[Slide 1]

Hello, I'm Georgiana Esquivias and today I will be presenting the method for developing a building exposure database for Tanzania.

[Slide 2]

We will go step-by-step through a series of flowcharts that assists in organizing and determining the type of data that is required for a Level-3 building exposure.

[Slide 3]

A Level-3 Building Exposure improves upon country level data by identifying sub-national variations in construction patterns, such as building types or building density, based on climate or cultural region-specific norms. As well, a Level-3 building exposure identifies major urban areas and enhances building counts or structural mapping scheme.

[Slide 4]

To develop the building exposure, the METEOR team utilized various sources of Earth Observation imagery, national statistics, reports, and journals. Additional source reference are found in the individual flowchart section.

[Slide 5]

The building exposure process is broken down into 7 flowcharts. Each flowchart asks a series of questions that helps identify the data and process step required for determining the Population, Structural distribution, number of buildings, if Dasymetric is possible, replacement cost, building heights, and building areas.

[Slide 6]

Here is a modified population flowchart that traces the data processing steps taken for Tanzania.

[Slide 7]

For Tanzania, the population spatial units was obtained by administrative units. There were 3 sources of population data provided by the Tanzanian government partners (NBS Tanzania 2012(1) and NBS Tanzania 2012(2)). An Excel spreadsheet and GIS dataset tracking population at the ward-level.

[Slide 8]

Along with a separate National Statistics report that contained the ‘persons per household’ data (NBS Tanzania 2013), which was important in estimating total number of buildings and building area.

Additionally, The Integrated Public Use Microdata Series (IPUMS) micro census data contains a population estimate, however IPUMS was used solely for determining building characteristics in Tanzania (Minnesota Population Center, 2019).

[Slide 9]

As we follow along the flowchart, it illustrates the merging of the GIS data with the tabulated population data.

The population data was then spread from the ward-level to a 3”-arcsecond grid, which is approximately 100m, using dasymetric mapping (FEMA 2015b).

In the spreading process, several Earth Observation data sets are used including global population data sets, which was ultimately used to estimate number of buildings.

[Slide 10]

The population data is from 2012, and there was no scaling to adjust for population growth.

[Slide 11]

The distribution of buildings by structure type typically receives the most concerted attention in a risk study, and this project was no exception. The sources used to develop the exposures include infield surveys, micro-census data, and expert opinion.

[Slide 12]

There were two methods for developing the structural distribution:

The first used IPUMS district census data to develop the mapping schemes using a combination of the building roof, wall, and floor material types. Each combination was mapped to a GEM construction type, and the number of households were used to develop a structural distribution by ward. The mapping schemes were developed at the individual ward level. In the few areas where IPUMS district information was not provided, the neighboring district was used as a proxy. The IPUMS floor, wall, and roof combinations were mapped to PAGER (USGS, 2008) and GEM (Brzev et al., 2013) classes by ImageCat staff and reviewed by Tanzanian measurement engineer Frank Mushi.

[Slide 13]

The second method includes an infield data survey collected by Humanitarian OpenStreetMap [or HOTOSM] (Humanitarian OpenStreetMap Team, 2019) which was used to establish the mapping scheme for each development pattern.

The process was conducted in the urban area of Dar es Salaam, as well as the neighboring Pwani region.

HOTOSM used a stratified sampling method and Bayesian updating approach (Porter, et al. 2014). Approximately 400 building locations were surveyed in each of the development patterns type, totaling 2,900 buildings.

The sites were surveyed using the HOTOSM tools by local staff trained in GIS techniques. Engineers were not present in the initial infield data collection. However, a second team of 4th year students majoring in Building Economics trained by Frank Mushi were present. This team served as validators of the data; whose main aim was to verify the tagging of buildings in priority areas.

[Slide 14]

Photographs of each structure accompanied each building assessment. The HOTOSM building survey data was verified by an ImageCat engineer Michael Eguchi and HOTOSM Mapping Supervisor Emmanuel Kombe. All of the mapping schemes were then mapped to the PAGER standard structural types and overlaid with the development patterns, shown in the next slide, to create a refined mapping scheme.

[Slide 15]

Based on a review by HOTOSM and ImageCat, it was determined that the IPUMS data captured the diversity of the construction practices throughout the country and the HOTOSM infield survey data captured the diversity of highly urbanized areas. Thus, the IPUMS mapping schemes were applied to Development Pattern 1, and the HOTOSM mapping schemes were applied to all other development patterns. The exception being Zanzibar, where due to the unique history of the region and homogeneous construction type, the IPUMS mapping schemes were used for the entire island.

[Slide 16]

Here we see the final results of the mapping schemes, mapped by administrative district for all of Tanzania. Note, the grey districts did not have counts in the IPUMS database, and so the nearest neighbors were used as proxy.

[Slide 17]

In Tanzania, the accuracy of the number of buildings was aided by several factors, which include OpenStreetMap building footprint database and the ward level person per household dataset.

[Slide 18]

The total number of buildings per structural class per 3-arcsecond grid cell was inferred using a combination of census data, development pattern sampling, building count samples, and aggregated OpenStreetMap building counts.

[Slide 19]

Population was converted to estimated number of households and disaggregated to the district level from the ward level using the data provided by Tanzanian Prime Minister's Office (Tanzania NBS(2), 2012; Tanzania NBS(1), 2012).

Ultimately, the number assumed at the district level will equal the estimated number of buildings established by the census.

However, many data sources are used to estimate the number of buildings at any given cell.

[Slide 20]

For each development pattern, a project analyst reviewed 3-arcsecond grids selected using a random number generator and identified cells where at least 90% of the buildings were represented in the OpenStreetMap database (Humanitarian OpenStreetMap Team, 2019).

[Slide 21]

The selected grids were used to train a regression model using the 3-arcsecond Sentinel-1 Global Human Settlement Layer (GHSL-S1).

In addition, WorldPop(1) (WorldPop & Center for International Earth Science Information Network [CIESIN], 2018), and the High Resolution Settlement Layer from the Facebook Connectivity Lab & CIESIN, (Facebook Connectivity Lab & Center for International Earth Science Information Network [CIESIN], 2016), and the Global Urban Footprint (GUF) from DLR (DLR Earth Observation Center, 2016) were used to spread the remaining number of buildings, or used in places where it appeared to be a better indicator than GHSL-S1 (Corbane et al.,(2) 2018). Sentinel-1 GHSL proved more effective in highly developed regions, whereas the GUF was more effective in rural areas.

[Slide 22]

The model was constrained using an aggregated 3-arcsecond OpenStreetMap building count (Humanitarian OpenStreetMap Team, 2019; ImageCat, 2019) grid as a minimum cell value.

As mentioned, the people per household were obtained from the ward level data.

[Slide 23]

Dasymetric mapping was used to spread the number of buildings from census administrative units (Tanzania NBS, 2012(1) and Tanzania NBS, 2012(2)) to the 3-arcsecond grid cell by way of statistical assessment of the moderate resolution Earth Observation (EO) data. To collect Earth Observation indicators of settlements and the density of buildings, various remote sensing data sets were used.

[Slide 24]

The following remotely sensed data product were used for the dasymetric mapping, the data was obtained from:

The US National Oceanic and Atmospheric Agency (NOAA), NASA, US Oak Ridge National Laboratory, The European Commission Joint Research Centre, The German Aerospace Centre, WorldPop, and OpenStreetMap.

[Slide 25]

These remotely sensed earth observation products and building footprint aggregates establish the distribution statistics for dispersing buildings by urban density and development pattern type. Each of the Earth Observation products, individually or in a combination, act as a weight to disperse the known population per Tanzanian district by structural type and determined development patterns to 3-arcsecond grid cell.

To prevent unpopulated areas from being considered as settlements, a mask was created using a minimum threshold value determined by visual inspection in Google Earth and with various specific Earth Observation datasets, including Sentinel-1 (Syrris et al., 2018), Global Urban Footprint (DLR Earth Observation Center, 2016), VIIRS (Earth Observation Group NOAA-NCEI, 2015), LandScan (Oak Ridge National Laboratory, 2009), and High Resolution Settlement Layer (Facebook Connectivity Lab and Center for International Earth Science Information Network [CIESIN], 2016). These minimum values correspond to even the sparsest human presence. This mask was used to subset other high-resolution

Sentinel-1 SAR mosaic which goes into the machine learning process that comes up with the development patterns. The Global Urban Footprint (GUF), Sentinel-1 based GHSL, and CIESIN-Facebook High Resolution Settlement Layer (HRSL) were not subset because these were good indicators for locating rural settlements.

[Slide 26]

To characterize building density in more populated areas, analysts digitized development pattern training polygons in the top 10 most populated cities in the country.

[Slide 27]

For these cities, the digitized vector data was used directly, rather than the classified development pattern grid. The delineations are digitized using the Google Earth area tools and saved as KMLs. The basemap vintage and source vary by region and zoom level. However, the most current high-resolution satellite images are used. The training polygons and the moderate resolution Earth Observation products described above are used in a machine learning process for assigning the development patterns throughout the country, which informs the estimated building density. The intensity of urbanity correlates to both the building density and the structural distribution. For Tanzania, the ImageCat engineer characterized 6 development pattern types.

[Slide 28]

Replacement cost for Tanzania are based on estimates from a study conducted by ImageCat for several African Nations which more accurately characterize the per unit building area replacement cost, particularly in rural Africa.

[Slide 29]

Replacement cost was determined by profiling structural durability by development pattern, as determined via remote sensing segmentation and digitization. Structural durability was classified into 3 tiers: temporary, semi-permanent and permanent.

The replacement costs for temporary and semi-permanent structures were obtained primarily from the resettlement action plan, and the replacement costs for permanent structures were taken from construction cost manuals.

The replacement costs for each development pattern was determined by applying a weight by mapping scheme for each structural type. Work was done under a separate grant for the Global Facility for Disaster Reduction and Recovery (GFDRR) and applied to Tanzania (Huyck & Eguchi, 2017).

[Slide 30]

Additionally, here a list of the sources used for determining the replacement cost.

[Slide 31]

A ratio was calculated between the durability classes, and expressed as a percentage of engineered construction.

For example, semi-permanent construction for a given country may represent 40% of the total permanent construction cost, and temporary construction may represent 20% of the total permanent construction cost.

The ratio approach is mathematically equivalent to taking the average value by durability class but has the added advantage of providing a basis for comparing the consistency of the approach across countries.

Once these prices were established, a matrix of the expected durability class given development pattern was established for all structure types.

[Slide 32]

For Tanzania, the permanent material replacement cost was $339 US dollars, semi-permanent is $141, and temporary material replacement cost is valued at $61 US dollars.

[Slide 33]

Here we have the estimated replacement cost values for Dar es Salaam.

[Slide 34]

The building height values for all development patterns were directly obtained from the field survey data collected by HOTOSM (Humanitarian OpenStreetMap Team, 2019).

[Slide 35]

This established the building height by structure type per development pattern type. Rural areas are assumed to be 100% low rise. The HOTOSM survey was reviewed and validated by an in-house engineer using images provided of the individual buildings.

[Slide 36]

The building area values were directly obtained from the field survey data collected and the aggregated building area raster.

[Slide 37]

The total building area was calculated using a combination of the Humanitarian OpenStreetMap Team (HOTOSM) survey data (Humanitarian OpenStreetMap Team, 2019), and 3-arcsecond aggregated OSM building area raster data sets created by ImageCat (ImageCat, Inc., 2019).

For the rural development pattern, the average building footprint area from the 3-arcsecond aggregated OSM raster was used to determine the average total building area; assuming all rural developments are single story.

For all other delineated development patterns, the HOTOSM in-situ building survey was used to establish the total building area by structure type per development pattern types using the surveyed building footprints and height values.

[Slide 38]

And finally, there was an adjustment of building size to reflect the urban development pattern.

[Slide 39]

There is a great deal of information that goes into creating a building exposure. The aim of these flow charts is to:

1) Assist future researchers in updating or refining the exposure database as a whole or finely tune an individual section with improved data.

2) Help assist in organizing metadata collection

3) Help work through the logic of the processing steps.

[Slide 40]

That concludes the flowchart section. We hope that these flowcharts assist you and future teams in conducting, updating, or refining a level-3 building exposure.

[Slide 41]

Thank you and have a good day.