International Journal of Disaster Risk Reduction 10 (2014) 442–451

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr

A global exposure model for disaster risk assessment

Andrea De Bono [a,b,*], Miguel G. Mora [c]

[a] Institute for Environmental Sciences, University of Geneva, Geneva, Switzerland
[b] United Nations Environment Programme, Global Resources Information Database, Geneva, Switzerland
[c] International Centers for Numerical Methods in Engineering (CIMNE), Universitat Politècnica de Catalunya, Barcelona, Spain

ARTICLE INFO

Article history:
Received 9 December 2013
Received in revised form
7 April 2014
Accepted 5 May 2014
Available online 20 May 2014

Keywords:
Urban
Exposed assets
Urban produced capital
Population
GAR
Downscale

ABSTRACT

The Global Exposure database is being produced for the Global Risk Assessment 2013, part of the Global Assessment Report 2013 (GAR 2013). It aims to map at a granular geographical level the world's capital stock in urban areas. It is designed primarily to assess the risk of economic losses as consequence of natural hazards at a global scale.

The Global Exposure database for GAR 2013 (GEG-2013) is an open exposure global dataset at 5 km spatial resolution which integrates population and country-specific building typology, use and value. It is currently suitable mainly for earthquakes and cyclones probabilistic risk modeling using the CAPRA platform (http://www.ecapra.org).

This paper describes the development of the GEG-2013. The database is based on a top-down or "downscaling" approach of national/regional socio-economic and building type information. These information are transposed onto a regular raster dataset (grid format) using a geographic population distribution model as a proxy.

1. Background

Exposure is the collection of the elements at risk to potential losses [1] or that may suffer damage due to a hazard impact. This paper describes how a global exposure database was generated and used for the quantification of both the exposure and the vulnerability, to support the earthquakes and cyclones probabilistic risk modeling in the Global Assessment Report on Disaster Risk Reduction (GAR) process [2].

Models for assessing quantitative risk from natural hazards use building counts, as well as statistic aggregations of buildings at different areal units, to estimate physical damages. The exposed elements include people, resources, infrastructures, production, goods, services or ecosystems and coupled social–ecological systems. In the Global Exposure database for GAR 2013 (GEG-2013) the physical exposures is represented through the inventory of buildings in urban areas, called here "the building stock".

During the last ten years, various remarkable products, related to exposure at a global level, have been released into the public domain.

Among those, HAZUS is multi-hazard loss estimation software [3]. It mainly focuses on the United States environment, assigning structural classes to a grid representing the occupancy. These structural classes are then directly linked to damage functions.

The PAGER inventory database [4] provides the distribution of housing/dwelling units rather than the distribution of buildings types.

The Global Exposure Database for Global Earthquake Model [5], as yet still unpublished, will certainly be the most advanced global exposure database for earthquake risk assessment. GED4GEM is a global, multi-scale, and regionally/locally driven exposure database [6]. The GED will feature data at four different geographical scales:

- Level 0 (at country-level), the GED is a gridded dataset, at 30″ resolution, including data compiled through

---

* Corresponding author at: International Environment House (IEH) 11, chemin des Anémones 1219 Châtelaine Switzerland.
E-mail address: debono@unepgrid.ch (A. De Bono).

http://dx.doi.org/10.1016/j.ijdrr.2014.05.008
2212-4209/

national census surveys. It contains statistical information about the number and distribution of dwellings and/or buildings fractions in a country, classified by building type and residential/non-residential. Where available, the information about the value of buildings per meter square is also provided.

- Level 1 (at sub-country level) the GED is a gridded dataset, at 30″ resolution, including comparable information to level 0, but with data coming from sub-national or district-level surveys. Where available, the information about the value of buildings per meter square is also provided.
- At Level 2 (the local level), the GED consists of a gridded dataset based on datasets obtained by aggregating sub-grid/finer data, coming from building-by-building surveys.
- At Level 3 (single buildings) a vector GIS dataset includes single buildings/dwellings level information that is provided when available.

At a regional level, the Pacific risk exposure database[1] co-funded by the Global Facility for Disaster Reduction and Recovery (GFDRR, World Bank), and the Japanese Government, provides a comprehensive geo-referenced catalog of facilities for 15 Pacific countries. It includes, besides population distribution, more than 400,000 high-resolution satellite images, including building footprints for structural classification and use. The database also includes information on main infrastructures such as roads, bridges, dams, ports and airports, as well as on utilities and major crops.

More recently, the GFDRR is working on an exposure model for natural catastrophe risk assessment in Latin America and the Caribbean Region [7]. The aim is to produce a global open exposure dataset based upon population, country specific building type distribution and other global/economic indicators such as World Bank indices that are suitable for natural catastrophe risk modeling purposes.

The most accurate way to produce an inventory of exposed assets consists in collecting individual georeferenced data within the economic evaluation of each asset: this is a classical bottom-up approach. It is usually applied for local (sub-city–city level) models and partially for larger scaled models (OpenDRI [8], GED 4GEM levels 2 and 3) which consider the characteristic of each of the exposed components such as buildings and more generally critical infrastructures. Many datasets have been and are constantly produced, by the World Bank's staff especially at a local scale, throughout the Bank's regions' projects; however, these projects have not produced, or yet made available, global scale datasets i.e. dataset covering the entire world. Even putting together the existing high resolution datasets, they would not provide a global coverage. Furthermore, these dataset does not present a uniform spatial distribution that would allow the buildup of bottom-up global exposure grids [9].

An exposure database that includes a global inventory of critical facilities based on a pure bottom up approach would require considerable human and economic efforts, and is beyond the scope of this project.

As the bottom-up approach is not available, we employed a spatial disaggregation, which consists in downscaling the available data (e.g. by administrative units) by means of auxiliary information and statistical techniques.

The development of GEG-2013 is based on a top-down or “downscaling” approach, where information including socio-economic, building type and capital stock at a national scale are transposed onto a regular grid, using geographic population and Gross Domestic Product (GDP) distribution models as proxies.

The exposure indicator is the number of persons subdivided by socio-economic class living in a specific construction type in a determined geographical location. This is used as a base for distributing the exposed economic value of the building stock.

2. Objectives

The objective of the GAR global risk assessment is to provide comparable disaster risk metrics for all countries and territories in the world, through a set of risk classes representing the likely order of magnitude of loss.

The GEG-2013 is a key element in the estimation of the risk developed in the GAR 2013 process. Disaster risk is considered to be a function of hazard, exposure and vulnerability, expressed as the probability of loss of life, injury or destroyed or damaged capital stock in a given period of time [2].

The purpose of GEG-2013 is to generate a global evaluation of exposed assets, in urban areas, in order to provide specific exposure input data to be used in further, coarse grain type, risk assessments [10]. It only considers the direct physical damage to urban buildings, combining both national socio-economic information, as well as a geographic population and GDP distribution models, as the main sources of information. The GEG-2013 top-down approach has the advantage to guarantee comparability across all countries and territories worldwide.

In order to develop a consistent framework for the GEG-2013 three essential aspects of the exposed assets are analyzed, by responding to the following questions:

1. What are the “exposed assets”?
2. How can we classify them?
3. How we can establish a monetary evaluation of the exposed assets?

2.1. *What are the “exposed assets”?*

The exposed assets considered in GEG-2013 are the building stock including both dwellings and buildings, and the population living or working inside them.

In the 2013 edition of GAR, the global probabilistic risk analyses will include only the economic component of the losses affecting urban agglomerations; hence the exposure database will cover urban areas with generally more than 2000 inhabitants. This limitation can be justified by the fact that urban human settlements are the pivot point of

---

[1] PCRAFI: http://pcrafi.sopac.org/.

human life, concentering more than 50% of the population and 70–90% of economic activity [11].

At a global scale, Earth observations made by remote sensing are an efficient tool to define and locate built-up areas. Remote-sensed measurements are consistent and can be made comparable in space and time. Several datasets delimiting the built environment exist, including Globcover, GLC 2000, MODIS, GRUMP [12] and Impervious Surfaces Areas (ISA). As the MODIS 500 m, presently, represents the most optimum option [13,11] the building stock in GEG-2013 will be defined by the built-up areas classes of MODIS 500 m. It will be intended as a statistic aggregation of buildings.

The urban areas definition in the GEG-2013 context is based on physical attributes; that is areas that are dominated by the "built environment". The built environment includes all non-vegetative, human-constructed elements, such as buildings, roads, runways, etc. [11]. Therefore, in this context when we refer to urban areas or urban agglomerations we implicitly refer to the above definition of built-up environment.

The last component of the exposed assets is the population utilizing the building stock.

In our top down approach we will use a pre-existing model of the distribution of the population on the territory. The exposed people will correspond to those people who occupy the urban area defined above.

2.2. *How can we classify them?*

Datasets resulting from the definition and delimitation of the building stock are a binary representation of the reality who simply indicate the existence or not of a built-up area. Our aim consists now to characterize this "aggregation of buildings" in terms of their use and their structure.

The "use" means the subdivision of the building stock in different classes based on their occupancy: residential, industrial, commercial, institutional, educational, and health purposes.

The "structure" refers to the building type, conforming to the PAGER building type's classification.

In the GEG-2013 top-down process proxies are used, as direct data concerning the buildings do not exist with the same resolution for the whole earth. In the case of the building use, population data will be used to obtain information on the constructed area. For example, a certain percentage of people, employed in the industrial sector per a determinate country, will be related to the proportion of industrial buildings in that country.

The integration of the model of the world population distribution into the PAGER building types classification was generated by the World Agency of Planetary Monitoring & Earthquake Risk Reduction (WAPMERR). This model [14] subdivides the settlements into three categories based on their population size, according to the Satterthwaite classification [15] in order to define a level of complexity of the urban settlement. The level of complexity is a parameter that will be further used for the characterization of vulnerability functions based on building typology [16]. In cities highly populated (=high level of complexity) there are generally more controls and regulations and thus a better compliance with the design levels according to the hazard can be expected.

2.3. *How we can establish a monetary evaluation of the exposed assets?*

1. The "value" of an asset can be assessed as a "flow" or a "stock": value as a stock; a snapshot of accumulated assets – wealth/capital.
2. Value as a flow; a stream of benefits deriving from the asset – income/GDP/expenditure.

These two approaches are in essence "two sides of the same coin" and inherently inseparable; the flow is the benefit deriving from a stock, and the stock is likewise the sum of the discounted flows for a particular impact.

In the context of natural disasters, stocks represent the usual choice of unit for measuring exposure. This is especially true considering that a natural disaster could cause asset damage greater than the annual flow.

In the past editions of GAR (2009 and 2011), the economical exposure was based on the concept of flow. For GEG-2013 a stock (or "capitals") approach is used, focusing largely on produced capital in urban areas. It is important to emphasize that this does not consider indirect flow effects, which should anyway be captured in the capital valuation, or wider effects, which are explicitly out of the scope of GAR 2013.

The GEG-2013 uses World Bank data [17] and focuses directly only on buildings stocks without taking into account other typologies of infrastructures such as roads and bridges. Nonetheless the produced capital stock, as calculated by the World Bank (WB), also captures indirectly the costs of more general stocks at the city level.

3. Data and methodology

3.1. *Main components*

This section describes the main datasets and underlying databases that have been used to produce the GEG-2013 Fig. 1.

3.1.1. *Building stock definition: the urban mask*

The building stock was delineated by using the built-up areas class extracted from the MODIS 500-m [13].

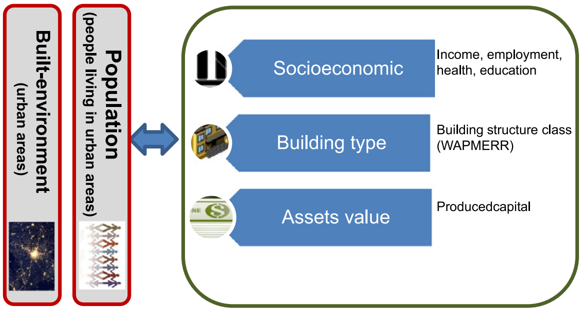

Fig. 1. The five thematic main components of GEG-2013.

3.1.2. Building stock characterization: socio-economic indicators (SEI)

Socioeconomic indicators will be used as proxies to estimate the sectorial use of the building stock in accordance with the levels of complexity of each urban area. They consist in a set-of tabular data at country level.

A first set of indicators, groups the resident population into four classes, based on the income per capita, according to the World Bank classification [18] and they will be used to distribute the population in categories of dwelling according to their income.

Original data provided by the WB are converted using the GINI (Lorenz) curve for income distribution in a population. The result is an estimation of the percentage of population by level of income according to the four WB classes:

% People belonging to each category

1) Low income $\leqq$ 1025 GNI per capita (US$).
2) Middle-low income $\leqq$ 4035 GNI per capita (US$).
3) Middle-high income $\leqq$ 12,475 GNI per capita (US$).
4) High income $>$ 12,476 GNI per capita (US$).

A second group of indicators, including the non-residents, is subdivided by economic activities (industry, services and government), health coverage (number of hospital beds in public/private) and education for private and public sectors (number of pupils). They will be used to estimate the population distribution into categories of buildings according to their activities Table 1.

Socio-economic statistical data are commonly released as relatively continuous time series, and cover at least the last 20 years. 2007 was by far the most complete year for a majority of indicators, it was therefore chosen as base year.

3.1.3. Building stock characterization: building structure typology

WAPMERR [14] provided all the information concerning the construction types in each country, for three size categories of settlement complexity, as percentage of peoples living in the determinate building structure. In other words, the distribution of structural types was carried out according to the population that lives in each of them and not in accordance to the number of construction type per number of buildings.

The sources for building types employed by WAPMERR were 40% from census data, 25% from the WHE/PAGER project, 25% based on research, 9% based on UN reports, and 1% on HAZUS data.

The information provided by WAPMERR needs to be distributed by level of complexity in residential/nonresidential use. This is carried out based on the information found in the structural systems distribution catalog for residential/nonresidential use [4].

3.1.4. Building stock evaluation: produced capital stock

The produced capital stock [19] tends to be the most readily understood form of capital due to its tangibility and the quality of data collected on investment levels. The World Bank [17] published a dataset for 152 countries that provides broad estimates of the current (2005) capital stock of machinery and structures, based on the Perpetual Inventory Method (PIM) and historical Gross Capital Formation (GCF) data. Furthermore, the World Bank scale-up this estimate by 24% to account for the value of Urban Land.

According to PwC [20] the World Bank methodology appears to be the most consistent method of measuring produced and natural capital values at an international level, at the present time. Other valid alternatives can be represented by the UNU [21] and the [22] but they have estimations only for a limited number of countries.

3.1.5. Raster of population distribution

The primary source of global exposure information is a model that represents the distribution of people on the earth surface. A gridded population dataset is the result of this model and it is based on a regular grid, where each cell indicates the number of people living on it.

Presently only three (complete) gridded population datasets are available at a global scale: LandScan developed by The Oak Ridge National Laboratories (ORNL), [23–25],

Table 1
Socio-economic Indicators (SEI) and their sources.

Sector or use		Socio-economic indicator (SEI)	Official data sources
Population		Country population in 2007 and 2010	UN WPP 2011
Residents	Income	% Of people living below 1005$	WDI 2012
		% Of people living between 1005$ and 3975$	
		% Of people living between 3975$ and 12275$	
		% Of people living above 12276$	
Non-residents	Employment	% Of people employed in agriculture	WDI 2012, ILO 2012, UNSD 2012
		% Of people employed in industry	
		% Of people employed in services	
		% Of people employed in government	
		Number of employed/total population	
	Health	Number of beds per 1000 people private	WDI 2012
		Number of beds per 1000 people public	
	Education	Number of pupils primary and secondary (private)/total population	UNESCO 2012
		Number of pupils primary and secondary (public)/total population	

GPW and GRUMP from the Socioeconomic Data and Application Center (SEDAC) [12].

LandScan represents an ambient population distribution at 30″ resolution (approx. 1 km equator). The LandScan (unpublished) algorithm, uses a multi-layered, dasymetric, spatial modeling approach to reallocate populations based on layers representing land use/land cover, high resolution imagery analysis, transportation networks, elevation and slope.

Although GRUMP could be a quite good option for GEG-2013, it presents two major inconveniences: the delimitation of urban settlements is based on obsolete Nightlight DMSP imagery, and the distribution of rural population is done directly by administrative units without any modeling. Until the complete release of new open data from the WorldPop initiative [26], LandScan is a reasonable alternative for the requirements of the GEG-2013, as previous studies performed with both GRUMP and LandScan, showed a better correlation with LandScan [1,27].

3.1.6. Raster of GDP distribution

Similar to the gridded population dataset, the raster of GDP represents the distribution of the Gross Domestic product on the earth surface. We will use an unpublished dataset originated by the World Bank DECRG team for GAR and further extrapolated by UNEP/GRID – Geneva 2009, and successively updated for the 2011 edition [28]. The Global Regional Product (GRP) and national GDP data are allocated to 30″ grid cells in proportion to the population residing in that cell (based on LandScan).

3.2. Compilation and harmonization (missing data, assumptions)

Most statistical socio-economic indicators, as well as the capital stock from the World Bank, do not cover the full list (216) of countries/territories used in the GAR assessment.

In the case of the socio-economic indicators, in order to complete the missing values four different methodologies were employed (in order of preference):

1. Data were searched for through national statistical offices: this method was used particularly for France “DOM-TOM” by using the National Institute for Statistics (INSEE).
2. Data came from “global” unofficial databases such as the CIA Factbook this was extensively used in various cases.
3. Data were estimated using proxies (e.g. GDP versus capital stock).
4. Data were assumed to be equivalent to countries considered as “similar” in terms of geographic position, development, economy, and sovereigns.

As far as the capital stock is concerned, data provided by the World Bank only cover 152 countries. This had to be expanded to cover the remaining 66 economies. The International Institute for Applied Systems Analysis (IIASA) provided a list of data for 35 countries extracted from its unpublished internal database.

The missing data on produced capital for the rest of countries/territories were evaluated following the approach outlined by PwC [20], they suggest three main methodologies as follows in order of decreasing robustness:

a. Apply the World Bank algorithms using World Bank Gross Capital Formation13 (GCF) data.
b. Apply the World Bank algorithms using data on Gross Fixed Capital Formation (GFCF) from the IMF14 and EconoStats15.
c. Apply the World Bank algorithms on GCF: GFC data are calculated by using a fixed ratio between GDP and GFC.

The third methodology has been applied exclusively to smaller countries/territories where only GDP data were available. More detailed information and a list of estimated countries is available in De Bono [29].

3.3. Spatial disaggregation (downscaling) and data integration

Data on SEI, capital stock and building classes are generally provided at the national scale by administrative units. In this section we will illustrate the methodology and assumptions in order to spatially disaggregate them in smaller and uniform geographical units.

The uniform geographical units consist of a reference grid at 2′30″ (or approx. 5×5 km^2 at equator). The choice of the resolution is justified by three significant reasons.

Firstly a 5×5 resolution is considered as satisfactory in order to capture effects from large scale hazards such as earthquakes and cyclone wind.

Secondly to guarantee consistency in the results: SEI proxies are at a national scale, for certain large and non-uniform countries a disaggregation to smaller cell sizes is stretching the downscaling process too far.

Thirdly to reduce the size of the database in order to optimize the time of analyses needed for obtaining the results from the probabilistic risk calculation.

It is important to emphasize that the size of the cell is only one side of the “resolution” of a dataset, which also includes the thematic aspect, or in other words the amount and the quality of the information captured within it.

Essentially our distribution model will follow a multi-layered, dasymetric, spatial approach where data by administrative units are converted to a regular finer surface (reference grid) by means of ancillary data constituted by Population and GDP rasters.

The whole spatial disaggregation process and data integration into a geo-database is displayed in Fig. 2 and outlined in five steps.

3.3.1. Step 1: Extract the urban population

The “built-up” class from MODIS 500-m were selected, resampled to 30″ and used to build the mask or the so-called urban mask. This mask was successively employed to extract the population from the gridded population dataset (LandScan). Output includes people living in urban areas (Fig. 3). This first result needs more improvements: indeed urban populations in some geographical areas remain underestimated. In all OECD countries and most developing countries, urban agglomerations delineated by the urban mask, capture, in a satisfactory way the underlying population; the urban mask fits more than 90% of the cells containing at least 500 inhabitants. The worst associations occur in some regions of India and Bangladesh, where large sectors of high density populated cells remain excluded from the mask. This is principally related to remote sensing incertitude, together probably with the downscaling approach used in the LandScan process.

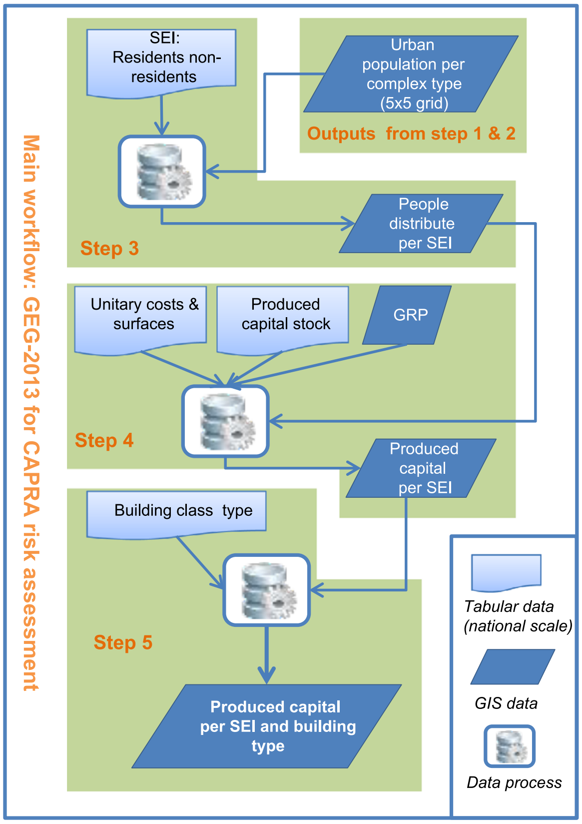

Fig. 2. Data processing workflow.

In order to avoid this inconvenience, clusters of contiguous cells with at least 2000 inhabitants per cell from LandScan were integrated into the MODIS mask.

The OECD and the European Commission [30] define urban areas as at least 1500 inhabitants per kilometer square; at global scales our tests suggest the use of a more conservative value of 2000 inhabitants per cell.

3.3.2. Step 2: Establish the “complex type” and aggregate data onto the 5 × 5 km^2 reference grid

The complex type characterizes the urban settlements based on their size according to Satterthwaite classification where [15]

- Complex type 1: $\leqq$ 20,000 inhabitants: upper urban.
- Complex type 2: between 20,000 and 2000 inhabitants: lower urban.
- Complex type 3: $\leqq$ 2000 inhabitants: rural.

This classification scheme was suggested and utilized by WAPMERR [14] in the characterization of building structural typology.

Contiguous cells were associated with the corresponding complex type class on the basis of the total population of the corresponding urban settlement. Final outputs of this step include 153,477 urban settlements (patches) subdivided into three complex types for the whole surface of the Earth.

The next operation consists to transpose and aggregate the already calculated urban population at 30″ to the reference grid at 2′30″ (or approx. 5 × 5 km^2 at equator).

3.3.3. Step 3: Distribution of the population per socioeconomic indicators (SEI)

With the objective of estimating goods dedicated for residential and non-residential use, the population by level of income, employment, instruction and the number of beds in the health sector needs to be estimated.

The detailed algorithms developed by CIMNE to manipulate and transfer these variables at subnational scale (5 × 5 grid cells) can be consulted in the CIMNE report [16] where the following sub-steps are outlined:

1) Population distribution by income level following the four income classes established by the World Bank [18].

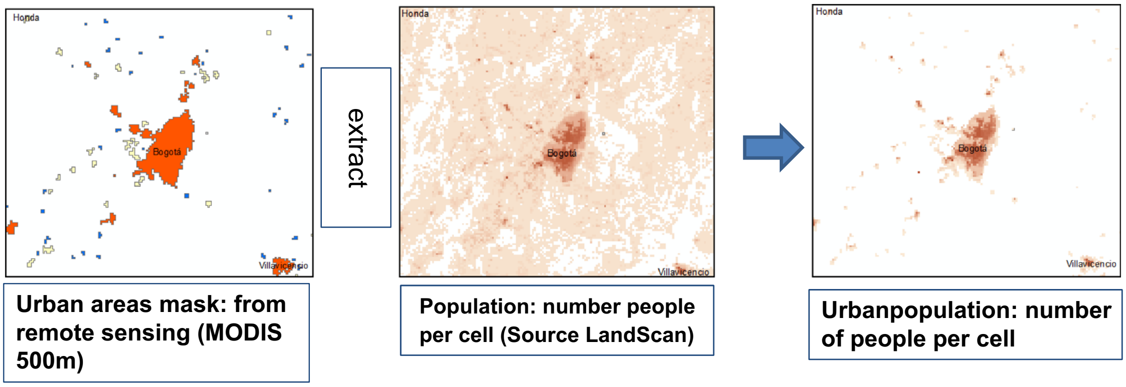

Fig. 3. From total to population living in urban area (region of Bogotá, Colombia).

2) Evaluation of the labor force by occupation in industrial, governmental and commercial sectors.
3) Estimation of the health service capacity in public and private sectors.
4) Estimation of the capacity of the education services in public and private sectors.

Mostly, the algorithm to transpose a socio-economic parameter at a national scale as a percentage of the total population to the grid cell using the population as proxy corresponds to

$$\mathrm{SEI}(x,y)=\frac{\mathrm{SEI(adm)}}{\sum_{\mathrm{(admn)}}\mathrm{Pop}(x,y)}\times \mathrm{Pop}(x,y)$$

where SEI(*x*,*y*) is the population per socio-economic parameter per cell at *x*,*y* position; SEI(adm) corresponds to the population per socio-economic parameter per administrative unit (general country level) Pop(*x*,*y*) represents the population living in the cell (extracted from LandScan).

After this operation, each cell represents the number of exposed persons for each of the eleven socioeconomic sectors (income, labor, instruction, and health) in a portion of an urban area.

The whole resident population corresponds to the sum of the four income classes.

Non-residents include workers in primary and secondary sectors, students, represented by the sum of peoples included in the labor and education sectors.

3.3.4. Step 4: Distribution of the capital stock per socioeconomic sectors

This downscaling process consists in transferring the produced capital from the administrative unit (country) to the 5 × 5 grid cells using the already estimated population per cell and sector as the main proxy. In order to refine the process and keep a more realist snapshot of the distribution of economic pattern of the country, the results were further weighted by two different sets of variables:

1) The Gross Regional Product (GRP) per capita.
2) The unitary values at national level and by socio-economic sector.

The GRP capita were derived by dividing the already mentioned raster of GDP distribution by the gridded population.

Essentially the process involving GRP consists in evaluating a coefficient of variation between the national values of GDP (capita) and those at subnational (regional) level. In other words the coefficient indicates how much a cell will differs from the national average of GDP. The coefficient of variation has been calculated using the following equation:

$$\mathrm{GDPcv\%}(x,y)=\frac{\mathrm{GRPc}(x,y)}{\mathrm{GDPc(adm)}}\times 100$$

where GDPcv%(*x*,*y*) is the coefficient of variation for the cell located at *x*,*y* coordinates, GDPc(*x*,*y*) the GDP per capita per cell and GRPc(adm) the value of Gross regional Product (GDP at sub-national level) per capita.

The unitary values give an indication about cost and surface of the built-up surface related to the building usage; they have been processed by CIMNE on the bases of data coming from the Global Construction Cost and Reference Yearbook [31].

The unitary values are successively used as weight factor and integrated in the process: basically they consist of an evaluation per country and socioeconomic sector of the building surface and its unitary cost. It is important to underline, that surface and unitary costs, are relative values across the country; and they have to be interpreted as factors used only, to differentiate the surface and its cost from one socioeconomic class to another. In other words for a determinate country a surface value of 2 for low income buildings and 8 for high income, only means that the latter have surfaces four times those of lower income.

The produced capital is downscaled to the cell level using population data and then multiplied by the already calculated coefficient GDPcv%. This will result in a database that moves from a pure population distribution to a goods type distribution (population per use buildings) according to the following equation:

$$\mathrm{PCSEI}(x,y)=[\mathrm{PCc(adm)}]\times \mathrm{PopSEI}(x,y)]\times \mathrm{GDPcv\%}(x,y)\times \mathrm{UVs}$$

where PCSEI(*x*,*y*) is the produced capital at cell size per seI, PCc(adm) is the produced capital per capita at national level, PopSEI(*x*,*y*) the population of the cell per sector, GDPcv%(*x*,*y*) the coefficient of variation for the cell located at *x*,*y* coordinates, and UVs the coefficient related to unitary costs/surfaces related to the socioeconomic sector.

3.3.5. Step 5: Distribution of the capital stock per socioeconomic sector and building type

Once the population and capital stock are estimated for residential and non-residential usage, it is necessary to distribute them in the different building classes present in the country, integrating the information provided by WAPMERR that contains the population distribution by level of complexity and by building type (or structural system). This last operation is simply performed by multiplying the already calculated sectorial capital stock (PCSEI)

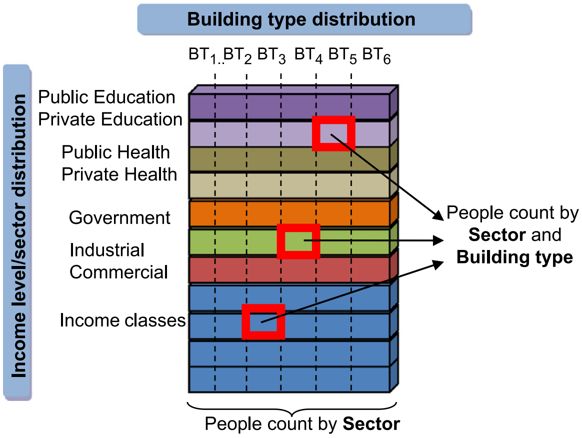

Fig. 4. People distribution into building types and SEI.

per capita by the corresponding population living in a certain structural system.

Finally, after this last step, each record will correspond to a building type, based on level of income/sector, specific to a localization of 5 km × 5 km of an urban area with an associated level of complexity Fig. 4.

4. Results

The main objective of GEG-2013 was to generate a global evaluation of exposed assets, in urban areas, in order to provide a specific exposure input data to be used in the CAPRA platform. The minimum information required and included in the GEG-2013 outputs comprises the five essential elements necessary for further probabilistic risk assessments:

- ID.
- Geographic location.
- Construction type for vulnerability classification.
- Exposed economic value.
- Human occupation.

This information is available for all the 216 countries included in the GAR 2013 process.

Each record (exposed value) in the GEG-2013 represents a certain building structural type of certain income level/sector in a certain urban area with a special point representation in the centroid of the 5 × 5 cell.

Moreover the GEG-2013 is an (off-line) geo-database developed using open source PostGreSQL/PostGis software. The database supports several kinds of queries including extraction, aggregation/disaggregation of both GIS and tabular data. The GEG-2013 also includes a global map (GIS format) of urban population and produced capital stock at a resolution of $5 \times 5\ \text{km}^2$ per residents and non-residents subdivided into eleven socio-economic sectors.

Since mid-May 2013 the exposure map is publicly available on the PREVIEW Global Risk Data Platform: http://preview.grid.unep.ch.

5. Discussions

The GEG-2013 was used in the GAR 2013 Global Risk Assessment on earthquakes and tropical cyclones [10]. The results were reviewed by a wide range of experts. This allowed the first characterization of Global Risk using a probabilistic approach. The format and resolution appears to be appropriate for the scope. The $5 \times 5\ \text{km}^2$ grid covers the need for those hazards that are affecting wide areas. The relative small weight of the dataset was well-suited for the multiple scenarios which were run for the probabilistic analysis. However, the dataset would need to be improved for hazards which are affecting smaller areas (e.g. landslides, tsunamis, floods).

GEG-2013 is limited in some important ways:

- It was fundamentally constructed using national indicators that were successively disaggregated onto a 5 × 5 grid. Several countries show important regional variations in terms of wealth, employment, sanitation, instruction, that are not captured in GEG-2013. This disadvantage is partially attenuated, at least for wealth by using a regional distribution raster of GRP as weight when distributing the national capital stock.
- Information concerning the unitary values (costs, and surfaces) are only related and consequently distributed according to the sectorial type of buildings/dwellings. Such kind of data will be strongly suitable for building typologies too.
- The capital stock in each cell is distributed on the basis of the number of persons living in that cell and does not take into account the real value of the assets of the cell. In this case a partial improvement is done by using a factor (unitary costs and surfaces) during the calculations of capital stock for each asset. In order to decrease this effect we propose two possibilities for the future(s) edition of GEG. The first is based on the integration of the detailed information on absolute unitary costs, surfaces and number of buildings (level 0 and 1) derived from the GED4GEMS when publicly available. The second will consist in using a new population independent GDP raster to weight the capital stock (Section 3.3.4). This approach will be used in the GEG for GAR2015.
- The raster of population distribution, LandScan, remains one of the most “fundamental pieces” used in the whole process of GEG-2013, but we have to keep in mind that, it has two main non-negligible, inconveniences.

The in-depth methodology and complete algorithms are not published, at present.

Its license does not allow it to be re-distributed, as all the other GAR datasets are, to the public domain.

With the aim to refine the final thematic (and eventually spatial) resolution of GEG, we suggest the systematic use of sub-national (regional) statistical datasets, at least for those countries that present major geographical variations in their socio-economic structure. Unfortunately at the global scale they are very scarce, with the exception of population (GPW-SEDAC) [12] and GDP (GRP, World Bank) [28]. Other data exist but with limited geographic coverage to specific group of countries as in the case of Eurostat, FAO coutrySTAT, or OECD.

Together with a classical top down approach, a gradual integration of some specific information from the bottom is advisable. This could be the case of road densities and power plant/industrial sites localization.

The forthcoming release of high resolution dataset in the public domain opens very interesting new perspective for the development of a future generation of exposure databases.

Several new high resolution essential datasets will be released in the public domain in the foreseeable future; they will include raster of population distribution at 100 m resolution with the WorldPop products [26] a very high resolution imagery for delimiting building footprints (GHSL, and AR imagery). Moreover a new global land cover map at 30 m resolution [32] has just been released.

Crowdsourcing effort will also play a more important role in terms of critical facilities inventories

(*OpenStreetMap* [33]), mapping (*Crowdmap* [34]), validation of datasets as in the case of land cover with *Geo-Wiki* [35], but also in the domain of disaster mapping as supported by USGS initiatives such as *Did You Feel It?* [36] and *Tweet Earthquake Dispatch* [37] (earthquakes) or *Did You See It?* [36] (landslides). Through the *National Map* [38], users can give detailed information about buildings and other manmade structures, unfortunately at present time is only available for US.

Together with this upcoming propagation of global high resolution datasets, a serious effort to standardize, harmonize and expand the existing local “bottom-up approach” datasets, could lead to the future generation of exposure datasets at a global scale.

6. Conclusions

The methodology described in this paper provides a framework for creating an open global sub national level (5×5 grid) exposure database, suitable for probabilistic risk assessment in urban areas at country level using the Capra platform (http://www.ecapra.org).

The exposure is represented as a value of a group of buildings in each $5 \times 5\ \mathrm{km}^2$ cell. Four socio-economic sectors were used for estimating the characteristics of the buildings at sub-national levels in accordance with the size of each urban area. The distribution of building type is related with the population that lives in each typology and not with the number of buildings in each one of the construction types. The produced capital (assets value) is distributed for each sector and building type per each cell of the $5 \times 5\ \mathrm{km}^2$ grid, according to the relative number of people living inside, also taking into account two supplementary elements: the occupation density and the unitary cost per sector.

The development of GEG-2013 is based on several assumptions and shows different points that can be improved (see Section 5) especially if new data become available. Its main purpose is to serve the realization of the “probabilistic risk estimation at the global scale”. The risk estimation is presented in relative terms, in the form of risk classes, and its purpose is to give an order of magnitude of the potential economic losses, in the perspective of a comparison between countries (GAR 2013). The consistency of the methodological approach used in the development of GEG, as well as the choice of the best data currently available for its implementation, have produced a product fully adapted to the needs of the global model of the evaluation of probabilistic risk.

Acknowledgments

This work has been funded by UN-ISDR. Andrew Maskrey and Manuela di Mauro at UN-ISDR are thanked for coordinating the whole process. We thank CIMNE for providing guidelines for the development process. PwC furnished an overview on economic aspects together with methodologies to evaluate missing urban produced capital data; IIASA provided supplementary data on urban produced capital, thanks to Pascal Peduzzi for the critical review, and Ruth Harding for substantial English improvement.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ijdrr.2014.05.008. These data include Google map of the most important areas described in this article.

References

[1] UNISDR. Global assessment report on disaster risk reduction: risk and poverty in a changing climate. United Nations International Strategy for Disaster Reduction. Geneva, Switzerland; 2009.

[2] UNISDR. Global assessment report on disaster risk reduction: from shared risk to shared value: the business case for disaster risk reduction. United Nations International Strategy for Disaster Reduction. Geneva, Switzerland; 2013.

[3] Schneider PJ, Schauer BA. HAZUS – its development and its future. Nat Hazards Rev 2006;7(2):40–4.

[4] Jaiswal K, Wald D, Porter K. A global building inventory for earthquake loss estimation and risk management. Earthq Spectra 2010;26:731–48.

[5] GED4GEM – Exposure Database 〈http://www.globalquakemodel.org/what/physical-integrated-risk/exposure-database/〉.

[6] Dell'Acqua F, Gamba P, Jaiswal K. Spatial aspects of building and population exposure data and their implications for global earthquake exposure modeling. Nat Hazards 2013;68:1291–309.

[7] Gunasekera R, Ishizawa O, Pandey O, Saito K. A globally consistent methodology for an exposure model for natural catastrophe risk assessment 2013;15.

[8] Open Data for Resilience Initiative (OpenDRI)|GFDRR 〈https://www.gfdrr.org/opendri〉.

[9] Kull D., Saito K., Soden R., Bacca A. World Bank-GFDRR activities on exposure mapping-modelling. 〈http://www.slideshare.net/GRFDavos/world-bankgfdrr-contributions-to-exposure-modeling-for-global-risk-modeling-initiatives-and-opendri-initiative〉; 2012.

[10] Omar-Darío Cardona, Mario G. Ordaz, Miguel G. Mora, Mario A. Salgado-Gálvez, Gabriel A. Bernal, Daniela Zuloaga-Romero, et al. Global risk assessment: A fully probabilistic seismic and tropical cyclone wind risk assessment, Int J Disaster Risk Reduction, Available online 26 May 2014, ISSN 2212-4209, http://dx.doi.org/10.1016/j.ijdrr.2014.05.006. (http://www.sciencedirect.com/science/article/pii/S2212420914000454).

[11] Schneider A, Friedl MA, Potere D. Mapping global urban areas using MODIS 500-m data: new methods and datasets based on urban ecoregions. Remote Sens Environ 2010;114:1733–46.

[12] SEDAC (Socioeconomic Data and Applications Center). Global Rural-Urban Mapping Project (GRUMP). Palisades, NY 10964 USA; 2012.

[13] Schneider A, Friedl MA, Potere D. A new map of global urban extent from MODIS satellite data. Environ Res Lett 2009;4:044003.

[14] Wyss M, Tolis S, Rosset P, Pacchiani F. Approximate model for worldwide building stock in three size categories of settlements. Geneva, Switzerland: ISDR; 2013 (Technical report prepared for GAR 2013).

[15] Satterthwaite D, Owen DL. Outside the large cities: the demographic importance of small urban centres and large villages in Africa, Asia and Latin America. London: IIED; 2006.

[16] CIMNE (Centro Internacional de Métodos Numéricos en Ingeniería), EAI, INGENIAR and ITEC. Probabilistic modeling of natural risks at the global level: global risk model. Geneva, Switzerland: UNISDR; 2013.

[17] Washington, DC: World Bank; 2011.

[18] World Bank's annual World Development Report (WDR). 〈http://wdronline.worldbank.org/worldbank/a/incomelevel〉.

[19] World Bank. Where is the wealth of nations? Measuring capital for the 21st century. Washington, DC: The World Bank; 2006.

[20] PwC (PricewaterhouseCoopers). UNISDR and PwC – Working together to reduce disaster risks. Geneva, Switzerland: ISDR; 2013.

[21] United Nations University. Inclusive Wealth Report 2012: measuring progress toward sustainability. Cambridge: Cambridge University Press; 2012.

[22] OECD. Land valuation in the National Accounts questionnaire; 2011.

[23] Bright EA. LandScan Global Population 1998 Database; 2002.

[24] Bhaduri B, Bright E, Coleman P, Dobson J. LandScan: locating people is what matters. Geoinformatics 2002;5:34–7.

[25] LandScanTM global population distribution data (Raster dataset). US: Oak Ridge National Laboratory, U.S. Department of Energy; 2007.
[26] Tatem AJ, Adamo S, Bharti N, Burgert CR, Castro M, Dorelien A, et al. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. Popul Health Metr 2012;10:8.
[27] Peduzzi P, Chatenoux B, Dao H, De Bono A, Herold C, Kossin J, et al. Global trends in tropical cyclone risk. Nat Clim Change 2012;2:289–94.
[28] Global Risk Data Platform: GDP. 〈http://preview.grid.unep.ch/index.php?preview=data&events=socec&evcat=1&lang=eng〉; 2010.
[29] De Bono A. The Global Exposure Database for GAR 2013. Geneva, Switzerland: UNISDR; 2013.
[30] European Union. Cities of tomorrow: challenges, visions, ways forward. European Commission, Directorate General for Regional Policy Unit C.2; 2011.
[31] Compass International Inc. Global construction cost and reference yearbook; 2012.
[32] Gong P, Wang J, Yu L, Zhao Y, Zhao Y, Liang L, et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. Int J Remote Sens 2013;34:2607–54.
[33] OpenStreetMap. 〈http://www.openstreetmap.org〉.
[34] Crowdmap. 〈https://crowdmap.com/〉.
[35] Welcome to Geo-Wiki Project. 〈http://www.geo-wiki.org/〉.
[36] DYFI Background – The Science Behind the Maps. 〈http://earthquake.usgs.gov/research/dyfi/〉.
[37] Tweet Earthquake Dispatch. 〈http://earthquake.usgs.gov/earthquakes/ted/〉.
[38] The National Map. 〈http://nationalmap.gov/〉.