Emergent Geospatial Data & Measurement Issues

Deborah Balk
CUNY Institute for Demographic Research (CIDR) & School of Public Affairs, Baruch College
City University of New York

Comments prepared for ‘Future Directions in Spatial Demography’
Santa Barbara, CA
12-13 December 2011
Where have we come from?

& where are we going to?
Transformation from tables to maps

- In little more than a decade, demographers have gone from a rather tabular view of the world to a spatial one
  - Spatial data have become seemingly abundant
- Spatial demography is not population geography
  - The former is typically based in the study of individual or population-level rather than the study of place.
  - These different traditions have lead to different data and methodological requirements.
    - Even as Demography has become more spatial it remains quite distinct from population geography
## Measures of Poverty

### Infant Mortality Rates

By Country

*Figures are for 2000 based on UNICEF (2003)*

<table>
<thead>
<tr>
<th>Country</th>
<th>IMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>168</td>
</tr>
<tr>
<td>Australia</td>
<td>5</td>
</tr>
<tr>
<td>Brazil</td>
<td>33</td>
</tr>
<tr>
<td>Cambodia</td>
<td>97</td>
</tr>
<tr>
<td>Cameroon</td>
<td>95</td>
</tr>
<tr>
<td>China</td>
<td>30</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>78</td>
</tr>
</tbody>
</table>

*IMR (Infant Mortality Rate): deaths per 1000 live births*
Measures of Poverty

Infant Mortality Rates

By Subnational Administrative Unit

Subnational mortality rates are adjusted to 2000 using national trend data. Original data for 96% of countries are from 1995 or later. All data are from 1990 or later.


Robinson Projection

National Boundary

Subnational boundaries have been removed from countries for clarity.
Survey cluster locations

Population density
per sq. km
- 0 - 1
- 2 - 5
- 6 - 25
- 26 - 50
- 51 - 100
- 101 - 250
- 251 - 92,837

Extreme cases present
- High
- Both
- Low
- Neither

Balk et al, 2004
Classify demographic rates by spatial features

- Infant and child survival by distance to city of 50K persons or more
  - Or by length of growing season
We now expect

- **Micro-data**
  - Publicly available
    - Some information about respondents’ location
      - Survey cluster and/or
      - Corresponding spatial boundaries
  - Restricted data
    - Full access to micro data though level of address matching varies

- **Aggregated data**
  - Increasingly fine resolution census (or other administrative) units

- Basic population grids that are constructed with demographically rigorous methodologies
New and (Re-emergent) Data & Methods

On what topics?
Demographic inquiry that require spatial data

- What are the dominant demographic issues of the 21st century?
  - Migration
  - Urbanization
  - Aging
    - Changing family and household structures that arise from these many demographic shifts
    - Migration and urbanization are intrinsically spatial phenomena
  - Associated characteristics
    - Vulnerabilities (including age and sex)
    - Inequality
      - Spatial inequality is often one aspect
Emergent data

And under utilized data
Cell phones

- Useful for measuring mobility, if not migration, and population distribution

- Concerns:
  - Analytical: how to use these data meaningfully?
    - Look at how daily temp and precip data for clues
  - Ethical: Privacy concerns would need to be addressed
    - But there are precedent for this
  - Computational: Volume of data are very large
  - Practical: Data ownership and stewardship
Night-time lights time series

- ‘Urban’ Spatial Change
- Compare change over time, annual data from 1992/3
  - Red = 1992
  - Blue = 2009
  - Annual data available

- Before using
  - Needs careful vetting
  - Method to calibrate between years and reduce blooming

Balk and Montgomery, 2011
New Methods

Primarily for data integration
(or for creating new data)
Remote sensing data is underutilized by demographers

- Main exception to this is subfield of pop & environment where moderate and high resolution satellite data have been coupled with household survey data, typically

For example, the night-time lights:

- Recent study uses night-time lights brightness to indicate seasonal migration and population density changes to predict measles outbreaks in Niger (see next slide, by Nita Bharti et al. *Science*, 2011)
Bharti et al. 2001
Urban change over time

- This is really low hanging fruit
- Requires satellites
  - Night-time lights time series
  - Landsat or higher resolution place-specific comparison
  - SRTM
- Would be great to have finely resolved census data to correspond (closely) with satellite views but this is not a prerequisite
  - Though some way to add names and population characteristics is essential
Urban *Spatial* Change: Landsat

**Jequié, Brazil**

<table>
<thead>
<tr>
<th>Measure</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>Annual % Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>144,030</td>
<td>147,439</td>
<td>0.19%</td>
</tr>
<tr>
<td>Built-Up Area (sq km)</td>
<td>18.23</td>
<td>36.55</td>
<td>5.66%</td>
</tr>
<tr>
<td>Average Density (persons / sq km)</td>
<td>7,899.16</td>
<td>4,033.64</td>
<td>-5.18%</td>
</tr>
<tr>
<td>Built-Up Area per Person (sq m)</td>
<td>126.60</td>
<td>247.91</td>
<td>5.46%</td>
</tr>
<tr>
<td>Average Slope of Built-Up Area (%)</td>
<td>5.79</td>
<td>7.66</td>
<td>2.46%</td>
</tr>
<tr>
<td>Maximum Slope of Built-Up Area (%)</td>
<td>31.58</td>
<td>43.05</td>
<td>2.48%</td>
</tr>
<tr>
<td>The Buildable Perimeter (%)</td>
<td>0.86</td>
<td>0.87</td>
<td>0.05%</td>
</tr>
<tr>
<td>The Contiguity Index</td>
<td>0.95</td>
<td>0.98</td>
<td>0.20%</td>
</tr>
<tr>
<td>The Compactness Index</td>
<td>0.36</td>
<td>0.25</td>
<td>-2.83%</td>
</tr>
<tr>
<td>Per Capita Gross Domestic Product</td>
<td>$5,623.89</td>
<td>$6,857.49</td>
<td>1.30%</td>
</tr>
</tbody>
</table>

Sheppard et al, 2008
Urban *Spatial* Change: SRTM

Scatterometer - Average

- **High**: -6.800000
- **Low**: -15.600000

Scatterometer - Standard Deviation

- **High**: 3.300000
- **Low**: 0.000000
Create better spatial aggregates

Combine census with survey data

- Poverty Maps
- Some have used this method for demographic rates

Muniz et al, 2008
Create better population grids:
Age-sex specific+

- Mapping the denominator
  - Malaria transmission classes (a)
  - Percent of ward-level population under age five (b)
  - Ward-level misestimation that would result from use of national-level age distribution (c)

Tatem et al, 2011
Quantify spatial uncertainty

- The more we mix and match data sets of differing underlying resolutions, the more we will need to quantify the uncertainty of resultant data products
  - This will require some additional methodological work
  - Greater transparency on how integrated data products are produced is an important first step
    - Spatial metadata are necessary but insufficient for downstream use.
    - Traditional codebooks that accompany data tables are also necessary
      - Along with clear descriptions of integration
Create flexible spatial aggregates of census micro-data

- Census micro-data availability
  - Fairly coarse admin units
  - In the US & Canada, in enclaves (RDC)
  - Fee-for use tabulations to census (at least, in USA)

- Greater flexibility in creating summaries by user-specific-aggregates
  - For example, demographic characteristics (beyond what is available in block-level data) of flood plains or narrow coastal zones

- Confidentiality issues: Enclaves or on-line?
  - Technological solutions
  - More common protocols across countries’ statistical offices
Consider new study designs and sample frames

- Except for some exemplary place-based work, we are largely retrofitting yesteryears’ study designs to meet our current needs.

- Do we need to rethinking our sampling frames?
  - If we are interested in sorting our results by various ecological units, why not treat ecological characteristics like other strata?

- Geographic data, especially RS data, can be helpful in constructing sampling frames.
  - Detection of slums
  - Emergence of new cities, town, or temporary dwelling (refugee camps)
Conclusions

Challenges?
Conclusion and a caution

- Do more with what we have
- Embrace new data and methods
- Embrace ‘google Earth’
  - Spatial data awareness and interest is much greater than in the past.
  - Double edge sword: Much investment and education is still needed to use these data rigorously
A cautionary note

- Non-demographers often want demographic data
  - If there is not high engagement from the spatial demographic community, non-demographers will create it anyway, often inadequately
  - One way to avoid this is through interdisciplinary collaboration

- McDonald et al., 2011 (PNAS)