World Population Data not so Simple

Max Wyss[1], Philippe Rosset[1], Andrea de Bono[2]

[1]World Agency for Planetary Monitoring and Earthquake Risk Reduction, Geneva,
[2]UNEP/DEWA/GRID-Geneva

Worldwide distribution of the population is necessary for many applications. We focus on estimating casualties due to earthquakes in real time. Contrary to popular opinion, very few countries give detailed data (down to single villages or census tracks). Census data are generally restricted to large cities. The LandScan team believes that their data set "is the community standard for global population distribution". However, a comparison with our QLARM data suggests that more useful data sets may exist or may be constructed. To supply first responders with population and casualty data in cases of disasters the LandScan data have two undesirable features. (1) The LandScan population units are given with a resolution of approximately 1 km, which is too small for calculating losses due to earthquakes. First, this makes it difficult or impossible to assign each unit an appropriate building stock because the latter depends on settlement size. Second, estimated losses can only be meaningful if large numbers of buildings are considered, such that errors due to unknown factors can average out. (2) The LandScan data indentify the units by numbers, whereas for first responders names of settlements are required.

Some groups are calculating losses due to earthquakes, using a subset of the LandScan data, called LandScan urban population, obtained using a mask based on MODIS satellite imagery. Here, we present a comparison of estimates of population affected by strong shaking at Intensities of VI and larger, where casualties result, using the three data sets: QLARM, LandScan, LandScan urban population. The comparison shows that, in the case of the M7.7 earthquake of 24/09/13 in Pakistan QLARM contains 34% more people than LandScan and 95% more people than LandScan urban population, which amounts to a difference of approximately 1 million. The conclusion is that data for the world population based on LandScan urban population should not be used for estimating losses in disasters.

The QLARM world population model has been constructed as follows. (A) The sum of the population in all countries are required to be within less than 3% of the number given by the CIA world fact book. (B) The available population in large cities is taken from internet sources. The sum of the population in these lists is always smaller than the population given by the CIA. (C) The missing population is distributed equally into the settlements for which Geonames gives coordinates and names. In the Pakistan test case, QLARM may overestimate the population in very small settlements in the test area at the expense of settlements elsewhere in Pakistan, or LandScan may underestimates the population in the test area assigning too many people to other parts of Pakistan. Both data sets contain the correct sum total of the population for Pakistan.

The solution to this problem may be scanning satellite images for locations of settlements with default population numbers and to adjust them, according to the image. We believe it would be time well spent, considering that the two good models for the world population, QLARM and LandScan, differ by 34% in case of a deadly earthquake.