Methodology for Hazard Exposure and Vulnerability Assessment in the Ganges-Brahmaputra (GBM) River Basin

Working Paper prepared for the ESCAP Asia-Pacific Disaster Risk Report, 2019

Daniel Clarke, Jiwon Seo, Yichun Wang

Purpose/Background:

For use in the forth-coming ESCAP regional Disaster Risk Report, a multi-country assessment study was commissioned for the Ganges-Brahmaputra (GBM) River Basin, with an emphasis on investigating the relationship between disaster risk and poverty. That is, to assess vulnerability (as approximated by economic poverty) of the population as a component of disaster risk, particularly for flood hazards in the GBM.

The GBM river basin contains regions of 5 countries (Bangladesh, Bhutan, China, India, Nepal). The GBM has some of the most densely populated cities of the world, especially along the rivers and in the lower portion of the river basin. GBM is also home to a very large rural population and extremely high number of households experiencing extreme poverty.

As a large multi-country, yet geographically well-defined catchment area,[1] the GBM is a good case for testing new approaches to compilation and integrated analyses of official data sources for cross-border analyses of disaster risk. The study provides an opportunity to find answers to an interesting and very important analytical question about vulnerability to disasters in the region – i.e. do vulnerable populations (like the poor) also have a higher likelihood of exposure to hazards like floods? This study also provided a chance to test and review new methods for integration of geospatial and statistical data across multiple data sources and across multiple countries towards harmonized statistics on hazard exposure and vulnerability.

Coverage:

Availability of data vary for each of the five countries with territories in the GBM. For practical purposes, the results of the study ultimately excluded the mountainous area in the Tibet region of China, although this area is included in some maps of the GBM. The entire areas of Bangladesh, Bhutan and Nepal are included in the scope of this study along with selected regions of India within the GBM basin. Since we did not have access to comparable household survey data with geo-referencing for Bhutan, the poverty exposure estimates portion of the results are limited to Bangladesh, India and Nepal only (in other words, for Bhutan we looked only at hazard exposure for the general population).

Data:

[1] Although there is not a single standard base-map setting boundaries for the GBM basin in this study, techniques are available for developing a precise biophysical-based definition for this large area. Moreover, coverage of access to comparable population and social data for the entire GBM region will improve over time.

The first objective of this methodology is essentially to integrate datasets of two broad types: remote sensing data in the format of grid-based (raster) geospatial files with counts of population and results of household surveys indicating wealth of households.

The project benefitted greatly from recent availability of Demographic and Health surveys (DHS) with geo-referencing and production of geo-covariates from the international DHS Programme.[2] The programme provides free access to random cluster-sampled household data for three of the four countries (Bangladesh, India and Nepal) all from within the last 3 years.

In addition to DHS Programme household survey data and associated geospatial information for Bangladesh (2014), India (2015-16) and Nepal (2016), other key data sources used are:

- Official population statistics as published by national statistics offices of the governments of the four countries and available by district (or equivalent, usually admin. Level 02 or 03) based on most recent census in each country.
- Global Urban Footprint (GUF, gridded global map of built-up areas (from the German Space Agency (DLR), which has changed names to Global Human Settlement Layer
- Global Climate Change Initiative (CCI) Land Cover map, global gridded map of land cover made available by the European Space Agency using a classification system consistent with UN-adopted standards developed by FAO
- The Earth Observations Group (EOG) at the National Oceanic and Atmospheric Administration (NOAA) and National Aeronautics s and Space Administration (NASA) of the United States composite images of visible light at night (VIIRS) Day/Night Band (DNB)
- Probabilistic hazard maps (raster layers) from UNISDR GAR 2015 Risk Data Platform

Methodology:

Our method involves a 2-step process, the first is to estimate population density by grid-cell location of for each administrative area and the second step is to build on the results from step one to estimate location of poverty via grid-assimilation and extrapolation of poverty information from DHS surveys in each country.

Step 1: Predicting location of population, grid-based estimation of population density

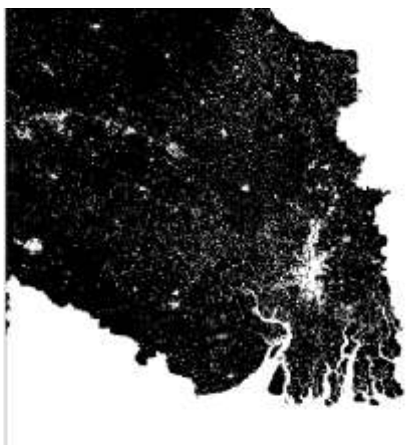
Censuses are conducted at national scale and there are no multi-national household surveys in the GBM. Most recent population figures, based on census are available from official sources at roughly the district level (or equivalent, Adm. 02 or Adm 03). Administrative areas (shapes or vectors in GIS terms) do not have standardized size or features and population is distributed unevenly across space. Thus, a grid-based assimilation model is needed to disaggregate the population data and assimilate the statistics into a standard grid system (raster) according to our best estimate of location of that population within each district. It is a method of disaggregation of census statistics, based on a probabilistic assessment of population density within the grid system for each administrative area.

[2] https://dhsprogram.com/

Aggregated census results are the starting point for the grid-based modelled distribution of the population and the grid-based assessments, thus the results aligned with the official population statistics if re-summed at the administrative regions level. The main inputs for this distribution of population density are built-up areas (as interpreted from remote sensing imagery) and other land use and land cover characteristics of the landscape.

A complete description of the methodology and a step-by-step manual for this gridded assimilation and estimation of gridded population density is available on the website for the Disaster-related Statistics Framework (DRSF).[3] This method was applied using the best available district-level (adm. 02 or adm. 03) population data for the four countries in this study and then merged into a common gridded population system for the GBM basin.

Sample result from gridded population estimate (population density per grid cell):



The result is a grid system in which each cell contains continuous values greater than or equal to zero, representing the estimate population density per cell. The estimation is conducted using a 100mX100m standard grid system[4] for the region (other grid sizes, including higher resolution, are possible, but this resolution is shown to produce reasonable results in previous pilot studies).

The portion of the population exposed to flood hazard is defined simply as the counts of population (per grid cell or summed across grid cells for each administrative region) within areas with high probability of impact from a given flood scenario. Thus, the overlay between the gridded population density and the flood hazard area is the population exposure to flood hazards. This overlay produces maps (images or raster files) and can be summarized into counts of population exposed to flood hazard at the district level -(or other scales, as relevant – an advantage of this model is scale of analysis can be fully flexible).

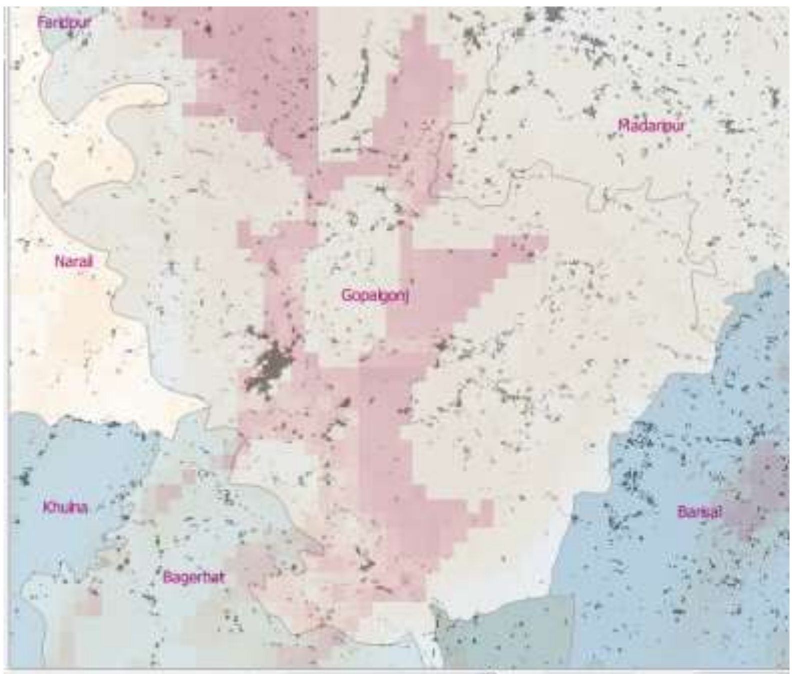
Defining hazard areas

[3] http://communities.unescap.org/asia-pacific-expert-group-disaster-related-statistics/content/drsf

[4] Results are later aggregated (via bilinear interpolation methodology) for integration with flood hazard areas and final presentation of the results at resolution 1kmX1km

The GAR 2015 Risk Data Platform 100-year flood scenario map is used for the analyses of exposure in the GBM river basin for this study. The probabilistic flood hazard map is measured in terms of meters (depth of water). Thus, in order to define the hazard area, a threshold value (depth in terms of meters) must be decided. Different hazard exposure studies have defined the threshold differently (and they are not always documented). The selection of hazard threshold depends somewhat on the purpose of the analysis. For this study, the threshold was set at 1 meter. One meter or more inundation of flood water would result in serious, if not complete, losses for the exposed population, especially for the most vulnerable populations (i.e. the poor) and their assets.

Result of gridded population (grey-black) overlaid with flood hazard areas (pink—red):

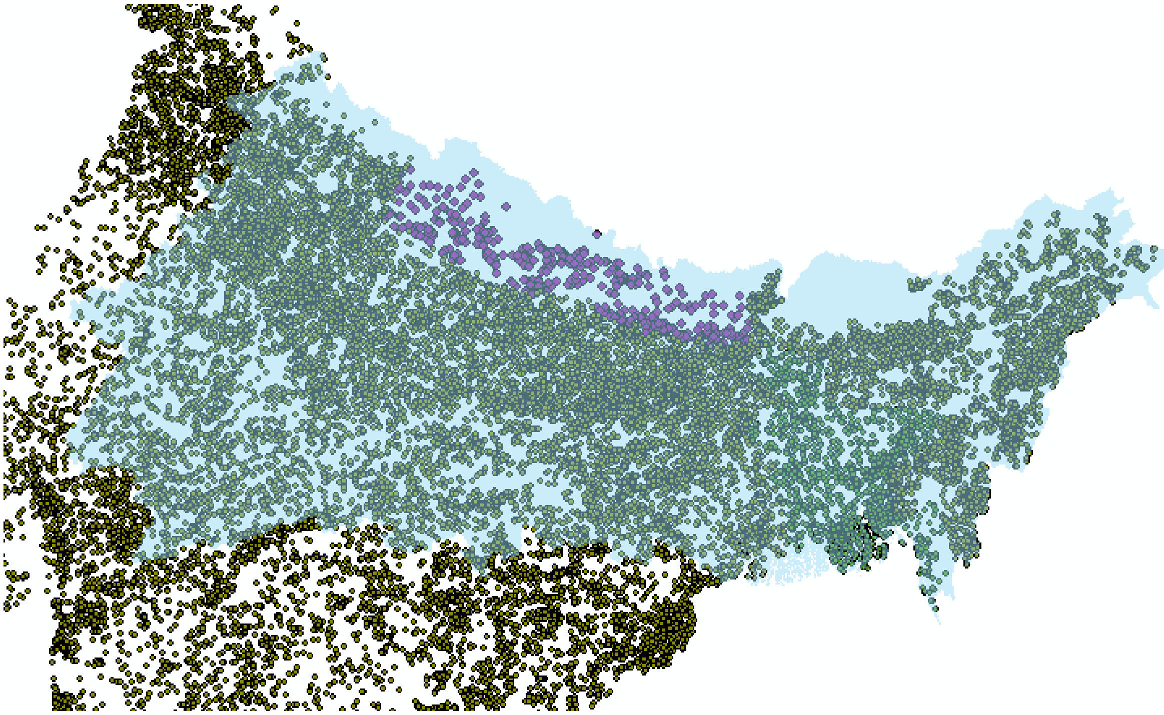


The resolution for the GAR 2015 Data Platform gridded hazard map datasets is 1km by 1km cells, therefore in the final analysis the gridded population density estimates are aggregated into this resolution for calculating the summary statistics.

Step 2: Predicting location of poverty

In the DHS Programme, published survey results include geographic referencing for the sampling clusters, or primary sampling units (psu's), which are geographic areas of clusters of households which were selected randomly for the survey using stratified random sampling methodology. Location of the clusters in the DHS stratified cluster sample are provided as a GIS vector file, with a small amount (2-5 km) shift (or distortion) to the precise coordinates as an added protection for confidentiality of respondents within each PSU.

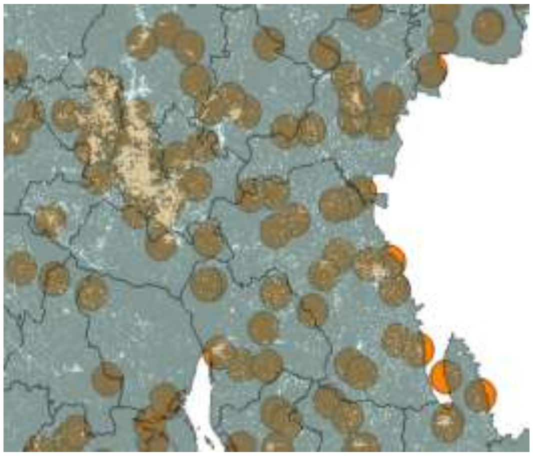
PSUs (clusters) for DHS Bangladesh (2014), India (2015-16), and Nepal (2016)

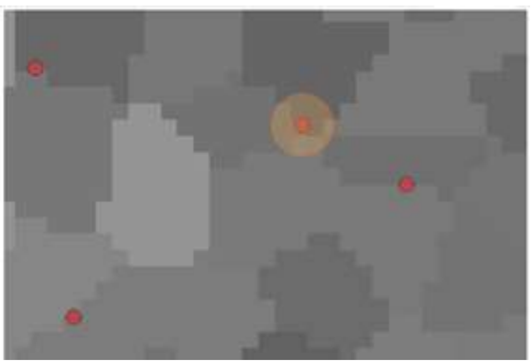


To accommodate for the intentional shift to the precise location of the sampling units, a circular buffer area is created (at 1- 5 km radius) around the reported locations of the sample. This area is used for developing a smoothed assessment of characteristics of the landscape in sampled areas based on earth observation data (i.e. the nightlights and built-up areas).

The assessment for each PSU buffer area is used as the inputs in the extrapolation model for predicting the location of relevant results from the survey.

Approximating location (buffer area) of DHS Sample clusters for overlay with earth observation data:





The buffer areas (orange circles) around sampling units (clusters) do not cover the entire areas of the countries. Thus, the aim is to use the information from these areas and extrapolate to the remaining areas of each administrative area. Since these sampling units are designed to be representative (for the purposes of the survey), we expect reasonably good results from the extrapolation to the areas not covered by the survey directly.

The extrapolation method will work best for variables with a significant variability across the geo-covariates (i.e. built-up areas and nightlights). In general, poverty is a good candidate for this kind of extrapolation and previous studies (see, e.g., Neal Jean et.al., 2016) have demonstrated a possibility for using correlations between poverty and nightlights for probabilistic mapping via machine learning techniques.

The gridded population estimates from Step 1 provide a base map for identifying locations of poor households as modelled using earth observation data sets to extrapolate location and then fitting the summed values for each district to the official aggregates. The latter adjustment helps correct for a potential inherent bias from using nightlights as a predictive variable, for example it may underestimate urban poverty (the model tended to underestimate poverty for relatively urbanized districts, but this was corrected by fitting the results to official aggregates).

Poverty Threshold

Multiple definitions or thresholds are used to define and measure poverty at national and international scales. As this is a multi-national study, we decided to use the well-known international poverty line, which was developed by the World Bank and is used for measuring extreme poverty in the Sustainable Development Goals ($1.90 PPP per day). As the international poverty line is based on consumption expenditure, we make the assumption, for the purpose of this analysis, that the household-level wealth index produced by the DHS survey is generally a good predictor for the shares of households experiencing poverty in the consumption measure. Even if the correlation between the two poverty measures do not hold up in all individual cases, the relationship need only hold on average for the aggregated populations. Moreover, this assumption only affects the estimated total number of poor used for fitting the model to the survey results and does not affect the extrapolation model itself.

Extrapolation Model

Several machine-learning models for predicting geographic variance in the survey results for household wealth were tested previously for the Bangladesh datasets.[5] Ultimately, it was found that the visible nightlights dataset is a strong predictor of location of poverty in the GBM using the Random Forest machine learning technique.

Essentially the model mines wealth data by household and within each sampling cluster and remote sensing data describing the landscape in each cluster area (in this case visible light at nighttime) to predict probable location of poverty in gridded population across the three countries. The output are gridded probability factor for distributing results from the survey across space.

Based on the above two-step modeling procedure, and a few simple arithmetic calculations across the modelled geospatial layers overlaid with the hazard area map produces estimates for 2 main indicators:

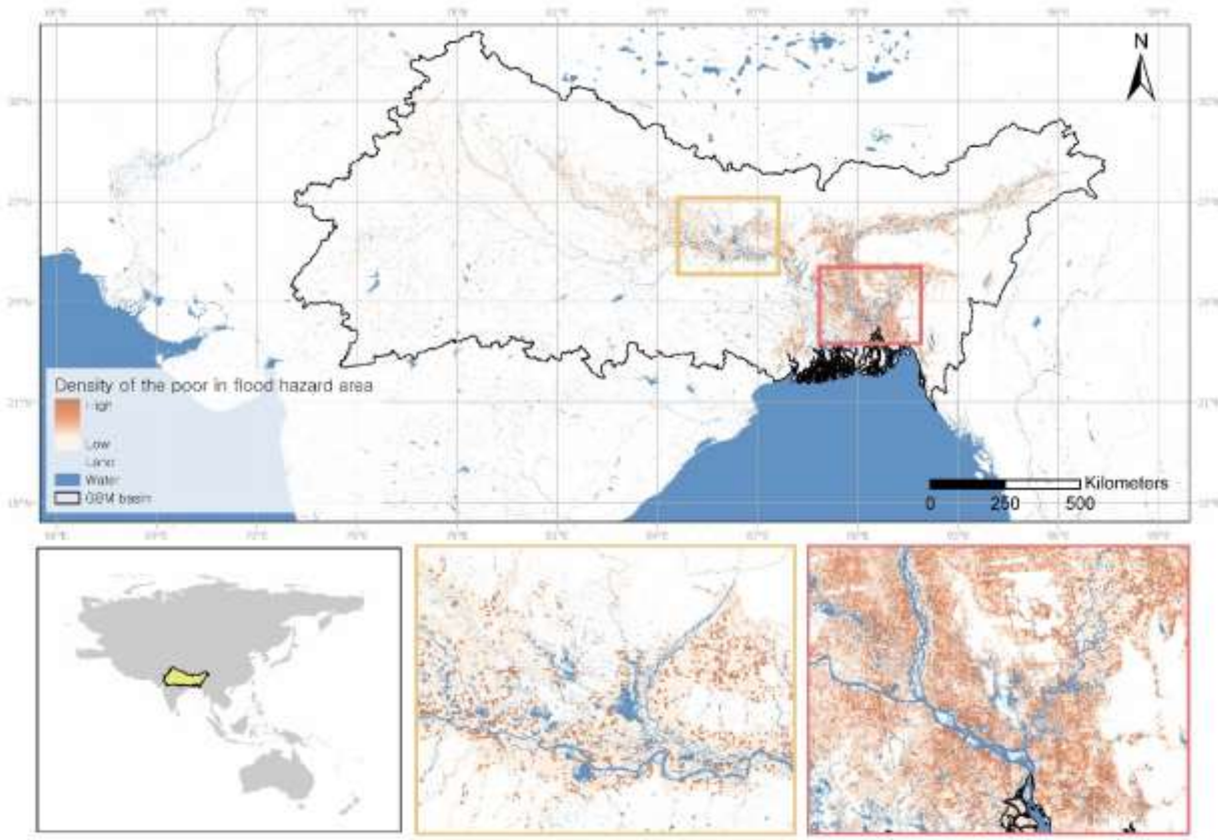
1. **Population (counts) and rates (percentages) exposed to hazards**
2. **Population exposure, which is also experiencing poverty (vulnerable populations, counts and percentages)**

For the cases where the percentage of poor exposed to the hazard area is greater than the exposure of the general population, it means that the poor are more likely, on average, to be exposed to flood hazard. This is the result for the GBM as a whole and for most, though not all, or its administrative regions.

[5] See explanation and results of pilot tests for Bangladesh in Yichun Wang (2019)

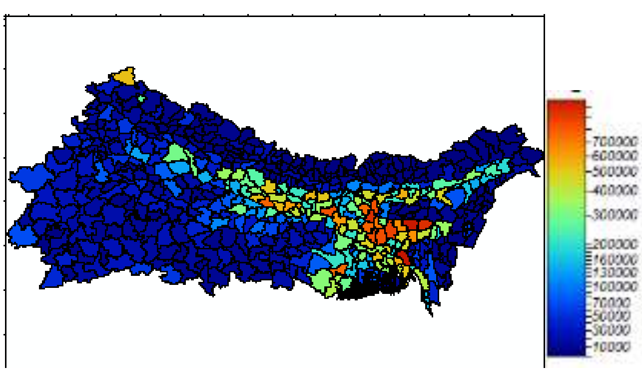
Summary of results for exposure to flood hazard areas by administrative regions:

			flood hazard exposure rate, general population (%)	*flood hazard exposure rate, among poor (below intl' povery line)*				*flood hazard exposure rate, general population (%)*	*flood hazard exposure rate, among poor (below intl' povery line)*
Bangladesh	*Division*				***Nepal***	*province*			
	1	*Barisal*	*17.396*	*20.144*		*1*	*Kosi*	*19.830*	*11.874*
	2	*Chittagong*	*73.653*	*79.561*		*2*	*Jankpur*	*10.668*	*7.310*
	3	*Dhaka*	*74.608*	*83.843*		*3*	*Bagmati*	*11.985*	*13.140*
	4	*Khulna*	*40.650*	*43.207*		*4*	*Gandaki*	*14.237*	*20.574*
	5	*Rajshahi*	*73.957*	*73.260*		*5*	*Lumbini*	*18.707*	*22.867*
	6	*Rangpur*	*65.623*	*68.389*		*6*	*Karnali*	*8.595*	*5.622*
	7	*Sylhet*	*84.151*	*89.089*		*7*	*Far West*	*11.864*	*24.510*
		BGD	*66.624*	*71.398*			*NPL*	*14.114*	*15.431*
Bhutan	*Prefecture*				***India***	*State*			
	1	*Bumthang*	*10.249*			*1*	*Arunachal Pradesh*	*20.914*	*16.989*
	2	*Chhukha*	*16.328*			*2*	*Assam*	*69.333*	*68.678*
	3	*Dagana*	*14.965*			*3*	*Bihar*	*46.564*	*44.514*
	4	*Gasa*	*11.336*			*4*	*Chhattisgarh*	*5.130*	*5.241*
	5	*Haa*	*6.325*			*5*	*Haryana*	*7.546*	*10.629*
	6	*Lhuentse*	*9.466*			*6*	*Himachal Pradesh*	*41.934*	*45.965*
	7	*Monggar*	*5.920*			*7*	*Jharkhand*	*10.632*	*8.883*
	8	*Paro*	*16.398*			*8*	*Madhya Pradesh*	*5.294*	*6.029*
	9	*Pemagatshel*	*8.163*			*9*	*Manipur*	*39.939*	*38.395*
	10	*Punakha*	*15.336*			*10*	*Meghalaya*	*21.884*	*22.151*
	11	*Samdrupjongkhar*	*4.281*			*11*	*Mizoram*	*8.369*	*10.497*
	12	*Samtse*	*15.734*			*12*	*NCT of Delhi*	*38.943*	*46.416*
	13	*Sarpang*	*4.441*			*13*	*Rajasthan*	*5.383*	*6.637*
	14	*Thimphu*	*21.980*			*14*	*Sikkim*	*0.708*	*14.837*
	15	*Trashigang*	*8.324*			*15*	*Tripura*	*29.545*	*27.341*
	16	*Trongsa*	*9.864*			*16*	*Uttar Pradesh*	*15.606*	*19.253*
	17	*Tsirang*	*8.508*			*17*	*Uttarakhand*	*15.629*	*23.149*
	18	*Wangduephodrang*	*9.840*			*18*	*West Bengal*	*25.793*	*26.251*
	19	*Yangtse*	*11.595*				*IND(GBM only)*	*23.305*	*26.903*
	20	*Zhemgang*	*8.677*						
		BTN	*12.221*			***GBM***		***33.801***	***38.089***



Disclaimer: The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations.

The gridded results presented in the box excerpt from the ESCAP 2019 Asia-Pacific Disaster Report above can also be summarized in tabulations as shown in the sample table above and in the district-based map below.



Suggestions for Further Research

The results previewed above are calculated with one common flood hazard threshold for the hazard map, a 100-year flood scenario and a depth of 1 meter. Other possible threshold values were also tested and changing threshold values obviously affects the results in absolute terms because changing the threshold effectively changes the size (grows or shrinks) the hazard area. Other flood scenarios and also other depth threshold values could be tested towards development of a standardized approach (or standard principles) for measuring population exposure for flood hazard. Moreover, hazard maps for other hazard types, such as earthquakes, which are also prevalent in this region, should be assessed for exposure analyses as the data have different units and are measured differently from flood or other hazards. Are the poor more likely than the general population to be exposed to earthquake hazard, or is this only the case for floods? Also, the geographic scope (GBM river basin) could be defined more precisely and, if feasible, more data should be included in the measurement from Bhutan and areas of southern China to improve the coverage of the study.

A further potential refinement, which might strengthen accuracy is to fit the grid-based modelled assessment to the results of small area estimation studies, that have matched census data with data from the same survey used in the grid-assimilation modelling. The outcome may be considered a hybrid approach**,** incorporating advantages of SAE studies with the advantages from the predictions based on earth observation data.

Building on the results of this study and connected research by Yichun Wang, (2019), further testing of use of the tree model in combination with geospatial covariates and the wealth index from DHS surveys in multiple countries. Replicating the methodologies with other countries is likely to reveal improvements to calibrate the model and help develop recommendations for a harmonize and transparent methodological guidance for disaster exposure and vulnerability assessment.

References

ESCAP. 2018. *Disaster-related Statistics Framework* Final Draft from the Expert Group on Disaster-related Statistics in Asia and the Pacific. May, 2018. http://communities.unescap.org/asia-pacific-expert-group-disaster-related-statistics/content/drsf

German Aerospace Centre (DLR): Global Urban Footprint. Accessed in 2018, https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-9631/16580_read-40465/

ICF. 2004-2017. Demographic and Health Surveys (various) [Datasets]. Funded by USAID. Rockville, Maryland: ICF [Distributor].

NASA Earth Observatory/NOAA NGDC Visible Light at Nightime Composite image of earth. Accessed Jan. 2019, https://www.nasa.gov/mission_pages/NPP/news/earth-at-night.html

Neal Jean, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell3, Stefano Ermon. 2016. *Combining satellite imagery and machine learning to predict poverty.* Science Magazine 19 Aug 2016: Vol. 353, Issue 6301, pp. 790-794

UNISDR.2015. Global Risk Assessment (GAR) Data Platform 2015.

Yichun Wang. 2019 DHS and Geo-covariates data integration Case study for Bangladesh survey 2014 , Yichun Wang (Feb. 2019), research for UNESCAP http://comunities.unescap.org/asia-pacific-expert-group-disaster-related-statistics/content/drsf