Remote Sensing of Environment 187 (2016) 230–240

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Consistent yet adaptive global geospatial identification of urban–rural patterns: The *iURBAN* model

Christoph Aubrecht [b,*], Rashmin Gunasekera [a], Joachim Ungar [a], Oscar Ishizawa [a]

[a] The World Bank - Social, Urban, Rural & Resilience (GSURR), 1850 I St. NW, Washington, DC 20006, USA
[b] European Space Agency (ESA-ESRIN), Seconded to the World Bank, 1818 H St. NW, Washington, DC 20433, United States

A R T I C L E I N F O

Article history:
Received 14 February 2016
Received in revised form 18 August 2016
Accepted 13 October 2016
Available online xxxx

Keywords:
Urban–rural
High-resolution built-up area data
Population reallocation
Disaggregation
Global scale
Geospatial modeling
Adaptive

A B S T R A C T

The main motivation of this paper is to shed new light on the problem of spatial identification of urban and rural areas globally, and to provide a compatible disaggregation framework for linking associated country-specific, non-spatial data compilations, such as building type inventories. Existing homogeneously set-up global urban extent models commonly ignore local-level specifics. While global consistency and regional comparability of urban characteristics are much strived-for goals in the global development and remote sensing communities, non-conformity at the national level often renders such models inapplicable for effective decision-making. Furthermore, the focus on identifying 'urban' leads to an ill-defined 'rural', which is simply defined by contrast as 'everything else'; a questionable definition when referring to strongly spatially localized residential patterns. In this paper we introduce the novel *iURBAN* geospatial modeling approach, **i**dentifying **U**rban–**R**ural patterns in **B**uilt-up-**A**djusted and **N**ationally-adaptive manner. The model operates at global scale, but at the same time conforms to country specifics. In this model, high-resolution, satellite-derived, built-up data is used to consistently detect global human settlements at unprecedented spatial detail. In combination with global gridded population data, and with reference to national level statistical information on urban population ratios globally compiled in the annually-released UN World Urbanization Prospects, *iURBAN* identifies matching urban extents. Additionally, a novel reallocation algorithm is introduced which addresses the poor representation of rural areas that is inherent in existing global population grids. Associating all of the population with inhabitable, built-up area and conforming to national urban–rural ratios, *iURBAN* sets a new standard by enabling careful consideration of both urban and rural as opposed to traditional urban-biased approaches.

1. Introduction

1.1. Defining urban

What is urban? This is a seemingly simple question for which there is, however, no simple answer. When analyzing different geographical regions urban characteristics materialize very differently in different parts of the world. Looking at urban and rural as a holistic concept, that concept clearly addresses the human environment. Both urban and rural fundamentally refer to certain geographical areas which are shaped by human activities; in this context, urban points to nonagricultural tasks (Weeks, 2010) and focuses strongly on the built environment.

Although people at a local level usually find it entirely self-evident as to whether the place in which they live is considered urban or rural, there is no international agreement of how urban areas are defined. In defining urban spaces, national governments first aim at best matching their own situational circumstances and internal needs. The United Nation's biannually-released World Urbanization Prospects (WUP), in its latest edition (UN-DESA, 2015), compiles national definitions of urban as defined by >230 countries in statistical terms. More than half of these use a variety of minimum population size and density thresholds to identify areas as urban. Further criteria utilized refer to social, economic, and functional characteristics and services, such as water supply, the sewerage system, and access to electricity (Aubrecht et al., 2014). In line with the WUP, the World Bank's World Development Indicators (WDI) database annually reports the urban proportion of population at the national level for every country. These global data compilations providing aggregated urban statistics are widely referred to for illustrating global trends. One of the most popular conclusions referring to those statistics points at the year 2008, when for the first time in history the 50% urban population milestone was crossed according to WUP. That means the majority of the human population are now living in what is considered urban environments. The latest numbers published in WUP-2014 indicate a 54% current global urban ratio, with a projected increase to 66% by 2050 (UN-DESA, 2015). The virtual crossing of the 50% milestone in 2008 has prominently been proclaimed

* Corresponding author. Tel.: +1 202 473 5417.
E-mail address: caubrecht@worldbank.org (C. Aubrecht).

http://dx.doi.org/10.1016/j.rse.2016.10.031
0034-4257/

the start of the 'urban age' (Katz et al., 2006; Kreibich, 2010), or the 'rise of *homo urbanis*' (Saunders, 2010). While such constructed milestones compellingly attract public and political attention, it must always be kept in mind that in fact incomparable and often incompatible definitions of urban are used for those types of aggregated statistics. The 'urban age' thesis has even been referred to as a "flawed basis on which to conceptualize world urbanization patterns, an empirically untenable statistical artifact and a theoretically incoherent chaotic conception" (Brenner and Schmid, 2014).

Human activities and population obviously play a central role in characterizing urban areas, and all these activities happen at certain identifiable locations. Weeks (2011) defined urban as "a characteristic of a place, rather than of people". A spatial approach to identifying and describing urban can therefore be considered indispensable. With respect to spatial parameters for urban extent identification, basic administrative criteria are still prominently used by >50% of the WUP-listed countries in making the distinction between urban and rural; for approximately 30%, it is even the sole criteria (UN-DESA, 2015). Attempting to circumvent the locally-biased and heterogeneous nature of administrative boundaries, urban researchers and the international development community have been active in trying to find a globally homogeneous and commonly accepted definition of urban by using spatially-explicit criteria independent of administrative parameters (see Potere et al., 2009; Gamba and Herold, 2009 for a comparative overview). Global consistency, and regional as well as international comparability, are major goals in better understanding and addressing cross-boundary urbanization issues, and enabling corresponding global policy development. The main challenge therefore is the different understanding of 'urban' in different parts of the world. Current global approaches to a homogeneous definition of urban extents mostly ignore local specifics for the sake of methodological consistency.

1.2. Spatial delineation of urban areas

There are several approaches described in the literature, addressing spatial delineation of urban areas from different perspectives and using different kinds of input data. It is essential to give a detailed overview of these approaches at the beginning to highlight methodological gaps and thereby illustrate the need for the novel model approach introduced in this paper.

Remote sensing data and derived products have played an important role in guaranteeing global methodological consistency from a data acquisition perspective. Satellite technology applied to observation of the Earth surface enables the detection of basic physical parameters of reflectance or emission. These physical parameters consequently need to be transformed into meaningful information about objects on the ground. In the identification of the distribution patterns of human activity, two features have been most prominently referred to; one being artificial light at night, and the other being built-up area.

Satellite-observed nighttime lights data was first identified as a potential proxy measure for human activity in the early 1970s (Croft, 1973), when this 'hidden' source of information was discovered in imagery from the Operational Linescan System (OLS) onboard the Defense Meteorological Satellite Program (DMSP). The DMSP platform was initially designed to monitor cloud cover and cloud temperatures, but due to its unique capability to detect low visible light levels, it was found that the sensor also records light at night from artificial and anthropogenic sources such as human settlements, gas flares, and various combustion sources. In those early studies by Croft, it was already stated that "inhabited areas are clearly outlined" (Croft, 1973). Subsequent research then found strong correlation between the illuminated areas and census population distribution data (Welch, 1980). However, it was not until the late 1990s that nightlights data from the (by then) digital DMSP-OLS archive was first used to specifically delineate urban areas in spatially-explicit form (Imhoff et al., 1997). That study investigated applying uniform thresholds to data on cumulative percentage of light occurrence in a multi-month OLS composite product for the continental United States, to convert lights into 'urban cover'. Different thresholds empirically derived to extract urban extents of different metropolitan areas were averaged to come up with one country-wide value (89% cumulative percentage of light occurrence), and overall good agreement was found at the national level compared to urban metrics reported in census data. At the state level, however, large variations of mismatch were observed (i.e. false positives of wrongly identified urban areas and false negatives of unidentified urban areas), questioning its applicability at coarser spatial scales. Further studies over the years tested thresholding stable nighttime lights data for urban delineation at the global scale, identifying different thresholds in different geographical settings. The main conclusion was that there is no single threshold applicable for different countries, or even for different cities within one country (Henderson et al., 2003; Small et al., 2005; Sutton et al., 2010). Despite these findings of methodological inapplicability at regional scale, nighttime lights data have been heavily used as proxy measures for urban areas. Early studies include the GLC 2000 Database (Global Land Cover) (Bartholome et al., 2002) as well as the Global Rural–urban Mapping Project (GRUMP) that uses primarily DMSP-OLS-derived urban extents in combination with ancillary data, such as populated places from the Digital Chart of the World (DCW) for refined population disaggregation purposes (Balk et al., 2005).

Artificial surface or built-up area is an obvious alternative proxy feature in the delineation of urban areas. Following up on the early global modeling efforts of GLC 2000, more recent global land cover products, such as the European Space Agency's GlobCover (Arino et al., 2007) and CCI Land Cover (Bontemps et al., 2015), analyze artificial surface area in optical medium resolution data (e.g. 300 m ENVISAT-MERIS satellite imagery) as an alternative to night lights for the classification of associated urban areas. Other notable efforts of global built-up area mapping include Boston University's (BU)-MODIS approach, based primarily on 1 km MODIS data from NASA's Terra/Aqua satellites (Schneider et al., 2003). Due to spectral similarities between built-up and bare soil at the 1 km level, ancillary data sources had to be used to define urban, including DMSP nightlights and population density from the Gridded Population of the World (GPW) product. Subsequent efforts using higher-resolution (500 m) MODIS data as exclusive data source for urban extent delineation resulted in an urban mask dataset (Schneider et al., 2010), that has widely been considered the most detailed available option, and which thus became the global standard for the scientific modeling community. Most recent developments have further refined the scale of analysis in terms of spatial resolution, for example using MODIS-250 m data (Mertes et al., 2015). At another order of magnitude higher spatial resolution, the Chinese GLC-30 mapping project resulted in a global 30 m product, extracted from Landsat imagery (Chen et al., 2015).

With the earlier BU-MODIS efforts already hinting at the potential value of joint use of built-up and population density measures for refined urban area identification, gridded population density data has since become a much-used, additional input data source. In particular, the regularly updated LandScan data set has been widely implemented in multiple application contexts since its first release that illustrated global population distribution patterns for the year 1998 (Dobson et al., 2000). While there have been various attempts to use gridded population density data as sole input for areal identification of urban patterns (Gallego, 2004; Dijkstra and Poelman, 2014; Christenson et al., 2014a), other studies integrate population with land cover and specific built-up parameters. The global Geopolis database, for example, defines urban agglomerations based on satellite-derived contiguous built-up areas containing a population of >10,000 people (Moriconi-Ebrard et al., 2008). This is one of the first approaches where built-up contiguity is explicitly mentioned as a defining factor. In that study, population numbers are derived from sub-national local administrative census units. Contiguity is also accounted for as a parameter of urban-ness in a joint effort of European Commission (EC) and OECD (Dijkstra and

Poelman, 2014) where gridded population is clustered (in terms of high-density cells), as opposed to identifying contiguous built-up area. A cutoff-point of 5000 people is thereby defined as the threshold for urban clusters in developed countries (Europe, United States, Canada, Japan). Christenson et al. (2014b) perform a similar analysis for a case study in Africa (Nigeria), although their study applies a much higher threshold (20,000 people), being subject to the respective national statistical definition. Focusing on the spatial extraction of the urban share of a country's total population, various integrations have been proposed for extracting LandScan population numbers from spatially pre-defined urban areas, as identified via nighttime lights (Salvatore et al., 2005), MODIS-500 (de Bono and Mora, 2014), the more recently released built-up reference layer (BUREF) (de Bono and Chatenoux, 2014), or fusion of several of those (Kasimu and Tateishi, 2008).

A limited number of studies have attempted to include further parameters for wide-scale urban identification modeling, in particular focusing on dynamic aspects as opposed to the commonly-used static residential patterns and physical attributes. In an attempt to conceptually extend the above-cited EC-OECD approach, commuting information was integrated as an additional factor, resulting in a notion of functional urban areas (OECD, 2012). Travel time was also one major input variable in the setup of the World Bank's so-called agglomeration index, proposing an alternative measure of urban concentration (Uchida and Nelson, 2008). Furthermore, another study presented the extraction of urban footprints by applying population density thresholds to a dynamic population distribution grid for the United States that accounted for daytime work fluctuations (Burian et al., 2006).

Using one globally uniform definition of urban has its benefits, especially in terms of efficiency, transparency, and comparability across regions. However, as outlined above, basic socio-economic and urban composition vary significantly from region to region, in particular between the developed and the developing world. It has therefore been largely concluded that applying fixed thresholds at the global scale, be it to nightlights, built-up, or population distribution data, is not appropriate if applicability at national and local level needs to be ensured. Criticism on the 'urban age thesis' and on taking the United Nations' compiled urban statistics as a given without questioning their composition is entirely justified, legitimate, and understandable, when approaching from a global perspective. Nonetheless, at the national level, the respective individual national definitions are likely to represent the urban character of a country better than a globally imposed standard. Salvatore et al. (2005) highlight that whether to use nationally-defined statistical figures on urban–rural population counts or uniform globally-modeled approximations depends on the objectives of a specific research application. That study was the first in trying to combine global spatial and national statistical approaches by experimenting on iterative thresholding of nighttime lights intensity data to reach compliancy with national WUP-released figures. It was concluded that the approach seemed too simplistic, with the resulting urban mask being very fragmented. More recent research (Christenson et al., 2014a) builds upon that initial concept, replacing nightlights as spatial basis with LandScan population distribution data.

Residential population is directly associated with settlement structures. Built-up area approximating building structures has therefore proven to be the main spatial linking variable. Novel global high-resolution built-up data, such as the German Aerospace Center (DLR)'s Global Urban Footprint (GUF) (Esch et al., 2012), or the Global Human Settlement Layer (GHSL) (Pesaresi et al., 2013) of the European Commission's Joint Research Centre (JRC), enable for the first time a globally-consistent look not only at urban agglomerations, but also at the flipside of urban. The identification of previously undetected small settlements in rural regions thus allows a conceptual refinement of the urban–rural spectrum, and a move from a binary 'urban vs. rural' to a more spatially-explicit 'urban and rural'. Spatial identification of settlements is, however, only one step towards urban–rural pattern characterization. Population figures need to be associated with those settlements for further considerations (see for example, Freire et al., 2015). Notably, existing global population grids are not designed or calibrated for the analysis of rural settlements. The evaluation of sub-national figures on urban–rural proportions and cross-validation with existing population grids' spatial allocation within pre-identified built-up area, reveals a clear and statistically significant correlation of the degree of urbanness of a certain region and how many people existing global distribution models allocate to built-up area in the disaggregation process (see Section 2.2 for further detail on the related population-urban-built-up regression model). The identified strong linear correlation leads to the conclusion that while existing global population grids capture urban population well within built-up area, rural population patterns are unreliable and vague in terms of their explicit spatial distribution. More precisely, total population numbers add up correctly at the administrative base unit level used as disaggregation input, but population is dispersed widely outside of built-up area in rural regions. The main reason is that for the set-up of current population distribution models the type of high-resolution settlement information described above has not been available at the global scale as disaggregation input. Current standards, such as the earlier-outlined urban extents derived from MODIS-500 or DMSP nighttime lights, operate at an order of magnitude different scale levels, and are thus much less sensitive to detection of smaller built-up areas.

1.3. Introducing the iURBAN model

In this paper, we present a novel geospatial modeling approach - *iURBAN* - **i**dentifying **U**rban–**R**ural patterns in **B**uilt-up-**A**djusted and **N**ationally-adaptive manner. The model has been developed under the framework of the World Bank's Country Disaster Risk Profiles (CDRP) project implemented in Central America and the Caribbean (Gunasekera et al., 2015) under the Latin America and Caribbean Region Probabilistic Risk Assessment (CAPRA) Program. In that context, *iURBAN* provides the geospatial basis for the building stock exposure model which is consequently integrated with hazard and vulnerability modules for probabilistic risk and loss assessment. In several aspects, the model is novel in design and provides a refined perspective on the definitions and implementations outlined above. The most recent high resolution GUF built-up area product, as derived from the TanDEM-X radar mission (Esch et al., 2012), has been provided for the CDRP project prior to public release, in an internal collaboration agreement with DLR. In a first step, GUF serves as spatial basis to accurately delimit human settlements from uninhabited land. For classification of those detected settlements in urban and rural, GUF built-up contiguity and built-up density are used as contributing variables aside from LandScan population distribution. In a composite index approach, the degree of urban-ness is defined for each contiguous built-up patch. Patches are then ranked and iteratively added up until the patch-inherent population numbers match the national urban share, as provided in WUP. In order to address the weakness of current global population grids in accurately representing rural distribution patterns, the rural population is reallocated to settlements identified as rural in a counter-weighted, spatial disaggregation approach. The final *iURBAN* output is consequently fully compliant with WUP, not only with regard to its urban ratio, but also in its identification and allocation of rural population to actual inhabited land.

2. Materials and methods

2.1. Data

The Global Urban Footprint (GUF) product is used as built-up data. DLR provided a preliminary binary mask at 75 m resolution for CDRP implementation. GUF is derived from TanDEM-X Radar imagery, which implies significant advantage with regard to distinguishing built structures from flat terrain, compared to optical satellite imagery.

LandScan (v2012) is used as population data. Compared to other existing global population grids, LandScan has a strong focus on urban areas that shows in clear peak density patterns. Annual urban ratios and national population counts and projections are taken from UN-WUP. While perhaps not comparable globally, the use of national definitions provides a consistency in terms of local-level applicability. *iURBAN* consequently follows the same path, i.e. applying a global methodological design while keeping national plausibility.

2.2. Population-urban-built-up regression model

The hypothesis that existing global population grids, such as LandScan, spatially allocate population well to built-up area (as identified in high-resolution GUF data) in urban areas, but fail to do so in rural areas, is one of the main underlying assumptions of the *iURBAN* model (see Section 2.3 below). Based on that assumption, LandScan population numbers are not altered in identified urban areas, whereas a population reallocation model is proposed for rural areas. The hypothesis is grounded in the fact that high-resolution built-up data (such as the used GUF) has to date not been accessible for global population modeling efforts. More coarse-scale, publicly available built-up data (such as MODIS or nighttime lights data) has therefore been used, which leads to a vast under-representation of small rural settlements that are not detected in these data sets. Larger urban agglomerations, on the other hand, are also well identified in coarse-scale, built-up data products, and are therefore reasonably represented in existing population disaggregation models (such as LandScan).

Qualitatively, the hypothesis is outlined in a straightforward manner. For quantitative evaluation, however, the data correlation patterns need to be statistically analyzed. We consequently perform a weighted linear regression referring to the following parameters. GUF built-up data is aggregated from the initially available 75 m resolution to a 30 arc-sec grid (hence called built-up mask) in order to be spatially compatible with LandScan. For each province in every Central American country, the share of LandScan population counts that is allocated within that built-up mask is calculated. That relative share is labeled the capture ratio (i.e. the share of a province's population captured by LandScan within the GUF built-up mask). The assumption is that the more urban a province, the higher the capture ratio should be. The provinces' degree of urban-ness (urban ratio) is derived from the most recent sub-national census data for each country. In a basic linear regression, R^2 values vary from country to country between 0.75 (El Salvador) and 0.95 (Panama), with an overall value of 0.78 determined for Central America. Basic linear regression, however, does not represent the hypothesis in appropriate terms. It basically implies that every province is equally weighted, independent of its population size. For example, the province Emberá (Panama) with just over 10,000 inhabitants has the same statistical influence on the R^2 estimation as the province of Guatemala City (Guatemala), with almost 3.2 million inhabitants. To account for that inequality (given that more densely populated provinces are almost certainly more urban), we introduce population size as a weighting factor in the regression analysis. In this weighted linear regression model, each of the two parameters (capture ratio and urban ratio) is multiplied with the square root of the respective province's total population. The resulting R^2 value for Central America is 0.95, thus showing a highly significant correlation and quantitatively supporting the above-stated hypothesis, and accordingly, backing the *iURBAN* approach presented in this paper.

It should be noted that the result, while clearly statistically relevant in the current setting, could in fact be influenced by varying data acquisition times. The GUF built-up data represents the situation in 2012 (when the Radar imagery serving as input was acquired). Latest census collection years in Central America (and thus the used urban ratio values) vary in a range between 2002 (Guatemala) and 2011 (Costa Rica). LandScan, in turn, while not documented, also seems to represent different years (based on comparison of aggregated country totals to annual national projections as reported in WUP), ranging from 2006 (El Salvador) to 2014 (Honduras). For statistical analysis, it would be optimal to have all data sets harmonized to a 2012 reference year. However, this is not possible with sub-national urban ratio projections not being available annually, and LandScan not representing one homogeneous time stamp. The data used in this study is therefore considered the best input available.

2.3. iURBAN model assumptions

Assuming that, as outlined above, the locational population information from LandScan is reasonably accurate in urban areas, these urban areas need to be identified at the grid level. Population numbers can be maintained as input for further modeling steps, while the more 'rural' an area, the less 'trust-worthy' the population numbers are considered. In order to measure the urban-ness of an area at a 15 arc-sec grid scale, two assumptions are made:

(1) the larger a contiguous built-up patch, the more likely it is to be considered urban (larger in terms of the summed-up area of individual high-resolution built-up patches detected in the 75 m GUF data), and
(2) the greater a population (in absolute terms) clustered within a contiguous built-up patch, the more likely that patch is to be considered urban.

After a 30 arc-sec grid (corresponding to the final CDRP output cell size) was initially tested, a 15 arc-sec grid was eventually chosen as spatial modeling unit, thereby following Week's conclusion that such a grid size (i.e. around 500 m cells) would ensure proper understanding of "variability both between and within human settlements" (Weeks, 2010).

2.4. iURBAN model steps

In a first step, the nationally aggregated LandScan sum is compared to WUP country totals. There is no documentation on which population input data LandScan uses for a specific country, thus the reference year is also unclear. In other words, taking LandScan v2012 as an example, that version number does not necessarily indicate that population distribution is consistently shown for the year 2012. LandScan country data are updated following specific internal needs and priorities that are not revealed to the external user. For urbanization studies, this practice makes application of the data product difficult. In *iURBAN*, the respective reference year per country is identified by comparing population totals to the annual WUP figures, i.e. the year for which the WUP value and nationally aggregated LandScan value best match.

For the identification of urban areas, a composite, normalized weighting index is introduced, weighting every contiguous built-up patch in the 15 arc-sec grid, accounting for its:

(1) built-up area (inherent summed-up patches), and
(2) absolute population count.

Both indicators are normalized to a 0–1 range; the patch with the highest built-up area value within one country is assigned 1. Multiplicative composition is applied for the index in order to put stronger focus on particularly high values of individual indicators. The built-up patches are then ordered according to their index value and iteratively added up using their respective population values until the national-level urban–rural split (according to WUP/WDI) is reached. The same reference year is thereby taken for the urban ratio as identified before in LandScan for a specific country. The selected patches are consequently classified as urban. The remaining contiguous built-up patches are considered rural.

Fig. 1 shows a fictitious 'country' with several contiguous built-up patches. Table 1 illustrates the *iURBAN* methodology for identifying

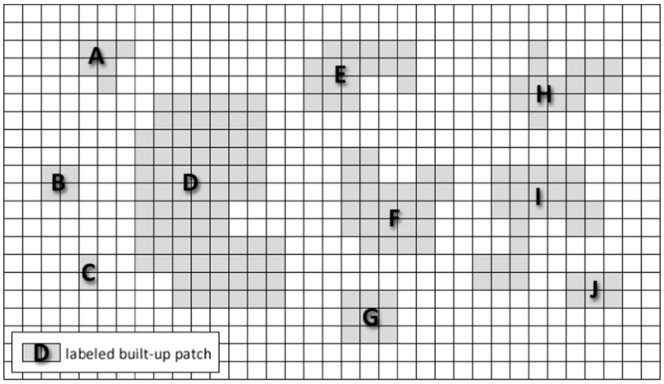

Fig. 1. Fictitious 'country' with various contiguous built-up patches, labeled A–J.

urban areas. After normalization of both the built-up area and the population value for each patch, the index is calculated by multiplication of the two normalized values. Subsequently, all patches are ranked according to their index value. The total population within the identified built-up patches is 134,400. However, the actual population count of this fictitious country is 162,000 people (considering that global population grids such as LandScan also distribute outside of built-up area as identified in the GUF data). Referring to a given national urban ratio of 60% (as reported in WUP), roughly 97,000 people are living in urban areas. To reach that value, the contiguous built-up patches are iteratively added following the ranking order. In this illustrated case, that would imply adding patches D, E, and I, after which a population count of 97,000 is reached. Those patches are then classified as urban, and all other patches considered rural.

As existing global gridded population data only allocate or capture a minor share of the rural population within the identified built-up area (for example, around 35% in Guatemala and Costa Rica, and <30% in Panama), a novel reallocation approach has been developed to overcome this limitation. *iURBAN*'s objective is to have the entire population allocated to built-up area in order to subsequently be able to use population counts as a proxy for other related parameters, such as building distribution. Absolute population counts in existing global grids are not considered trustworthy for rural areas, considering the poor population representation within built-up area in rural regions. As a starting point for built-up-adjusted rural population reallocation, the entire national-level rural share is taken as disaggregation basis. The contiguous built-up patches that were pre-identified as rural serve as disaggregation target zones. Following the assumption that the larger and denser such a patch (in terms of its inherent summed-up built-up area derived from the original high-resolution GUF data), the more likely larger population counts are; all patches are weighted and normalized accordingly. Those

Table 1
Illustration of *iURBAN* methodology for identification of urban areas.

Built-up patch	Built-up area		Population		Index	Ranking
	Absolute [km^2]	Normalized [0–1]	Absolute [people]	Normalized [0–1]		
A	1.10	0.07	1100	0.03	0.0019	8
B	0.70	0.05	500	0.01	0.0005	9
C	0.20	0.01	700	0.02	0.0002	10
D	15.00	1.00	43,000	1.00	1.0000	1
E	4.10	0.27	39,000	0.91	0.2479	2
F	5.30	0.35	8700	0.20	0.0715	5
G	1.85	0.12	3500	0.08	0.0100	6
H	3.30	0.22	22,000	0.51	0.1126	4
I	6.00	0.40	15,000	0.35	0.1395	3
J	1.60	0.11	900	0.02	0.0022	7
			134,400	Identified within built-up		
			162,000	Actual national population		

Table 2
Illustration of *iURBAN* methodology for rural population reallocation.

Built-up patch	Built-up area		Population	
	Absolute [km^2]	Relative weights	Original [people]	Reallocated [people]
A	1.10	0.08	1100	5073
B	0.70	0.05	500	3228
C	0.20	0.01	700	922
F	5.30	0.38	8700	24,444
G	1.85	0.13	3500	8532
H	3.30	0.23	22,000	15,220
J	1.60	0.11	900	7379
Rural population in BU			**37,400**	
Actual rural population				**64,800**

weights are then used to disaggregate the national level rural share of the population, resulting in a new patch population number.

Referring again to the fictitious country in Fig. 1 for illustrative purposes, Table 2 shows the *iURBAN* patch-level population reallocation approach for rural areas. In that case, the used population distribution grid (e.g. LandScan), allocates 37,400 people to the identified rural built-up patches. The actual rural population, however, is 64,800 people. <60% of the rural population is therefore captured correctly within built-up area. The inherent built-up area values of the rural patches are used as weights (in relative terms) to disaggregate the total rural population of 64,800 people to the rural patches. Consequently, the entire national population, both urban and rural, is allocated to built-up area.

As the next step, patch-internal population distribution patterns at the cell level need to be determined. This approach refers back to the aspect that population is being better captured in areas that are more likely urban, as well as to the prior assumption that the larger a contiguous built-up patch, the more likely it is to be urban. It is therefore concluded that for the largest contiguous rural patches – despite not relying on absolute population counts – the patch-internal distribution patterns can be considered as reallocation weighting factor (see relative relevance of the 'population' indicator for the largest rural patch [F] in Table 3). This means the cell with the highest population value (in the original population grid) within a contiguous built-up patch is correspondingly weighted strongest for the patch-internal disaggregation. However, the smaller a patch becomes (thus the higher its potential degree of ruralness), the less reliable even the relative internal distribution patterns of existing global population grids are considered. Therefore, the individual cells' built-up ratio is introduced as a counter-weighting factor, complementing the above relative population distribution weight in normalized manner (see relative relevance of the 'built-up area' indicator in Table 3). Taking the fictitious small rural patch A as an example, the 'population' weighting factor (or relative relevance) is determined at 0.21 and the complementing 'built-up area' counter-weight is consequently set at 0.79. In other words, for such a patch at the cell level, a large built-up ratio is considered a higher influence for the disaggregation process, compared to a high original population value. Fig. 2 illustrates the steps of the patch-internal reallocation for the rural

Table 3
Illustration of *iURBAN* methodology for determining relative relevance of contributing factors (population and built-up area) for patch-internal rural population reallocation.

Built-up patch	Built-up area	Population	Relative relevance	
	Absolute [km^2]	Reallocated [people]	Indicator 'population'	Indicator ' built-up area'
A	1.10	5073	0.21	0.79
B	0.70	3228	0.13	0.87
C	0.20	922	0.04	0.96
F	5.30	24,444	1.00	0.00
G	1.85	8532	0.35	0.65
H	3.30	15,220	0.62	0.38
J	1.60	7379	0.30	0.70

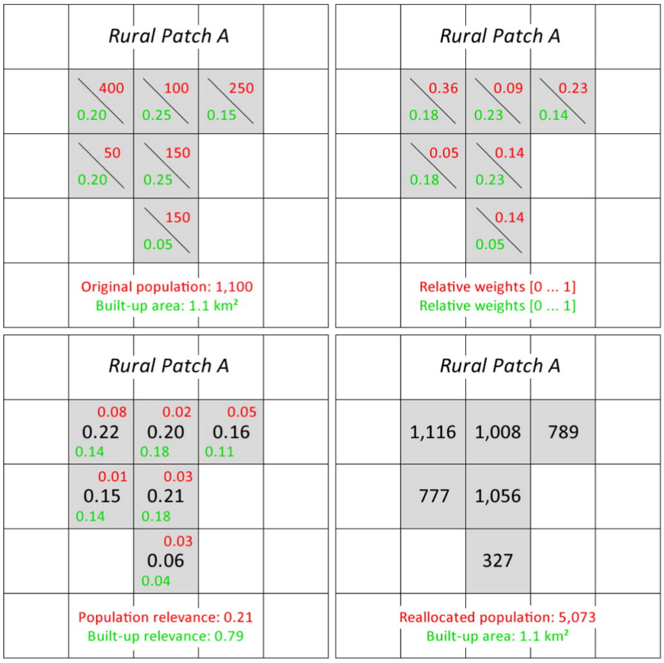

Fig. 2. Patch-internal 2-factor weighted population reallocation for the fictitious rural patch A.

patch A. First, the patch-internal weights for population and built-up area are calculated. Then, those weights are weighted according to the indicators' relevance for the respective patch, and summed to a final cell weight value. That value, in turn, is used for internal disaggregation of the pre-determined patch population.

Applying that 2-factor weighted disaggregation approach results in a new built-up-adjusted population distribution that, when summed-up from individual cell level to country level, equals the sum of the used original LandScan grid for the respectively identified reference year. In order to come up with a temporally consistent output, in an iterative process, *iURBAN* adjusts every country's reference year to the year 2012, which is determined as output time stamp. 2012 was chosen as it reflects the status of the GUF data which is derived from Radar imagery acquired during 2011/12. The adjustment works in a way that first population counts at cell-level are linearly projected from the identified reference year to 2012, using WUP's respective country totals. Then, the urban identification is re-run using the 2012 urban ratio.

3. Results

To date, *iURBAN* has been implemented for the countries modeled under the umbrella of the World Bank's CDRP project. These include all Central American countries (excluding Mexico) as well as several islands in the Caribbean, the latest additions being Jamaica, Saint Lucia, and Grenada. With easy transferability and upscaling in mind, a fully-automated geoprocessing tool has been developed that allows consistent implementation of *iURBAN* for other countries in a fast and straightforward manner. The ultimate goal of *iURBAN* in the context of CDRP is to provide a nationally consistent and spatially disaggregated basis for building stock analysis, in terms of building type-specific replacement value calculation and structural vulnerability assignments for consequent probabilistic risk assessment. In this paper we do not focus on those follow-up products (see Gunasekera et al., 2015), but rather on (see Fig. 3):

(1) the urban–rural classification as main output, and
(2) refined population distribution patterns (after urban–rural reallocation) as side product.

It should be clarified that, regarding the population pattern refinement, *iURBAN* is not a population distribution model per se. Rather, it refers to an existing modeled grid as input (both LandScan and WorldPop have been tested) and refines distribution patterns, in particular for rural areas, by integrating very high resolution built-up data as target features for reallocation. Rural population distribution patterns are problematic in existing global grids due to the hitherto unavailability of such local-level settlement data at global scale. As outlined in detail in Section 2.4, *iURBAN* eventually allocates all of the population to identified built-up structures, with resulting national level population sums and urban ratios compliant to WUP figures.

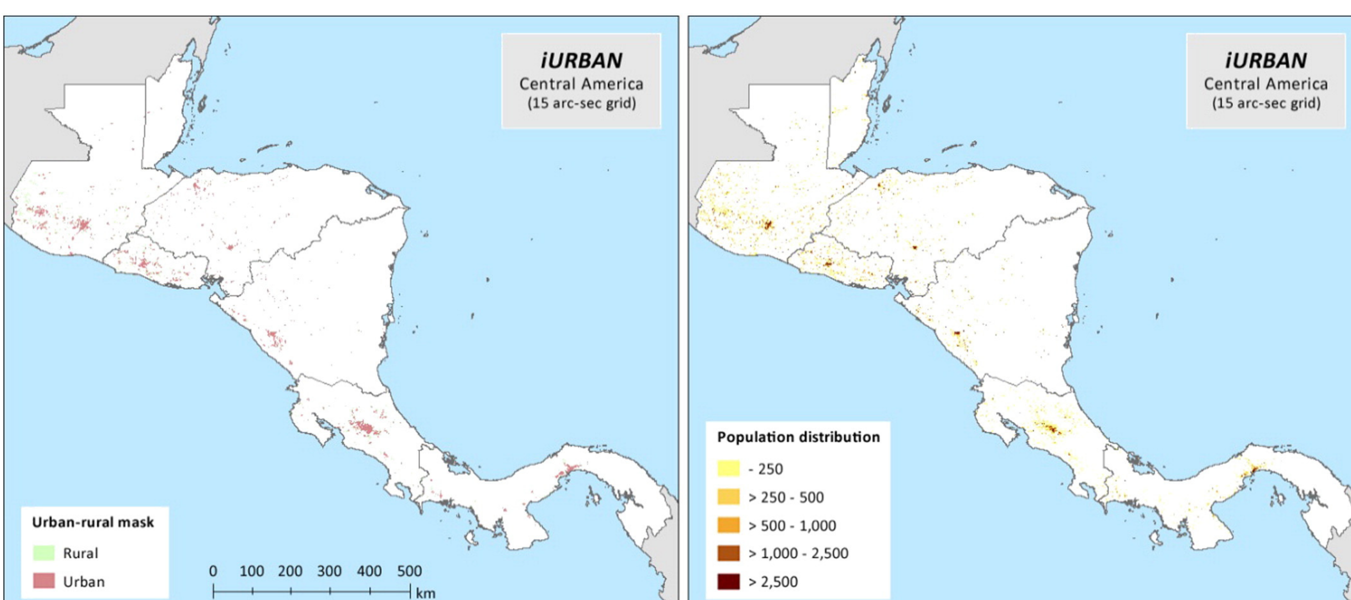

Fig. 3. Spatial overview of *iURBAN* output for Central America: 1) urban–rural mask (left), and 2) population distribution (right).

3.1. Urban–rural mask

Fig. 4 illustrates the output of the *iURBAN* model, taking Panama City as an example, in comparison to other existing urban identification approaches outlined above. Other urban layers include GRUMP urban extents (Balk et al., 2005), an urban mask based on uniform stable nighttime lights thresholding (>14DN) as replicated from Small et al. (2005), MODIS-500 urban areas (Schneider et al., 2010) which were e.g. used in the Global Assessment Report 2013 (de Bono and Mora, 2014), an urban mask based on uniform BUREF thresholding (>10%) as used in the Global Assessment Report 2015 (de Bono and Chatenoux, 2014), a WUP-adjusted urban mask using LandScan as replicated from Christenson et al. (2014a), and an urban mask as replicated from the EC-OECD approach presented by Dijkstra and Poelman (2014), also using LandScan. While at this stage it is not yet possible to provide evaluative true-false statements at cell level due to the lack of a corresponding reference source, it is clear that GRUMP and the nightlights-based approach overestimate the urban area (including large portions of non-built-up, uninhabited land), whereas MODIS and the BUREF-thresholding approach underestimate the built-up extent (failing to detect small settlements in rural areas in particular). The method presented by Christenson et al. (2014a) is the only one (other than *iURBAN*) that adjusts the urban classification to the WUP national urban ratio. Urban patches are rather scattered, with a number of falsely identified urban cells likely resulting from the LandScan-inherent road network bias (i.e. strong population disaggregation weight on road vectors, whereas high-resolution built-up data has not been integrated). The EC-OECD approach provides a higher degree of urban clustering, but does not conform to national WUP statistics. None of the presented approaches had access to high-resolution built-up data, which makes *iURBAN* vastly superior in the initial identification of settlement structures in urban environments, and even more so, in rural environments. The index-based approach at contiguous patch-level also ensures the proper clustering and compactness of derived urban extents.

3.2. Population distribution

Fig. 5 shows the *iURBAN* population reallocation output in comparison to six existing global gridded population distribution modeling approaches, again taking Panama City as the example. Exemplified population grids include the GPW version 3 (Balk et al., 2010) and version 4 (Doxsey-Whitfield et al., 2015) data sets, the follow-up GRUMP with urban refinement (Balk et al., 2010), WorldPop (Sorichetta et al., 2015), LandScan (Dobson et al., 2000), as well as the regional Latin American and Caribbean Population Database (LACD) (Hyman et al., 2000). The GPW data sets uniformly distribute population counts from administrative units to a predefined output grid. In that regard, GPW version 4 provides major improvement in terms of a much higher output resolution (i.e. 30 arc-sec compared to the 2.5 arc-min of GPW version 3) due to the increased availability of sub-national census data. GRUMP uses the GPW (version 3) population data as input, and further

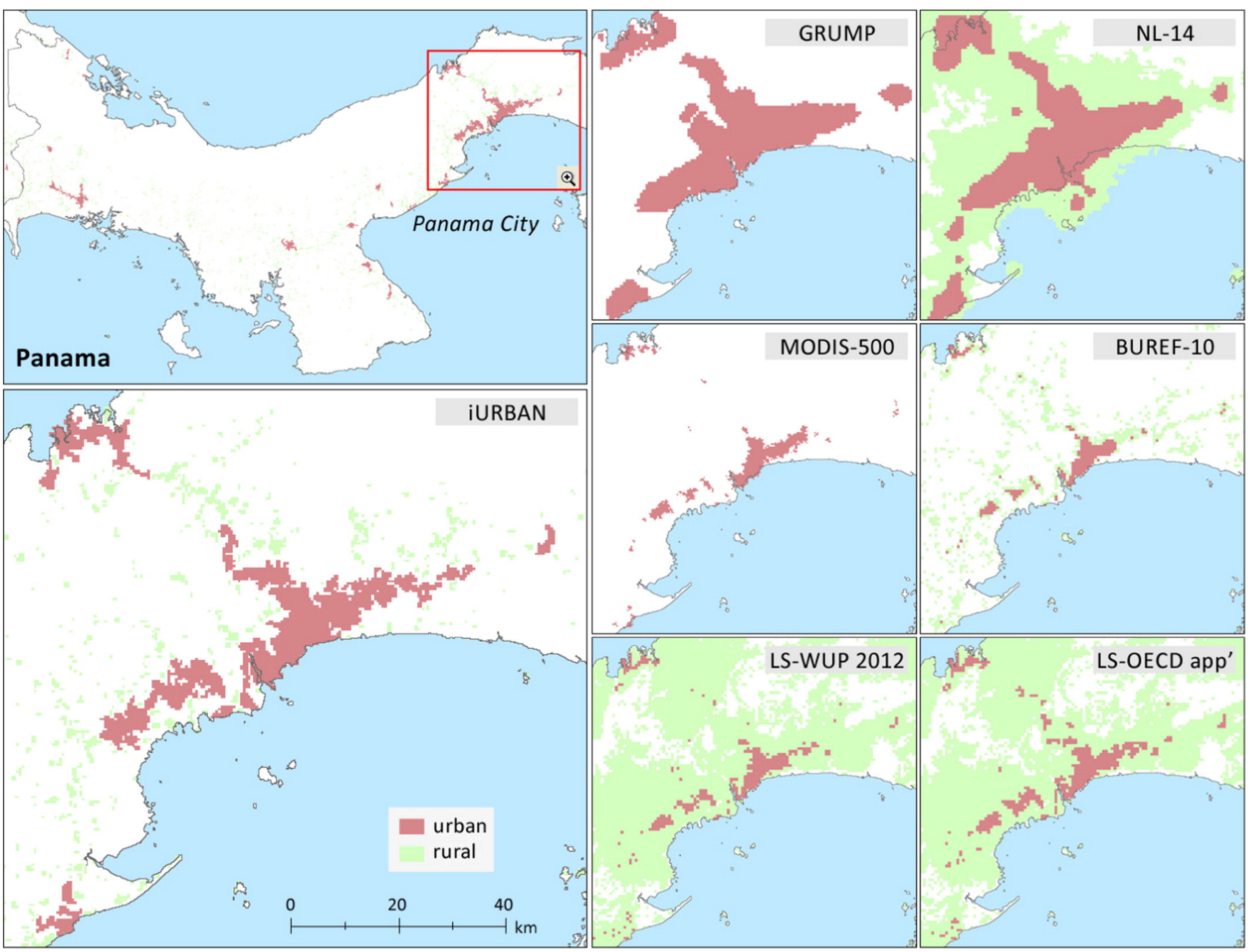

Fig. 4. Comparison of urban masks, Panama City example. *iURBAN* vs. six existing global approaches.

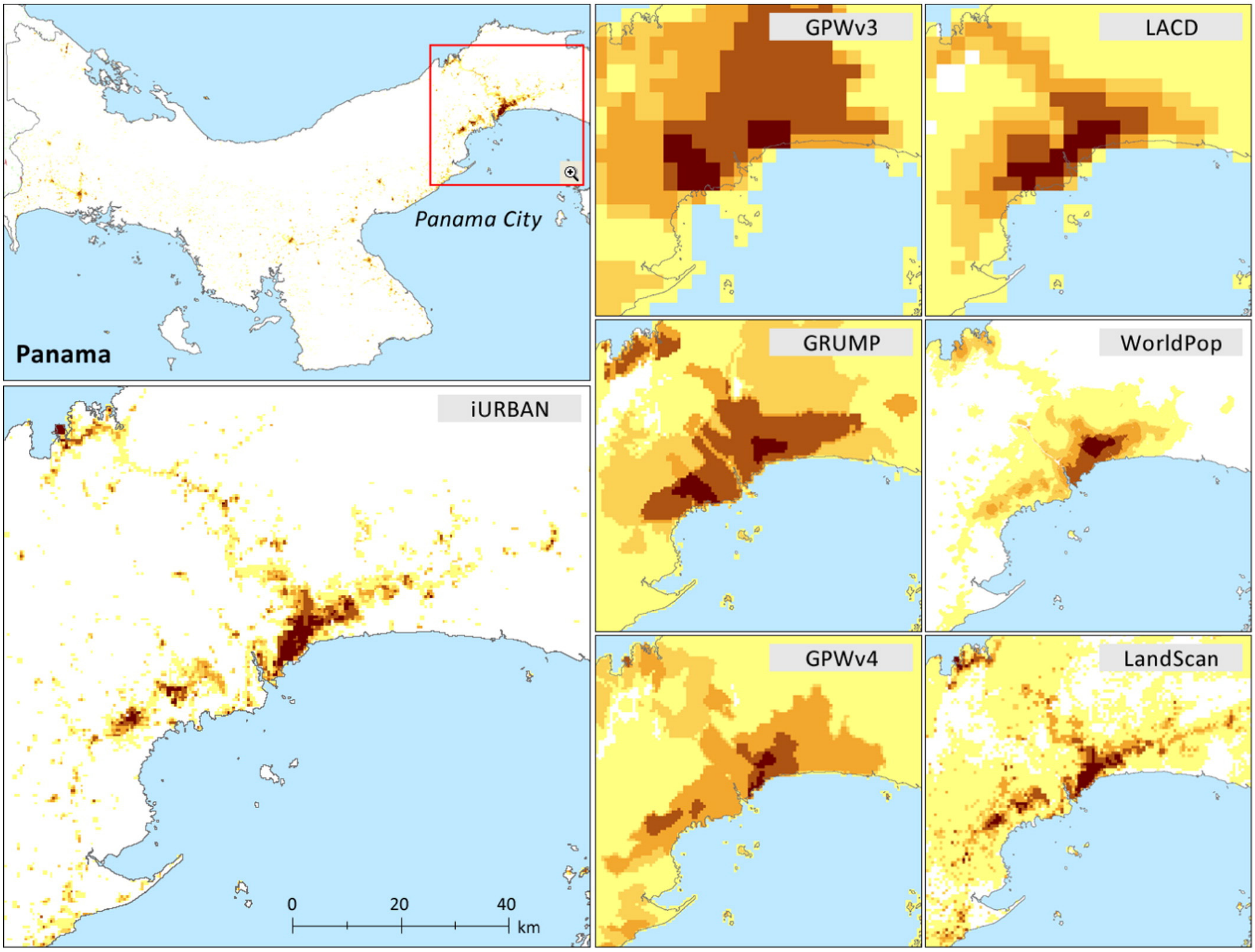

Fig. 5. Comparison of population distribution data sets, Panama City example. *iURBAN* vs. six existing global/continental population grids.

refines the distribution patterns by putting stronger weight on pre-identified urban areas. LACD, WorldPop, and LandScan all apply more complex modeling techniques to derive population disaggregation weights. To date, WorldPop and LandScan provide the most fine-grained distribution patterns. WorldPop clusters population more than LandScan, thus excluding more actual uninhabited land. LandScan, on the other hand, puts more focus on allocating peaks in urban centers, the latter being favorable for the *iURBAN* implementation approach. One aspect unites all of these data sets: very vague representation of rural population patterns. *iURBAN* provides significant improvement in that context by reallocating population to actual built structures as identified in the high resolution GUF data.

4. Accuracy and uncertainty

Accuracy and uncertainty are two crucial aspects contributing to the overall quality of every geospatial model. In remote sensing science, accuracy usually refers to the degree of 'correctness' of a map or classification (e.g. land cover), and is commonly analyzed using a confusion matrix, showing, amongst other measures, omission and commission errors (Foody, 2002). Respective quality assessment of *iURBAN* is, however, not that straightforward. The reason is there is simply no ground reference for what *iURBAN* is attempting to measure at the relevant scale level, i.e. whether a particular pixel is either urban or rural. There may be individual government maps of selected urban areas, but to our knowledge, there is no consistent reference at continental or even national scale available in that regard, that would allow adequate statistical accuracy assessment. National and regional land cover maps (such as the European CORINE or the US National Land Cover Database NLCD 2011) commonly depict built-up as opposed to urban land. It is not the intention of this paper to evaluate the accuracy of the built-up data we use (i.e. GUF) as there have already been various publications on that (e.g. Klotz et al., 2014; Felbier et al., 2014).

Urban–rural ratios are usually reported at the sub-national administrative level, referring to urban and rural population distribution from census collections. Inter-model comparative statistics can be calculated in that context, aggregating individual cells to sub-national units and using the census figures as reference. Taking Panama as an example, we compared modeled urban–rural ratios from different approaches at sub-national level to census figures (see Table 4). In addition to *iURBAN*, we selected three of the urban identification approaches illustrated in Section 3.1, namely GRUMP (using nighttime lights), BUREF-10 (using MODIS and LandScan), and LS-OECD (using certain clustering methods), for their inherent methodological differences. It should be specifically noted in this context that the reference years for all the various models can differ significantly. The most recent census in Panama was conducted in 2010. *iURBAN*, however, provides output consistently for the reference year 2012 (i.e. projected to match WUP 2012 total population counts and urban–rural ratios). GRUMP refers to population data from 2000, whereas LandScan v2012 (for Panama matching most closely the WUP 2007 total population count) is used for implementation of the BUREF-10 and LS-OECD approaches.

Table 4
Comparison of modeled urban ratios and deviations to census, Panama example. *iURBAN* vs. three existing global/continental urban identification models.

Province	Urban ratios					Deviation from census			
	Census	*iURBAN*	GRUMP	BUREF-10	LS-OECD'	*iURBAN*	GRUMP	BUREF-10	LS-OECD'
Bocas del Toro	0.3976	0.5016	0.4914	0.4068	0.5493	0.1039	0.0937	0.0092	0.1517
Chiriquí	0.5069	0.4725	0.3899	0.4598	0.5880	−0.0344	−0.1170	−0.0471	0.0811
Coclé	0.3446	0.4895	0.1828	0.2508	0.5195	0.1449	−0.1619	−0.0938	0.1749
Colón	0.6841	0.5793	0.8162	0.5884	0.7529	−0.1047	0.1321	−0.0956	0.0688
Darién	0.0832	0.0000	0.0000	0.0000	0.0000	−0.0832	−0.0832	−0.0832	−0.0832
Emberá	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Herrera	0.5414	0.4831	0.4681	0.3130	0.4472	−0.0583	−0.0732	−0.2284	−0.0942
Kuna Yala	0.0000	0.0000	0.3694	0.0000	0.0000	0.0000	0.3694	0.0000	0.0000
Los Santos	0.3180	0.2124	0.3756	0.1466	0.2802	−0.1056	0.0575	−0.1715	−0.0378
Ngöbe Buglé	0.0000	0.0000	0.0000	0.0000	0.1542	0.0000	0.0000	0.0000	0.1542
Panamá	0.9008	0.8367	0.8075	0.7730	0.9137	−0.0641	−0.0933	−0.1279	0.0128
Veraguas	0.3264	0.2368	0.3015	0.2383	0.2755	−0.0896	−0.0249	−0.0880	−0.0509
Average absolute deviation						0.0657	0.1005	0.0787	0.0758
Population-weighted[a] sum of absolute deviations						0.0802	0.1962	0.2028	0.1059

[a] National population totals differ for census, *iURBAN*, GRUMP, LandScan.

Two statistical measures are chosen for describing model output differences. The average absolute deviation (AAD) of the modeled urban ratios to the census reference gives an indication on the statistical dispersion, or the general magnitude of error at province level. *iURBAN* has the lowest AAD at 6.6%, with GRUMP having the highest value at 10.1%. The AAD considers all provinces statistically equally relevant and does not account for the size of a province in terms of its population. This is, however, crucial for urban ratios that inherently refer to population numbers. Therefore, we also report the population-weighted sum of absolute deviations as an additional quality parameter. The lower that value, the better the measured model output accuracy. Populous provinces therefore have a stronger statistical influence on the output, compared to sparsely population provinces. *iURBAN* again outperforms the other models with a value of 0.08, whereby BUREF-10 is at the other end of the scale with a value of 0.20.

The above statistical comparison of urban ratio deviations does, however, actually only illustrate one extracted aspect. Urban definitions follow different rules in different countries, some of which have multiple facets and are very ambiguous. Some settlements could thus even be considered partly urban (which *iURBAN* does not account for due to its consideration of built-up contiguity as one defining parameter). Furthermore - and this moves the discussion to the aspect of uncertainty - *iURBAN* is conceptually designed to provide the best quality at national level (i.e. as input for the Country Disaster Risk Profiles initiative) by adjusting for national WUP values. The fact that *iURBAN* provides better statistical values (i.e. lower deviations) at sub-national level compared to other global/continental approaches, is obviously a very positive result. However, that is more a favorable side effect, as the sub-national unit was never intended to be a primary reporting scale level.

LandScan, one of the major input sources for *iURBAN* (and various other models too), provides a lot of detail regarding local physical settlement features for population disaggregation (thus including more inner-urban variation). However, at sub-national level, aggregated LandScan totals rarely match census counts. Assumedly, the reason for this is not only census but also spatially coarse projection data are used for disaggregation (details on the methodology are not documented). *iURBAN* reallocates rural population to within built-up area in order to provide a closer match to the reality on the ground in terms of distributional patterns. Also, absolute population counts are adjusted for high-resolution built-up density, which is assumed to be most accurate (following common practice in dasymetric mapping).

With regard to population distribution patterns, e.g., GPW - when aggregated to sub-national level - will certainly provide the best accuracy compared to census counts, simply because GPW uses the highest level census data as disaggregation input. That does not signify, however, that GPW would be suitable to use at grid cell level for urban–rural identification. Considering built-up contiguity as an essential parameter of an urban settlement, this parameter would be undetectable in GPW as population is uniformly distributed within an administrative area. On the other hand, GPW could be used as a reallocation basis for *iURBAN* (as has been done, e.g. for the currently developed GHSL-based population distribution data of Freire et al., 2015). That would, however, consequently mean losing out on the complex distributional inner-urban detail that LandScan provides.

Given all the above, the quality of *iURBAN* should be viewed from different perspectives. Accuracy and uncertainty issues are deeply intertwined in the model setup and design, a fact that basically applies to all global and continental urban identification models. As outlined above, *iURBAN* provides a statistically measureable improvement for sub-national urban ratio identification even though this was not the initial goal of the model. The model's main strength lies in the use of much finer detection of settlement patterns via high-resolution, satellite-derived, built-up area. This implies a completely different basis for urban identification as well as detailed distributional population reallocation that is hard to illustrate adequately in statistical terms, due to the lack of appropriate ground reference data.

5. Discussion and conclusions

Taking on board the various different perspectives outlined above, the question which arises is why the choice of urban definition is such a major concern. While ascertaining what urban consists of is neither obvious nor straightforward, it is clear that the very definition of urban has many implications and ramifications, and is therefore of utmost relevance and importance. For instance, the targeting of national development or educational assistance programs, criteria for eligibility, as well as allocation of funds from those programs, depend on whether an area is classified as urban or rural. National data compilations such as census, housing, health, and living surveys, use national statistical definitions as guidelines for data collection and reporting. Consequently, when ancillary non-spatial data needs to be integrated with spatial urban extents for a specific application, accounting for the respective national statistical delineation approaches is crucial in order to ensure valid and retraceable linking. Remotely sensed information is often not used as an exclusive input data source for urban and regional development studies, but commonly linked to ancillary data for more integrative information retrieval (Aubrecht et al., 2013). Ancillary data, as derived using traditional methods such as surveys and public records, can include contextual information such as building typologies and population characteristics, as well as more complex indicators, such as explicit vulnerability aspects (Aubrecht and Özceylan, 2013). The prerequisite for the proper linking of two related data sets is always the availability of a common parameter, a distinct key in terms of relational database management. Focusing on regional and global scale urban research, high-to-medium spatial resolution remote sensing data can provide information on land cover patterns in a straightforward manner. When it comes to illustrating functional relationships and different use categories however, proxy measures are often applied, thus physical parameters are identified via remote sensing that are correlated to functional aspects. Data compilations which are not explicitly spatial, such as national censuses, encyclopedias, gazetteers, etc., commonly use classification schemes which introduce a typological structure for large data volumes. Global and national level databases, such as the World Housing Encyclopedia and the USGS PAGER (Prompt Assessment of Global Earthquakes for Response) initiative, refer to simple inventory region schemes such as urban and rural for their data collection (Jaiswal et al., 2010). When attempting to link such information properly to spatially-explicit data, the same classification scheme needs to be applied at both ends, i.e. non-spatial data collected for urban areas need to be spatially linked to according pre-identified urban locations.

'Link-ability' was highlighted as major issue of concern in integrating inventory region based tabular data with a spatial base framework. All the above-presented approaches have their objective firmly set on the identification of urban areas. However, many of the global challenges, such as poverty or natural hazards, have significant impacts in rural areas. For example, more than three quarters of the extreme poor live in rural areas, with those areas often being the most vulnerable to disaster risk and associated multi-faceted consequences (World Bank, 2013). It is therefore an issue of urgent necessity to take a closer look at the 'flipside of urban' (Aubrecht et al., 2015) in the context of spatial identification or urban–rural patterns. Whether existing approaches derive urban extent via the thresholding of nightlights or population density data, or via contiguous built-up area data, one assumption remains the same: the entire remaining area outside of urban is consequently defined as rural (Balk et al., 2010). Binary definitions of urban–rural do not capture the gradual spectrum between a city and its agricultural surrounding, with a peri-urban perimeter or suburban transition zone. Furthermore, uninhabited areas are conflated with low population density areas in an overarching definition of rural (Christenson et al., 2014a). While the conceptual urban–rural dichotomy is specifically evident in non-spatial databases, a consequently linkable, spatially-dichotomous representation should at least aim at excluding uninhabited and uncultivated land, and distinctly depict rural settlements. *iURBAN* addresses this issue by classifying spatially-explicit settlement structures in rural and urban, and leaving non-built land unclassified. Furthermore, allocating the entire population to those built-up areas guarantees full link-ability and conceptual compatibility, for joining non-spatial data collected based on urban–rural inventory region schemes. This compatibility consequently refers to both building- and population-based data compilations. In the CDRP project implementation, both building and population components are relevant as building typology data is disaggregated via an inventory region schema to cell level, where population counts are required as one proxy measure for final asset value calculation. Caution is advised when considering rural cultivated land and agricultural productivity, given that *iURBAN* specifically addresses the built environment.

Acknowledgments

This study has been carried out in the framework of the Central America Country Disaster Risk Profiles (CDRP) project, under the LCR Probabilistic Risk Assessment (CAPRA) Program (P144982). This work was funded by an Australian Government Grant through the World Bank's Global Facility for Disaster Reduction and Recovery (GFDRR) Program (Trust Fund #TF014499). The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of its Executive Directors, or the governments they represent. The Global Urban Footprint (GUF) product used in this study represents a preliminary raw version (status 08/14) which, at time of writing, had not yet been subject to any systematic quality assessment and improvement. It was provided by the German Aerospace Center (DLR) to the World Bank's GSURR LAC-DRM team for scientific testing applications. The LandScan (v.2012) data has been made available for World Bank internal use through the GFDRR grant from the Australian Government to the LAC-DRM team. The authors acknowledge discussions with the entire CDRP team that helped improve the presented work. Specific thanks also go to Dr. Christoph Freudenthaler for providing statistical analysis support.

References

Arino, O., et al., 2007. GlobCover: ESA service for global land cover from MERIS. IGARSS 2007, IEEE International Geoscience and Remote Sensing Symposium, pp. 2412–2415.

Aubrecht, C., Özceylan, D., 2013. Identification of heat risk patterns in the U.S. National Capital Region by integrating heat stress and related vulnerability. Environ. Int. 56, 65–77.

Aubrecht, C., Özceylan, D., Steinnocher, K., Freire, S., 2013. Multi-level geospatial modeling of human exposure patterns and vulnerability indicators. Nat. Hazards 68 (1), 147–163.

Aubrecht, C., Gunasekera, R., Ishizawa, O., 2014. Urban characterization aspects and associated spatial scale considerations in the context of global-level exposure modeling: comparative data evaluation. 2nd European Conference on Earthquake Engineering and Seismology, Istanbul, Turkey.

Aubrecht, C., Gunasekera, R., Ishizawa, O., Pita, P., 2015. The flipside of 'urban' - a novel model for built-up-adjusted rural–urban pattern identification and population reallocation. CDRP Working Paper (Country Disaster Risk Profiles). The World Bank GSURR, Washington DC, USA.

Balk, D., Pozzi, F., Yetman, G., Deichmann, U., Nelson, A., 2005. The distribution of people and the dimension of place: methodologies to improve the global estimation of urban extents. International Archives of Photogrammetry and Remote Sensing Vol. XXXVI(Part 8/W27). 5th International Symposium Remote Sensing of Urban Areas (URS 2005). Tempe, AZ, USA.

Balk, D., Yetman, G., de Sherbinin, A., 2010. Construction of gridded population and poverty data sets from different data sources. European Forum for Geostatistics Conference, Tallinn, Estonia, pp. 12–20.

Bartholome, E., et al., November 2002. (2002) GLC 2000: global land cover mapping for the year 2000 - project status. European Commission, Joint Research Centre. Institute for Environment and Sustainability, Ispra, Italy.

Bontemps, S., et al., 2015. Multi-year global land cover mapping at 300 m and characterization for climate modelling: achievements of the land cover component of the ESA climate change initiative. ISPRS Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XL-7 (W3), 323–328.

Brenner, N., Schmid, C., 2014. The "urban age" in question. Int. J. Urban Reg. Res. 38 (3), 731–755.

Burian, S.J., et al., 2006. Emerging urban databases for meteorological and dispersion modeling. The 86th AMS Annual Meeting, 6th Symposium on the Urban Environment, Atlanta, GA, USA.

Chen, J., et al., 2015. Global land cover mapping at 30 m resolution: a POK-based operational approach. ISPRS J. Photogramm. Remote Sens. 103, 7–27.

Christenson, E., Elliott, M., Banerjee, O., Hamrick, L., Bartram, J., 2014a. Climate-related hazards: a method for global assessment of urban and rural population exposure to cyclones, droughts, and floods. Int. J. Environ. Res. Public Health 11 (2), 2169–2192.

Christenson, E., et al., 2014b. Examining the influence of urban definition when assessing relative safety of drinking-water in Nigeria. Sci. Total Environ. 490, 301–312.

Croft, T.A., 1973. Burning waste gas in oil fields. Nature 245 (5425), 375–376.

De Bono, A., Chatenoux, B., 2014. A global exposure model for GAR 2015. Input Paper Prepared for the Global Assessment Report on Disaster Risk Reduction 2015. UNEP/GRID, Geneva, Switzerland.

De Bono, A., Mora, M.G., 2014. A global exposure model for disaster risk assessment. Int. J. Disaster Risk Reduct. 10 (Part B), 442–451.

Dijkstra, L., Poelman, H., 2014. A harmonised definition of cities and rural areas: the new degree of urbanization. European Commission Directorate-General for Regional and Urban Policy (DG REGIO).

Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., Worley, B.A., 2000. LandScan: a global population database for estimating populations at risk. Photogramm. Eng. Remote. Sens. 66 (7), 849–857.

Doxsey-Whitfield, E., et al., 2015. Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4. Papers in Applied Geography 1 (3), 226–234.

Esch, T., et al., 2012. TanDEM-X mission—new perspectives for the inventory and monitoring of global settlement patterns. J. Appl. Remote Sens. 6 (1) (061702–1).

Felbier, A., et al., 2014. The global urban footprint - processing status and cross comparison to existing human settlement products. 2014 IEEE Geoscience and Remote Sensing Symposium (IGARSS), Québec, Canada, pp. 4816–4819.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80 (1), 185–201.

Freire, S., Kemper, T., Pesaresi, M., Florczyk, A., Syrris, V., 2015. Combining GHSL and GPW to improve global population mapping. 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milano, Italy, pp. 2541–2543.
Gallego, F.J., 2004. Mapping Rural/Urban Areas from Population Density Grids. European Commission, Joint Research Centre, Ispra, Italy.
Gamba, P., Herold, M., 2009. Global Mapping of Human Settlement: Experiences, Datasets, and Prospects. CRC Press.
Gunasekera, R., et al., 2015. Developing an adaptive global exposure model to support the generation of country disaster risk profiles. Earth-Sci. Rev. 150, 594–608.
Henderson, M., Yeh, E.T., Gong, P., Elvidge, C., Baugh, K., 2003. Validation of urban boundaries derived from global night-time satellite imagery. Int. J. Remote Sens. 24 (3), 595–609.
Hyman, G., et al., 2000. Latin America and Caribbean population database documentation. International Center for Tropical Agriculture (CIAT), United Nations Environment Program (UNEP), and the World Resources Institute (WRI).
Imhoff, M.L., Lawrence, W.T., Stutzer, D.C., Elvidge, C.D., 1997. A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area. Remote Sens. Environ. 61 (3), 361–370.
Jaiswal, K., Wald, D., Porter, K., 2010. A global building inventory for earthquake loss estimation and risk management. Earthquake Spectra 26 (3), 731–748.
Kasimu, A., Tateishi, R., 2008. Global urban mapping using population density, MODIS and DMSP data with reference of Landsat images. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol XXXVII (Part B7), ISPRS Congress 2008, Beijing, China, pp. 1523–1528.
Katz, B., Altman, A., Wagner, J., 2006. An urban agenda for an urban age. Urban Age Conference, Berlin, Germany.
Klotz, M., et al., 2014. Mapping global exposure from space: a review of existing products and comparison of two new layers of global urban extent. Second European Conference on Earthquake Engineering and Seismology, Istanbul, Turkey.
Kreibich, V., 2010. Metropolis: our common future in the urban age? - an introduction. Our Common Future, Hannover/Essen, Germany.
Mertes, C.M., Schneider, A., Sulla-Menashe, D., Tatem, A.J., Tan, B., 2015. Detecting change in urban areas at continental scales with MODIS data. Remote Sens. Environ. 158, 331–347.
Moriconi-Ebrard, F., et al., 2008. Africapolis - Urbanization Trends 1950–2020: A Geostatistical Approach - West Africa (Paris, France).
OECD, 2012. Redefining "urban" - a new way to measure metropolitan areas. Organisation for Economic Co-Operation and Development, Paris, France.
Pesaresi, M., et al., 2013. A global human settlement layer from optical HR/VHR RS data: concept and first results. IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens. 6 (5), 2102–2131.
Potere, D., Schneider, A., Angel, S., Civco, D., 2009. Mapping urban areas on a global scale: which of the eight maps now available is more accurate? Int. J. Remote Sens. 30 (24), 6531–6558.
Salvatore, M., Pozzi, F., Ataman, E., Huddleston, B., Bloise, M., 2005. Mapping Global Urban and Rural Population Distributions - Estimates of Future Global Population Distribution to 2015. FAO Food and Agriculture Organization of the United Nations, Rome, Italy.
Saunders, D., 2010. Arrival City: The Final Migration and Our Next World. Knopf, Toronto, Canada.
Schneider, A., Friedl, M.A., McIver, D.K., Woodcock, C.E., 2003. Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data. Photogramm. Eng. Remote. Sens. 69 (12), 1377–1386.
Schneider, A., Friedl, M.A., Potere, D., 2010. Mapping global urban areas using MODIS 500-m data: new methods and datasets based on "urban ecoregions". Remote Sens. Environ. 114 (8), 1733–1746.
Small, C., Pozzi, F., Elvidge, C.D., 2005. Spatial analysis of global urban extent from DMSP-OLS night lights. Remote Sens. Environ. 96 (3–4), 277–291.
Sorichetta, A., et al., 2015. High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. Scientific Data 2, 150045.
Sutton, P.C., Goetz, A.R., Fildes, S., Forster, C., Ghosh, T., 2010. Darkness on the edge of town: mapping urban and peri-urban Australia using nighttime satellite imagery. Prof. Geogr. 62 (1), 119–133.
Uchida, H., Nelson, A., 2008. Agglomeration Index: Towards a New Measure of Urban Concentration. The World Bank, Washington DC, USA.
UN-DESA, 2015. World urbanization prospects, the 2014 Revision. United Nations Department of Economic and Social Affairs, Population Division, New York, USA.
Weeks, J.R., 2010. Defining urban areas. In: Rashed, T., Jürgens, C. (Eds.), Remote Sensing of Urban and Suburban Areas, Remote Sensing and Digital Image Processing. Springer, Netherlands, pp. 33–45.
Weeks, J.R., 2011. Population: An Introduction to Concepts and Issues. 11th edition. Wadsworth Publishing, Belmont, CA, USA.
Welch, R., 1980. Monitoring urban population and energy utilization patterns from satellite data. Remote Sens. Environ. 9 (1), 1–9.
World Bank, 2013. Global Monitoring Report 2013: Rural–Urban Dynamics and the Millennium Development Goals. The World Bank and The International Monetary Fund (IMF), Washington DC, USA.