeconomic variables (Cai, 2006). Dasymetric mapping is commonly used for population downscaling when accurate ancillary variables with high correlation with population are available.

CASUALTY ESTIMATION IN AN EARTHQUAKE LOSS MODEL

Seismic risk might be assessed in regional and/or urban scale. Regional estimates of damage to built-environment and assessment of human losses can be achieved using region-specific theoretical/empirical vulnerability relationships, i.e. magnitude-casualty relations, in connection with regional inventories of physical and social elements exposed to risk. In urban scale, more detailed inventories of elements at risk are required, to be used with analytical vulnerability relationships for the estimation of earthquake losses. Grid based (geo-coded) inventories of building stock and demographic data are needed for urban earthquake loss assessment. For an urban level analysis, the data should include construction year, occupational type, construction material type and number of floors for each building class as well as the number of people residing at each geo-cell. Hence, the sophistication and completeness level of inventories of elements at risk will determine the level of analysis that will be used in loss estimation (Hancilar et al., 2010). Figure 1 presents an illustration of the seismic risk assessment approaches and the entry of population data for each methodology.
Casualty Assessment at Regional Scale: Empirical Approach

Population data, i.e. population density information, is required for the quantification of number of exposed people. Once the human exposure is assessed, casualties can be estimated through empirical models which rely on fatality rates of exposed population as functions of earthquake intensity, regional vulnerability level, regional growth rate and time of the day. Regional vulnerability depends on the level of economic development, general building types and building occupancy rates with various degrees of refinement and resolution.

Empirical models have been generally developed on the basis of the regression analysis between the number of casualties caused by building damages and the earthquake magnitude for different ranges of population densities over the affected areas (e.g. Samardjieva and Badal, 2002; Nichols and Beavers, 2003). Jaiswal et al (2009) proposes a global empirical model which utilizes historical earthquake casualty data and provides a country or region-specific earthquake fatality rate as a function of shaking intensity. Differently from the former models, the latter makes use of ground shaking intensity (e.g. MMI), a spatially varying parameter and an indicator of direct impact of ground motion on built environment.

Casualty Assessment at Urban Scale: Analytical Approach

Analytical approach estimates casualties (death and injury) on the basis of estimated damage probabilities of different building types in the exposed building stock. For a general building stock the following parameters affect the damage and loss characteristics: structural (system, height, and building practices), non-structural elements and occupancy (such as residential, commercial, and governmental).

Once the grid-based building inventory is classified with respect to pre-defined typologies, the exceedance probabilities of different damage levels (i.e. slight, moderate, extensive and complete) are computed for each of building classes at a given level ground shaking intensity. Then the number of causalities is obtained as a product of the casualty ratio for each building damage class and the number of occupants in buildings of that damage class. At this point, two inputs are needed:

i. Relationship between the general occupancy classes and the model building type with specific casualty inputs provided for each damage state in combination with occupancy data and time of the event.

ii. Total population at grid cell
INTEGRATING URBAN LAND USE AND POPULATION CENSUS DATA: THE CASE STUDY OF VIENNA

Within the SYNER-G project a disaggregation model was applied to downscale the census population data for the city of Vienna. This model disaggregates the residential population using an urban land use map as ancillary data. The model is based on a dasymetric approach that is also applied to produce the Population density grid of EU-27+ (version 5) (Gallego 2010).

This application uses the Urban Atlas (UA) land use map, which is derived from HR satellite imagery, with support of other reference data. The UA covers the central city area and good part of the hinterland (Large Urban Zone), at a nominal scale of 1:10 000. It discriminates, at the building blocks level (defined by road intersections), between five built-up density classes, two built-up uses and three transportation network classes (Urban Atlas, 2010). Two different sources of population census data were also used (Figure 2):

i. a local population census on the central city area (about 290km2), from 2001, available at the level of census polygon of approximately 25 ha in size;
ii. a country level population census from 2006, available as 1 km grid cell.

The study area was divided in a training area and a test area (Fig. 3). The training area was used for estimating the population density of each urban land use class (c) derived from the UA. The test area was used for testing two different population downscaling models. A limiting variable and a fixed ratio model were tested (Gallego et al., 2011; Eicher and Brewer, 2001).

In order to take into account the different population density that occurs in the central city district and in the hinterland, the study area was also divided in two different zones: the central area zone (stratum 1) and the hinterland (stratum 2) (Figure 3). The stratum 1 corresponds to the area where the local - polygon based - census is available. For the stratum 2 only the grid cell census was available. The modeling steps are described in the following.

Figure 2. Population census data for the Large Urban Zone of Vienna: Local population census (left side), country level population census (right side)
Figure 3. Population zones for the study area in Vienna

**Sampling of Population Density Classes**

The population density classes (c) used to downscale the population density value were derived from the UA classes through the class aggregation presented in Table 1. The class aggregation assumes that no population and no residential population should be distributed in the commercial and industrial areas and other public facilities. However, in those classes some residential population could be found due to the intrinsic limitation of the UA map which contains a high uncertainty level when it maps different building use. Therefore, those classes were aggregated as “Preliminarily Non Residential” areas.

In order to assign the population density to each urban land use class, a set of sample polygons was selected in the training area. Those polygons correspond to census units which are covered by homogeneous urban land use classes (pure c census zone), excluding the “no population” land use area. The “pure c” census zones where selected according to the following condition (Eq. 1):

\[
\frac{a_{cg}}{A_g - a_{no\ pop\ g}} > 0.80
\]

where:
- \(a_{cg}\) = area of the urban land use class c in the census unit g;
- \(A_g\) = total area of the census unit g;
- \(a_{no\ pop\ g}\) = total area of “no population” class in the census unit g.

In the hinterland area (stratum 2) the training samples (pure c census zone) correspond to grid cells (pure c cells), while in the central area the samples correspond to pure c census polygons.

The population density value of each class (Cd) was calculated as the average population density of the class samples, weighted by the area. The final population density for each class and each stratum are listed in Table 1. The UA classes “Discontinuous Very Low Density Urban Fabric” and “Isolated Structures” were aggregated because the difference in terms of building density was not significant in the stratum 2. This was verified by visually comparing and validating the UA map.
against the VHR satellite imagery available on Google Earth (GE). For those two classes no samples were found in stratum 1, therefore the density value was derived by linear extrapolation. The same approach was applied to assess the density value of the “Continuous Urban Fabric” class in stratum 2, in this case only one sample was found. Finally “Medium Density Urban Fabric” and “Discontinuous Low Density Urban Fabric” were considered as one single class in stratum 2, in this stratum in fact the difference in terms of building density, estimated by visual inspection on GE, was not significant.

Table 1. Population density values for each UA class and each stratum

<table>
<thead>
<tr>
<th>UA class</th>
<th>Population density class (c)</th>
<th>$C_d_{s1}^*$</th>
<th>$C_d_{s2}^*$</th>
<th>Density rank (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous Urban Fabric</td>
<td>Dense</td>
<td>250.800</td>
<td>1.306</td>
<td>1</td>
</tr>
<tr>
<td>Sample number</td>
<td></td>
<td>59</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Discontinuous Dense Urban Fabric</td>
<td>Discontinuous Dense</td>
<td>153.680</td>
<td>0.800</td>
<td>2</td>
</tr>
<tr>
<td>Sample number</td>
<td></td>
<td>24</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Discontinuous Medium Density Urban Fabric</td>
<td>Discontinuous Medium Density</td>
<td>74.160</td>
<td>0.272</td>
<td>3</td>
</tr>
<tr>
<td>Sample number</td>
<td></td>
<td>11</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Discontinuous Low Density Urban Fabric</td>
<td>Discontinuous Low Density</td>
<td>25.200</td>
<td>0.272</td>
<td>4</td>
</tr>
<tr>
<td>Sample number</td>
<td></td>
<td>8</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Discontinuous Very Low Density Urban Fabric;</td>
<td>Discontinuous Very Low Density</td>
<td>0.859</td>
<td>0.004</td>
<td>6</td>
</tr>
<tr>
<td>Isolated Structures</td>
<td></td>
<td>0</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>Sample number</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airports; Construction sites; Industrial,</td>
<td>Preliminarily non</td>
<td>17.560</td>
<td>0.040</td>
<td>5</td>
</tr>
<tr>
<td>commercial, public, military and private units;</td>
<td>residential</td>
<td>51</td>
<td>207</td>
<td></td>
</tr>
<tr>
<td>Port areas; Sports and leisure facilities.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample number</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fast transit roads and associated land;</td>
<td>No population</td>
<td>0.000</td>
<td>0.000</td>
<td>7</td>
</tr>
<tr>
<td>Other roads and associated land; Railways and</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>associated land; Agricultural and others**.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $C_d_{s1}$ = class density for the stratum 1; $C_d_{s2}$ = class density for the stratum 2

** Semi-natural areas; Wetlands; Forests; Green urban areas; Land without current use; Mineral extraction and dump sites; Water bodies.

The fixed ratio and the limiting variable methods were tested using a 2 km resolution map as starting point.

**Fixed Ratio Model**

The population count of each 2 km grid cell was disaggregated, per each stratum, at the level of the urban land use class units, according to the following Eq.:

$$P_{cg} = \frac{P_{g} C_d_{cxs}}{\sum_{c} A_{cg} \times C_d_{cxs}} \quad (2)$$

where:

- $P_{cg}$ = population density of the urban land use class $c$ in the census unit (2km grid cell) $g$;
- $P_{g}$ = total population in the census unit $g$;
- $C_d_{cxs}$ = population density for the land use class $c$, in the stratum $sx$;
$a_{cg} =$ area of the land use class $c$, in the census unit $g$.

**Limiting Variable Model**

The limiting variable model is described through the following steps (Gallego et al., 2011):

i. a uniform population density value derived from the original census dataset is assigned to all the land use polygons in each 2 km grid cell ($D_g$);

ii. the Land use classes are ranked and an index ($r$) is assigned from the lowest to the highest population density value (from “Continuous Urban Fabric”- $r=1$ – to ”No population” - $r=7$) (see Table 1);

iii. If the $D_g$ is above the population class density of the given stratum, the density value is modified: the population density become equal to $Cds1$ or $Cds2$, and the population in excess is redistributed among the more dense classes according to the following Eq. (4.4) (Gallego et al., 2011):

$$P_{cg}^{ef} = D_g + \frac{a_{cg}^{ef} \times (D_g - C_{cg})}{\sum_{r'=r}^7 a_{cg}^{r'}}$$  \hspace{1cm} (3)

This step is iterated for each $r' > r$, seven iterations in total, one for each land use class, were performed;

iv. at the end of the process if there is still some excess population it is redistributed to all the land use classes proportionally to the class density ($Cds1$ or $Cds2$).

**Selection and Application of the Model**

The disaggregated population maps resulting from the two tests were re-aggregated at 1 km grid and compared with the original 1 km grid of census population through a correlation analysis (Figure 4). This was done, in absence of additional reference data, to assess the models’ performance and select the most suitable model. The result showed a strong linear (positive) correlation and a similar performance for the two models, with a coefficient of determination ($r^2$) of about 0.93 for both the model outputs (Table 2).

The agreement between the reference and the model-simulated data was estimated. For each model output the Total Absolute Error (TAE), this was calculated as the sum of the residual absolute value. The TAE measures the average model-performance error in a way which is less ambiguous (Willmott and Matsuura 2005) and more robust in presence of outliers (Gallego et al. 2011, Legates and McCabe 1999) than the commonly used Root Mean Square Error (RMSE). The TAE of the limiting variable model was slightly higher than that of the fixed ratio model (Table 2). The fixed ratio model was therefore applied to the whole study area (test area plus training area), for each 1km grid cell.

Table 2. Statistic indicators for each model output comparison with the reference population data

<table>
<thead>
<tr>
<th>Statistic indicator</th>
<th>Limiting variable model</th>
<th>Fixed ratio model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r^2$</td>
<td>0.931</td>
<td>0.936</td>
</tr>
<tr>
<td>TAE</td>
<td>260241</td>
<td>242334</td>
</tr>
</tbody>
</table>
The map in Figure 5 shows the final disaggregated population map. The tested models preserve the pycnophylactic property (Tobler 1979) which means that the original total population set (from the 1km grid census) is preserved. The advantage of this modelling is to have a population density disaggregated to spatial units which have homogeneous size and are usually smaller than the original census dataset. This spatial detail can affect the information accuracy on human exposure when the population map is used in risk assessment models.

The main limitation of this method is the difficulty of getting a population reference dataset suitable to obtain a sufficient sample of census unit covered by a single land use class. Most of the census dataset, in fact, have a too coarse spatial unit. In this case study the two population census datasets were enough detailed to allow the collection of an appropriate sample number both in the central city area, and in the peri-urban area.
DISCUSSION AND WAY FORWARD

This application only deals with residential population, it is therefore suitable for casualties estimations in night-time scenarios. Day-time population maps are more useful for human exposure analysis in disaster risk assessment and risk management. The data that are necessary to perform population spatio-temporal analysis (mobile population database and activity location map) were not available for the SYNER-G case studies and scale of analysis. However the building vulnerability analysis performed in the SYNER-G project deals mainly with residential building typologies, therefore the disaggregated population map can be used, together with the available building fragility data, to perform population vulnerability analysis for night time scenarios.

In order to integrate the fine detail population data with the building fragility information the fragility and population information must be standardized in the spatial domain. In the syner-g project the building fragility data are collected at the level of single buildings, using building by building inventories or sampling approaches. Therefore, in order to be combined with the population data, the building fragility information should be statistically aggregated at the level of land use class.

CONCLUSIONS

The analysis and modelling of disaster outcomes require population total and population attributes at fine scale. Yet, most population datasets, mainly for privacy reason, are released in a much coarser aggregated form. The disaster modellers and crisis management specialists have no other option than to disaggregate coarse population figures into finer geographical units based on population distribution proxies that include land cover land use and built up information layers.

This paper provides a review the state of the art and theory and proposes one method based on residential population census and urban land use map at the city level. This method adapts the model used for the Population density grid of EU-27+ to the city of Vienna. The availability of the population census at a quite fine resolution justified the choice of Vienna as case study. The method disaggregates the population at the level of building blocks which are usually smaller than the original census dataset. When applied to case studies where the input census data is available at coarser resolution, the downscale result is certainly more important, however the example is useful to assess the methodology to be applied elsewhere.

The proposed approach is suited for the scope of the vulnerability analysis performed in the SYNER-G project. Different models of vulnerability and casualty estimation may require population data at different resolution and with different details, especially in terms of temporal resolution. Time-specific population data are very difficult to produce due to the challenging of obtaining a complete spatio-temporal database, however such data are necessary to predict casualties in any time. Therefore there is margin for improvement in the casualty modelling considering the urban population growth and the great challenges that this will pose for urban mobility in the coming years.

REFERENCES


