European Forum for Geostatistics Conference (EFGS)

The production & use of spatial statistics for sustainability

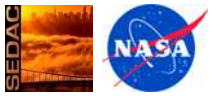
Statistics Portugal, Lisbon, 12-14 October 2011.

A Comparison of Vector and Raster Population Time Series: Population Change in Latin America and the Caribbean, 1990-2010

Susana B. Adamo and Liana Z. Razafindrazay

CIESIN, The Earth Institute, Columbia University

sadamo@ciesin.columbia.edu



Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Acknowledgements: this presentation is based on previous work by:

- NASA/SEDAC & CIAT
- National Science Foundation's Grant No. 0709606.
- Collaborators:
  - CIESIN/SEDAC: Marc Levy, Maria Muñiz, Sandra Baptista, Greg Yetman, Tricia Chai-Onn, Maria Elisa Lukang, Deborah Balk (currently CUNY)
  - Others:
    - University of Puerto Rico: T. Mitchell Aide, Maria Jose Andrade
    - Sonoma State University: Matthew L. Clark
    - University of California Santa Barbara: David Carr
    - Universidad Nacional de Tucumán: Ricardo Grau

Outline

1. Population censuses
2. Population spatial time series
3. Case example: Latin America and the Caribbean

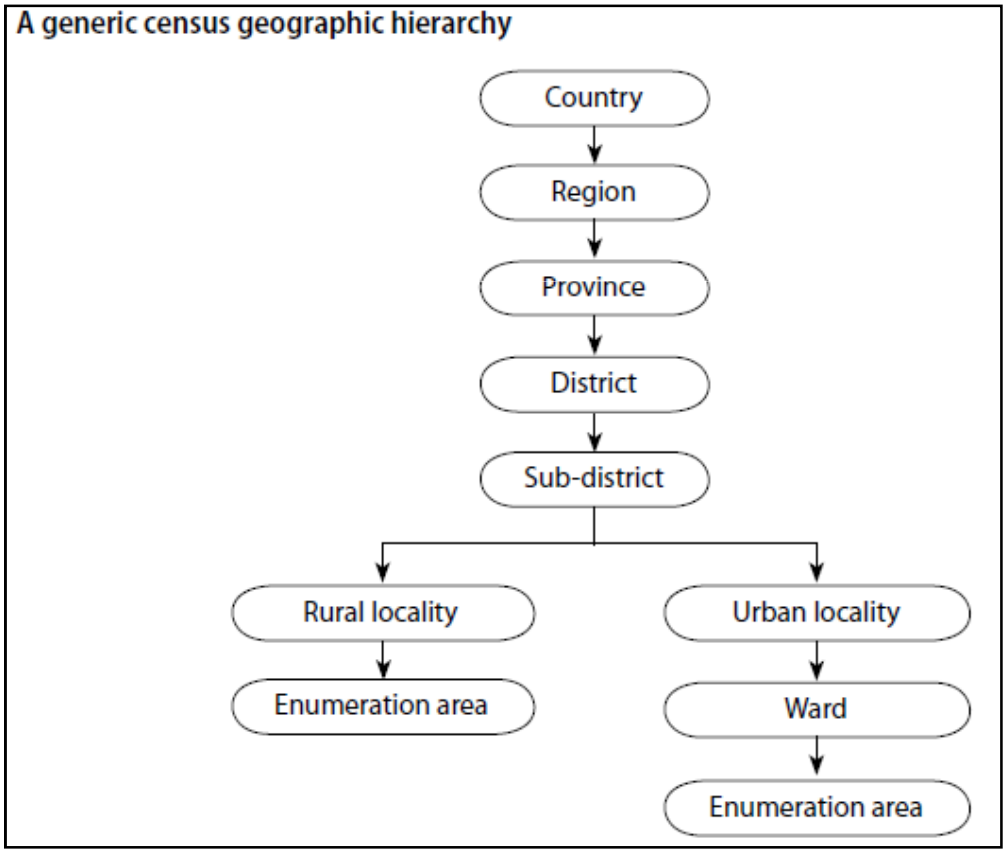
- ***Work in progress!***

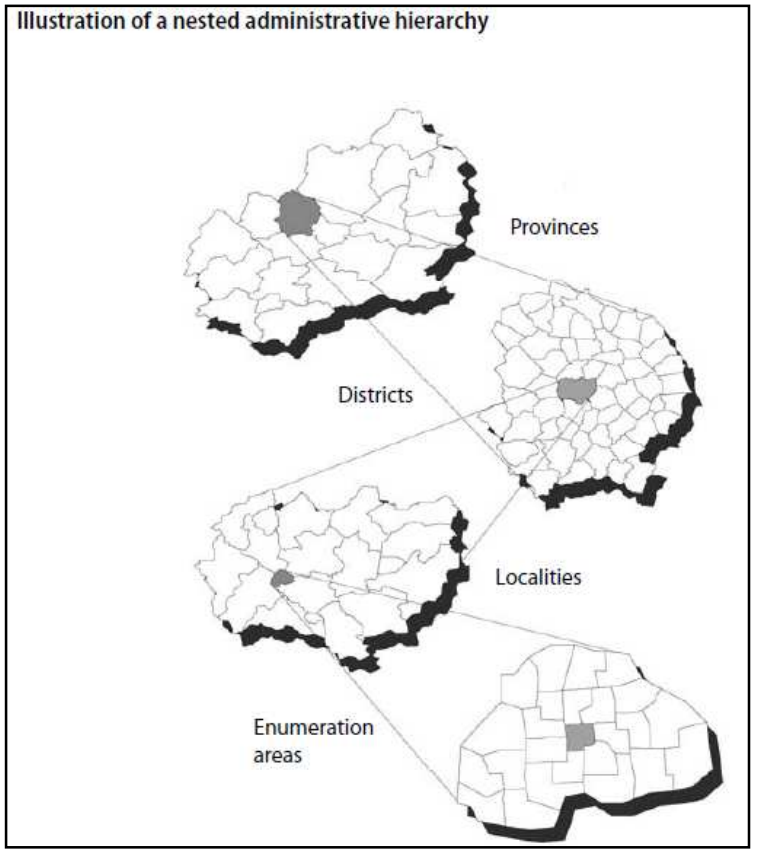
1. Population Censuses

- The total process of collecting, compiling and publishing demographic, economic and social data pertaining, at a specified time or times, to all persons in a country or delimited territory.
  - Aim: complete coverage of a country's population (population counts, demographic structure and characteristics) at a given time
- Set of questions reasonably comparable across countries and over time.
  - Growing standardization of practices (questionnaires, field activities, data capture, etc.)
- Every 10 years (although actual intervals vary across countries).
- Not all the collected information is actually available:
  - not processed because of budget constrains, confidentiality, etc.
  - some limitations or restrictions in terms of accuracy, privacy, immediacy, and coverage

Population Censuses (cont.)

- ***Large, exhaustive, rich geographic coverage when fielded*** (although spatial detail of publications could be very coarse).

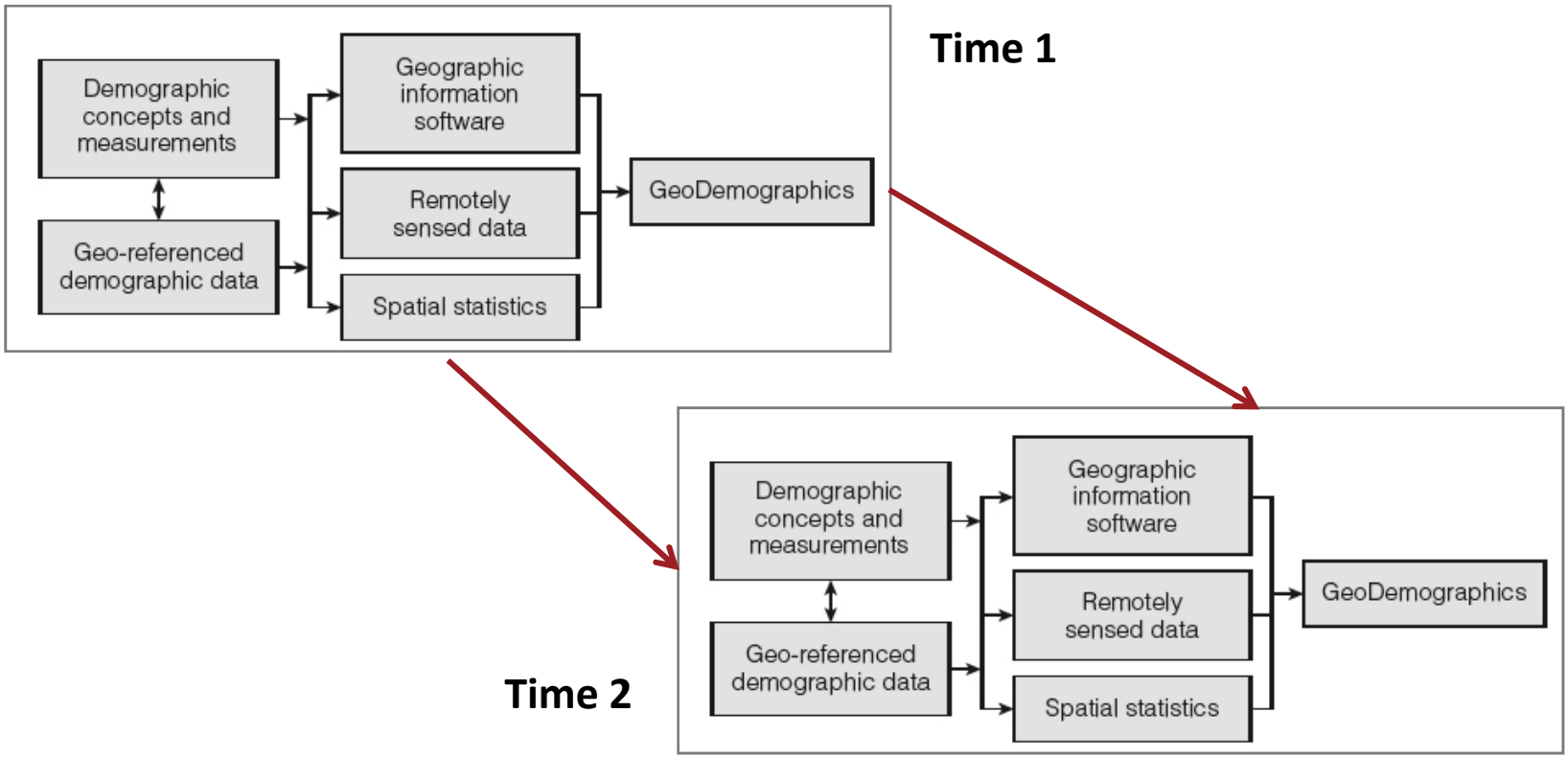




Source: UN. Statistics Division. 2009

2. Population Spatial Time Series

- Combination of GIS and demographic data over time



Graph source: Weeks 2008, Ch.4

Characterizing changes in spatial distribution of population over time is important:

- As driver of global change processes:
  - Land use;
  - Climate change;
  - Hydrology;
  - Emerging infectious diseases.
- Element of vulnerability to :
  - Global change;
  - Natural disasters;
  - Epidemics;
  - Political conflict;

Fundamental building block of (social and ecological) sustainability science.

2. Population Spatial Time Series (cont.)

- It is a hard task to develop at a large scale and at a higher level of detail (sub-national level).
  - Georeferencing of census data (basic data source) is not standardized across countries.
    - Each country does it differently.
  - Boundaries change over time
    - Complicates dynamic analysis of spatial data.
- As a result, spatial studies of population dynamics tend to be small-scale, local and regional
- Gap: continental-scale spatial analysis of population dynamics.

"There is a trade-off between geographical detail and chronological depth so traditionally we can look at how a phenomenon varies over space *or* how it varies over time but have only very limited capacity to look at both" (Gregory, I. and P. Ell 2005. Breaking the boundaries: geographical approaches to integrating 200 years of census. *Journal of the Royal Statistical Society.* 168:419-437:420)

Some issues of census data

Census geography changes over time, especially –but not only-- for the smallest units (census tract level or higher), usually defined in terms of number of households or people.

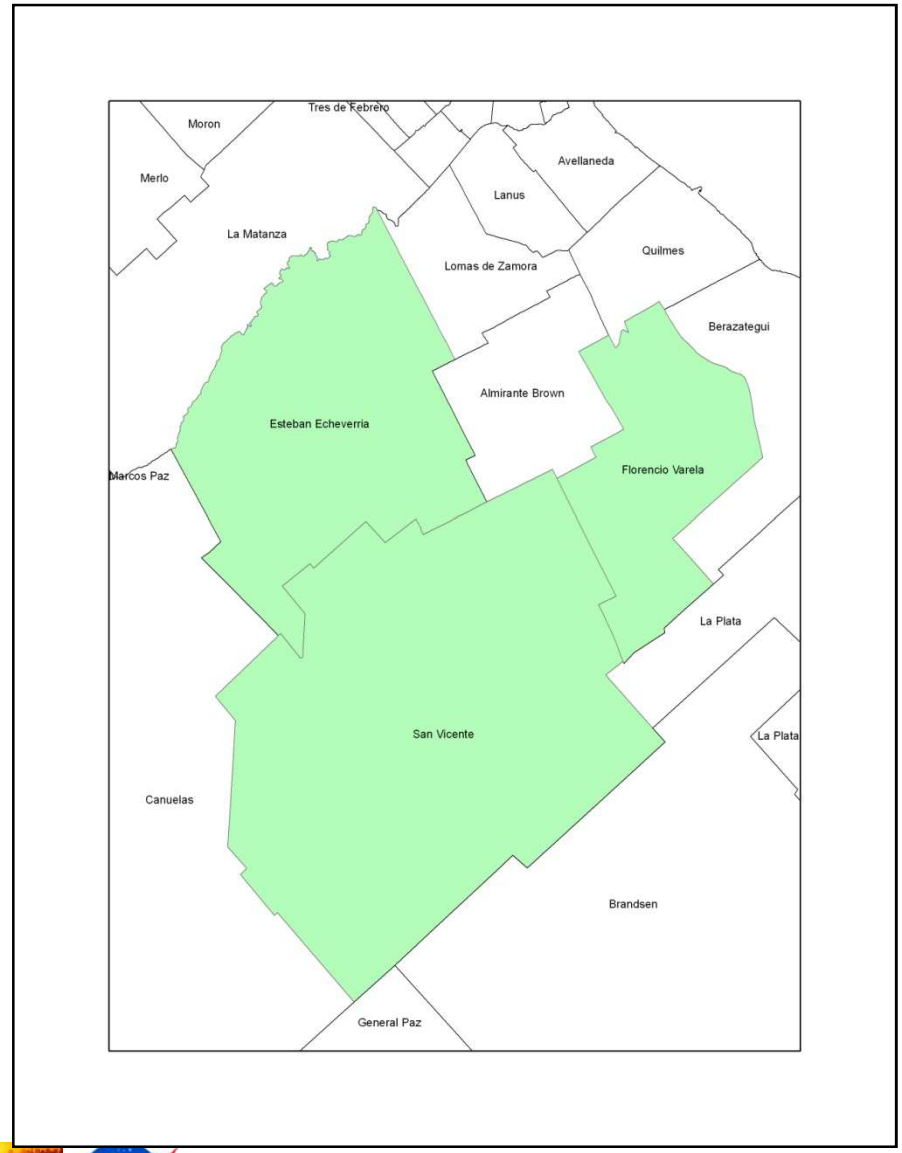
It is difficult to characterize with precision inter-annual changes in the spatial distribution of population because of these changes in administrative boundaries/census geography across censuses.

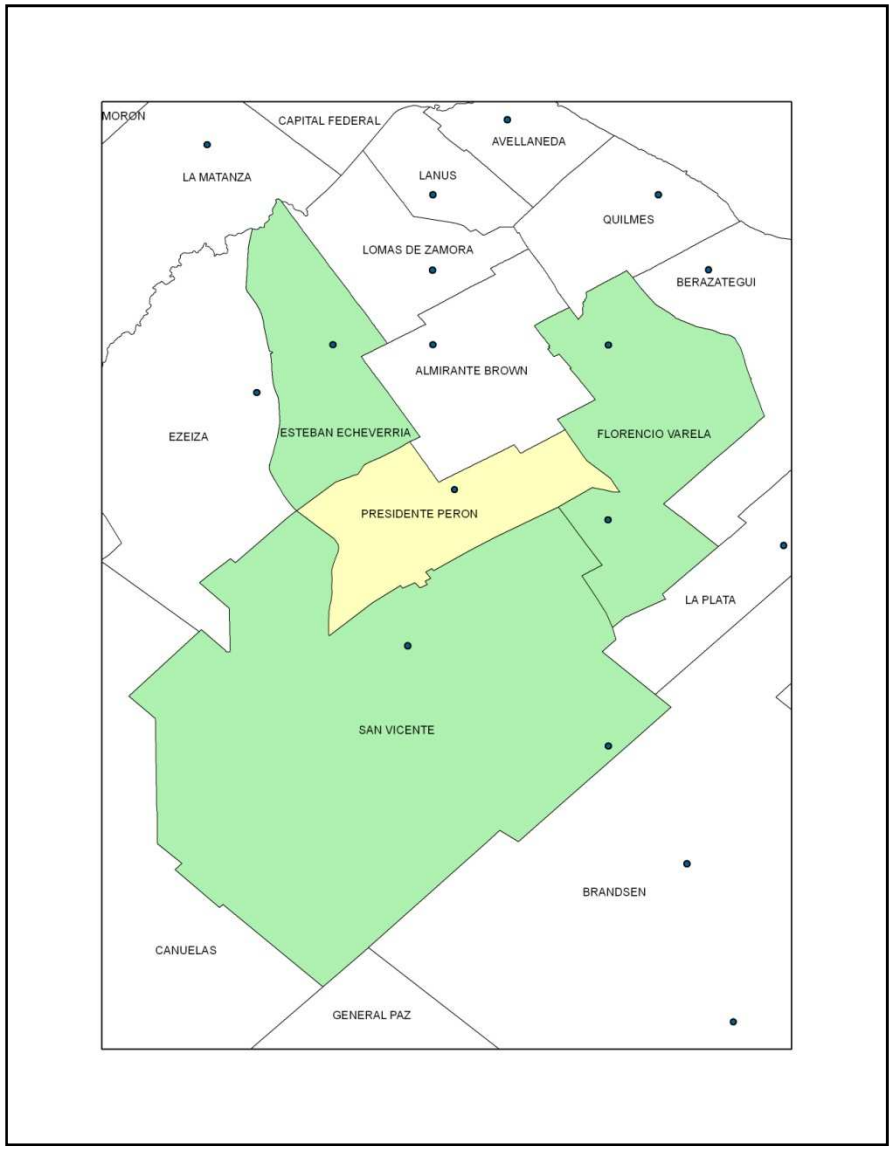
- Greater spatial precision in census units and tracking changes in the framework would permit comparisons over time.
- Commitment to "spatial backcasting" when census units are redrawn would contribute to the building of spatial census time series.
- Development of spatial time series over the last decades has taken advantage of:
  - more and better data;
  - methods for harmonizing information over time and space (getting future and past estimates, filling the gaps, etc.)

3. Case example: Latin America and the Caribbean population spatial time series

- Two datasets that rely on census data for population and cartography with different formats
  - Gridded population of the world version 3 (2005)
  - NSF-LAC version 1 (2010)
- Different perspectives
- Inputs are relative simple: population of administrative areas, usually in census years from paper or digital tables, and spatial boundaries of administrative areas
- Population and boundary data must match
  - Matching the inputs to one another is not as easy as it might seem
  - Boundaries change often and come in different scales
  - Population data may not match boundaries
    - We may have population values for different years at different levels (e.g., district-level one year, state-level another)
  - Population and boundary data may not match themselves

Example: a new municipality





Adjusting population data

- Census data as close as possible to the target years of 1990 and 2000.
  - Data from multiple census years.

- Matching to target years:
  - Calculate annual growth rates: $$r = \frac{LN[(P_2/P_1)]}{(t_2 - t_1)}$$
  - Extrapolate and interpolate to generate estimates for 1990 and 2000, applying the growth rate to the official population estimates:

$$e^{rt} \times P_1$$

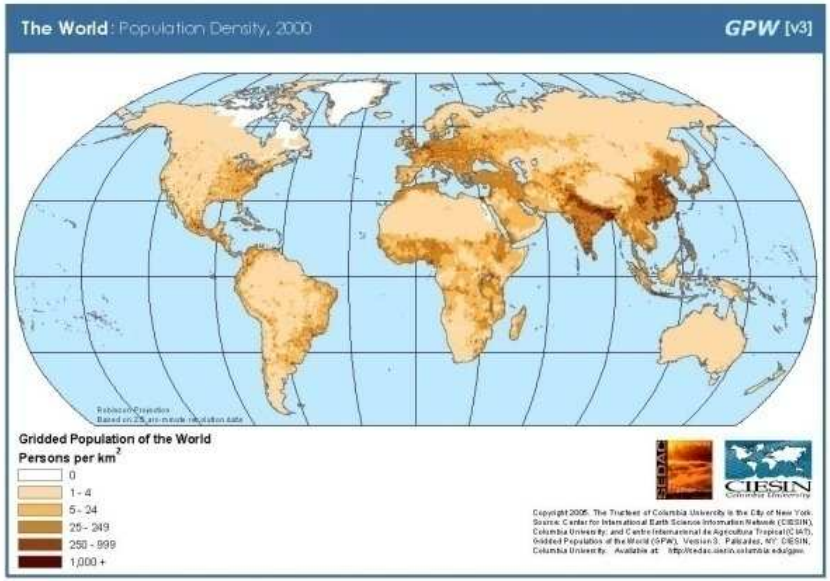
Adjusting spatial boundary data

- Often less available than population data.
- Main issues:
  - Changes in administrative boundaries between 1990 and 2000.
  - Spatial and population mismatches:
    - Outdated boundaries (prior to 1990);
    - Administrative unit names in the census that did not match the boundary file;
    - Administrative unit changes between the two censuses (splits and newly formed units) .
- Changes due to subdivision of census enumeration units, but also to administrative changes.

NSF- LAC version 1

- NSF-LAC database: built to characterize changing population distribution and growth in Latin America and Caribbean for 1990-2000 using replicable methods; spatial consistency; comparable units; spatial resolution suitable to global change analysis.
  - vector (polygon) format;
  - spatially-explicit census data; 1990 and 2000
  - density, distribution and growth
  - includes population counts, density, change and growth, 1990-2000,
  - municipio level or equivalent (level 2)

Gridded Population of the World version 3



- Distributes population across a common global grid;
- Raster dataset, rasterized to a 2.5' (4 km at the equator) grid;
- Population counts and density;
- Areal weighting used for population allocation to cells in the administrative units.
- Why gridding?
  - Allows aggregation of population by any other spatial and geographic phenomenon and facilitates analysis with these other data types, e.g., land cover, ecosystems, coastal proximity, etc.
  - Other subnational data is often proprietary, e.g population distribution by administrative boundaries

Basic Descriptives

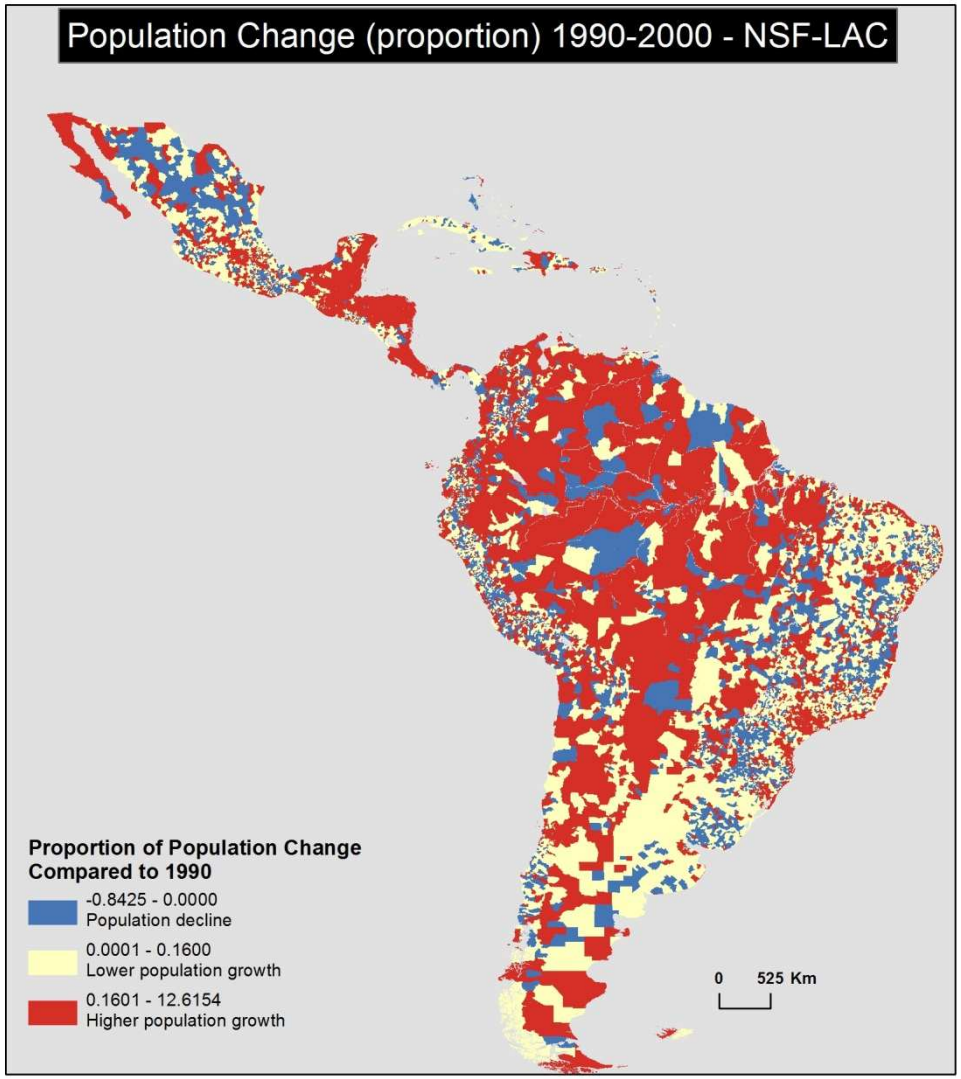
	NSF-LAC		GPW3	
Variables	1990	2000	1990	2000
Population Counts				
mean	26960	32144	227.87	286.32
max	9562381	10434252	492541.4	495677.5
min	18	16	0.00	0.00
standard deviation	124032	142792	3591.6	4024
Population Density				
mean	230.00	257.00	11.65	14.68
max	48000.00	42488.00	244334.00	24494.00
min	0.01	0.00	0.00	0.00
standard deviation	1500.00	1480.00	185.90	206.75

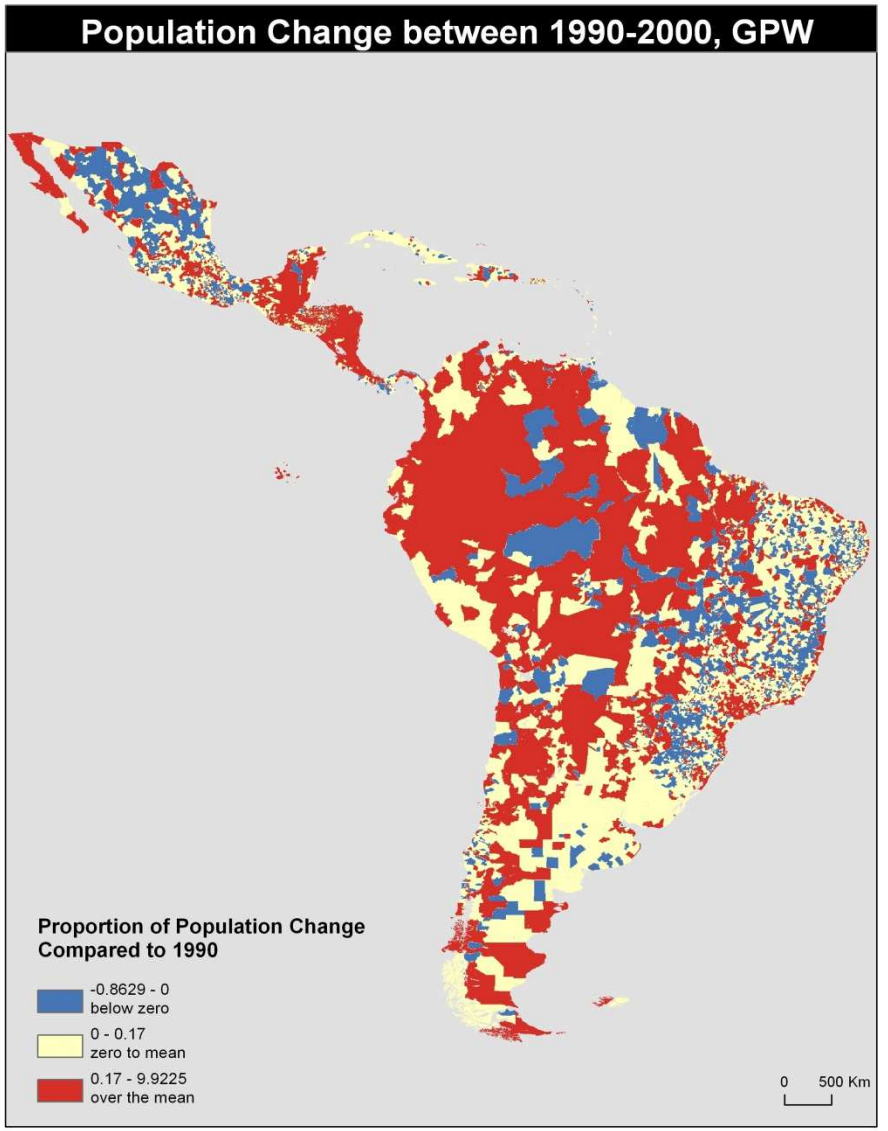
	NSF-LAC	GPW3
Population Change 1990-2000		
mean	0.1581	0.1778
max	12.6154	9.9225
min	-0.8425	-0.8629
standard deviation	0.3607	0.2445

Population Estimates 2010	NSF-LAC	GPW3
Population Counts		
mean	39888.55	233.42
max	11385617.75	494600.09
min	14.22	0
standard deviation	170012.20	3635.92

Population Density ,1990 and 2000

Population Change 1990-2000

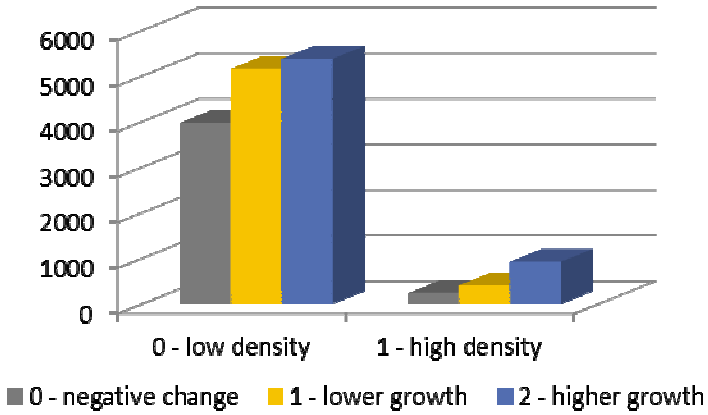




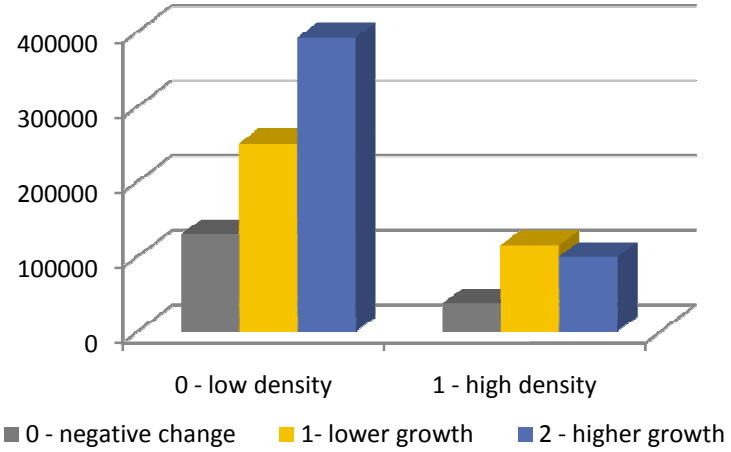
Population Change and Population Density

Population change 1990-2000 by population density in 1990

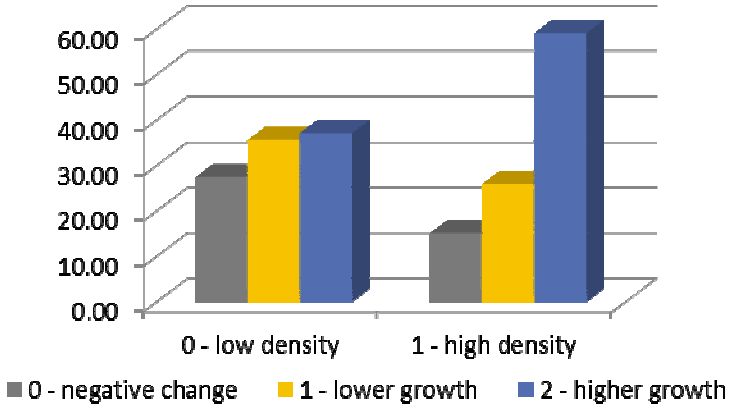
NSF-LAC

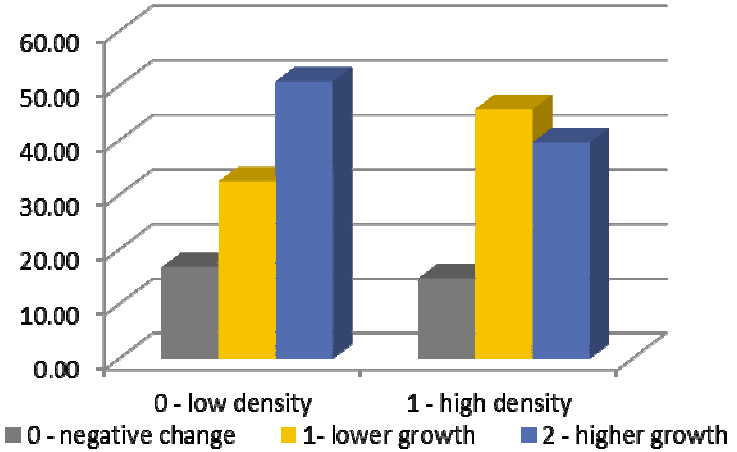


GPW 3



Population change 1990-2000 by population density in 1990 (%)





Future estimates