American Economic Review: Papers & Proceedings 2017, 107(5): 441–445
https://doi.org/10.1257/aer.p20171052

Validating Migration Responses to Flooding Using Satellite and Vital Registration Data†

By Joyce J. Chen, Valerie Mueller, Yuanyuan Jia, and Steven Kuo-Hsin Tseng*

Quantifying flood-driven migration is crucial for governments and donors, given the increasing frequency of such events under global climate change as well as their potential impact on host economies and international security. However, existing work suggesting the potential for mass exodus remains largely unsubstantiated over longer time periods and larger geographic areas. Indeed, current pioneering work in the environmental migration literature suggests that the capacity for migration is much more limited, given that many lack the means to finance relocation and the social networks needed for finding employment (Bryan, Chowdhury, and Mobarak 2014).

Gray and Mueller (2012) first challenged the conventional narrative of "environmental refugees" in Bangladesh, finding a larger effect of drought-related crop failure than flooding on permanent migration. However, the study is limited to selected sites, while environmental exposure and migration varies with local characteristics, such as proximity to inland/coastal locations. Furthermore, their measures of crop failure and flood events are self-reported, reflecting subjective factors such as recall bias and reference dependence.

Tackling the external validity problem, Lu et al. (2016) track population movements around Cyclone Mahasen in 2013 using mobile phone network data. They find that population flows are largely unchanged by this event. But, lacking knowledge of who is using these phones, this approach cannot identify vulnerabilities of specific populations, a key aspect of targeting social protection and relief. The focus on a single event additionally limits the generalizability of the findings to disasters with varying duration and intensity.

We build on these studies by linking nationally representative data with objective measures of flooding to shed additional light on the migration-flooding nexus in Bangladesh. Household-level migration data are drawn from vital registration records, which offer the advantage of monitoring mobility among communities spanning the entire country over nearly a decade. To construct objective measures of flooding at each household's subdistrict (*upazila*) of origin, we use satellite data, similar to Guiteras, Jina, and Mobarak (2015). Typical proxies of flood exposure are rainfall extremes, measured by converting raw precipitation data into anomaly or percentile variables (Mueller, Gray, and Kosec 2014). We show how inferences on flooding displacement change when using an objective flooding measure versus proxies commonly adopted in the literature.

I. Data

Migration.—Our data are drawn from the 2003–2011 Sample Vital Registration System (SVRS), an annual survey of over 200,000 households conducted by the Bangladesh Bureau of Statistics. Samples are nationally representative, in order to provide inter-censal

*Chen: The Ohio State University, 324 Agricultural Administration Building, 2120 Fyffe Road, Columbus, OH 42310 (e-mail: chen.1276@osu.edu); Mueller: International Food Policy Research Institute, 2033 K Street, NW, Washington, DC 20006 (e-mail: v.mueller@cgiar.org); Jia: The Ohio State University, 275 Mendenhall Lab, 125 South Oval Mall, Columbus OH, 43210 (e-mail: jia.102@osu.edu); Tseng: Center for Space and Remote Sensing Research, National Central University, No.300, Jhongda Road, Jhongli City, Taoyuan, 32001, Taiwan (e-mail: khtseng@csrsr.ncu.edu.tw). We thank Boris Braun, Craig Jenkins, Mushfiq Mobarak, and C. K. Shum for many helpful comments. Kathryn Dotzel and Mekamu Kedir Jemal provided excellent research assistance. Support from the National Science Foundation via the Belmont Forum/IGFA Program (ICER-1342644) is gratefully acknowledged.

† Go to https://doi.org/10.1257/aer.p20171052 to visit the article page for additional materials and author disclosure statement(s).

demographic statistics representative at the district (*zila*) level. Data on migration is recorded for all individuals who have either been away for at least six months or left due to household displacement or marriage. This under-states overall out-migration, as temporary moves and migration by entire households are not captured in our data.

Rainfall.—Data on rainfall are drawn from two gridded monthly products, the Tropical Rainfall Measuring Mission (TRMM) and the University of Delaware (Willmott and Matsuura 2012), as well as 34 in situ weather stations operated by the Bangladesh Meteorological Department. Subdistrict centroids are linked to the nearest grid point and weather station.

In situ data, when available, has the advantage of more accurately capturing rainfall, but only within close proximity of the station. And the placement of weather stations and temporal resolution may be correlated with omitted variables (Auffhammer et al. 2013). Gridded datasets have the advantage of using balanced panels of information from nearest weather stations, satellites, and climate models to fill in data gaps (Donaldson and Storeygard 2016). However, the accuracy of these products is sensitive to the underlying data. In the case of Willmott and Matsuura (2012), the 124 grid points for Bangladesh are based on only 10 weather stations. In the case of TRMM, underlying data are based on satellite images, so only moderate to high rainfall rates can be detected, due to sensitivity limitations (National Research Council of the National Academies 2007). When validated against in situ rain gauges, TRMM is found to overestimate precipitation during the pre-monsoon period and in dry regions and underestimate precipitation during the monsoon period and in wet regions (Islam and Uyeda 2007). We therefore construct flood proxies using each of the three rainfall measures.

Epanechnikov kernel densities of the subdistrict correlations of annual precipitation across data products show positive correlations for the majority of the distributions (see the online Appendix). The TRMM measure has an average correlation of 0.66 with the Delaware measure, and 0.63 with the weather station measure. There is a more modest correlation of 0.51 between the weather station and Delaware precipitation measures. We examine how these discrepancies might translate into different predictions for flooding displacement using the model described below.

Inundation Extent Measure.—Data are drawn from NASA Moderate Resolution Imaging Spectroradiometer (MODIS) satellites at 500m resolution. We construct the Modified Normalized Difference Water Index (MNDWI) (Xu 2006), which differentiates water and non-water features based on surface reflectance.[1] A pixel is defined as water if MNDWI > 0.1.[2] Upazila-level measures are based on the maximum percentage of water pixels over all eight-day composites in the period. To differentiate water bodies from inundation, we look at the difference in water coverage between the monsoon (July–December) and dry (January–March) seasons.

Using the TRMM measure as a reference point in Figure 1, we see in panel A that a significant portion of the sample exhibits either negative or low positive correlation between rainfall and the satellite-based measure of inundation. This likely reflects the complex hydrology of Bangladesh, a deltaic plain formed at the confluence of the Ganges, Brahmaputra, and Meghna Rivers. Water levels and flooding are, therefore, highly dependent on not only local but also upstream precipitation. However, in areas with high river density, local precipitation may play a relatively larger role. Indeed, in panel B, we see a much stronger correlation among the top 40 percent of subdistricts with respect to river density.[3] There are far fewer observations in the negative quadrant and an overall shift in the distribution to the right. This pattern is evident for all rainfall products (see the online Appendix). In the absence of satellite-based measures, this suggests rainfall proxies will perform better in areas with greater surface water coverage.

We also examine whether monsoon precipitation may be a better proxy for flooding, given that this season accounts for well over half of

[1] Because surface images are obscured by cloud cover, these pixels are first removed (Xiao et al. 2006).

[2] This measure has been found to provide the most accurate detection of flooded areas, compared to other commonly used band ratio indices and has the most stable threshold (Ji, Zhang, and Wylie 2009).

[3] Defined as river length as a proportion of total area. Derived from Global Lakes and Wetland Database.

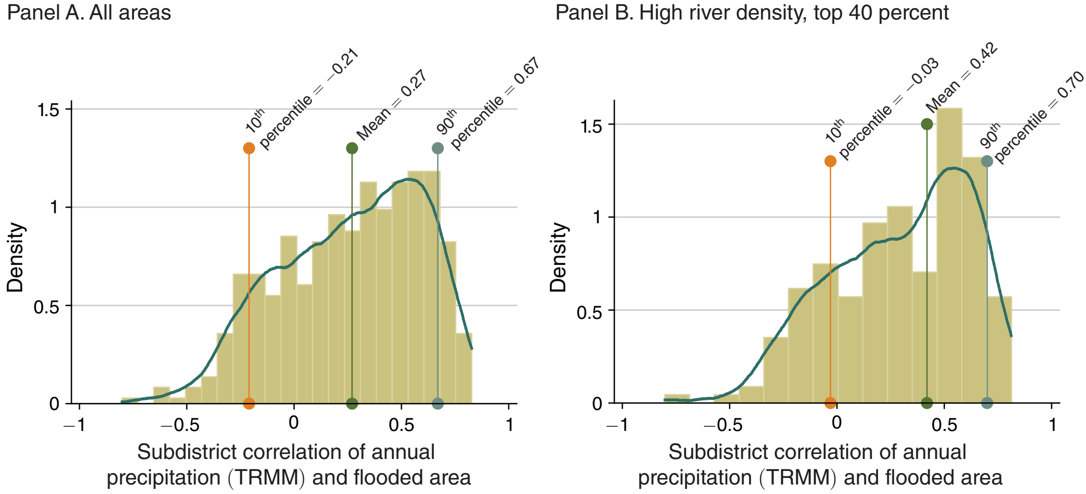

Figure 1. Epanechnikov Kernel Densities of Correlations between TRMM Precipitation and Inundation Measures

yearly rainfall in most parts of Bangladesh. In fact, total annual precipitation has a substantially stronger correlation to the satellite-based flood measure, with the exception of the University of Delaware measure (see the online Appendix). Longer-term precipitation measures better reflect overall water balance, and the limitations of satellite products in detecting rainfall across seasons suggest that annual precipitation is generally a better proxy for flooding.

II. Empirical Model

We employ a linear probability model to estimate the effect of flooding in location j at $t-1$ on the probability of a household h having at least one migrant, M, at time t:

$$(1)\quad M_{hjt} = \alpha \mathbf{X}_{hjt} + \sum_{m=2}^{5} \beta_m F_{mjt-1} + \gamma_t + \epsilon_{hjt}.$$

We adopt the convention of looking at quintiles of flooding, F_2, F_3, F_4, and F_5, to account for nonlinear impacts. Implicit in $\mathbf{X}$ are variables that affect migration decisions, such as household demographics and wealth (full list detailed in the online Appendix) and climate (lagged quintile categorical variables for growing degree days over the growing season, 30-year running averages for degree days and annual precipitation). We also control for competing time-specific influences on migration by including a time fixed effect γ_t. Standard errors are clustered at the primary sampling unit to allow for correlation in unobserved factors influencing migration.

III. Results

Table 1 displays the point estimates from (1) when including the rainfall-based proxies (derived from weather stations and gridded data products) and the preferred satellite-based measure of flooding extent. Looking at the full sample (columns 1, 3, 5, and 7), we find significant negative associations between migration and the fourth and fifth quintiles across all flooding measures. This corroborates earlier longitudinal analysis using self-reported flooding measures (Gray and Mueller 2012). The probability of a household having at least one migrant under an extreme flooding scenario compared to no flooding declines by 0.4 to 1.8 percentage points, which is sizable given a sample mean of 5 percent. The satellite measure (column 7), however, reveals effects of localized flooding as well, with significant negative effects observed throughout the distribution, albeit smaller in magnitude for lower quintiles.

Given stronger correlation between rainfall proxies and satellite-based measures in areas with high river density, we also consider this sample restriction in our regressions reported in columns 2, 4, 6, and 8. The significant negative effects in the fifth quintile are still evident for

Table 1—Migration-Flood Relationship

	Station	Station	Del.	Del.	TRMM	TRMM	MODIS	MODIS	MODIS
Q2	0.002	0.010	0.001	0.002	−0.002	0.000	−0.004	−0.003	−0.005
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
Q3	0.001	−0.002	−0.002	0.004	−0.001	−0.001	−0.003	−0.000	−0.003
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Q4	−0.004	−0.005	−0.007	−0.004	−0.009	−0.009	−0.006	−0.001	−0.003
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)
Q5	−0.006	−0.003	−0.016	−0.011	−0.018	−0.014	−0.008	0.004	0.004
	(0.003)	(0.004)	(0.003)	(0.005)	(0.003)	(0.005)	(0.003)	(0.004)	(0.003)
Sample	Full	HRD	Full	HRD	Full	HRD	Full	HRD	HRD

Notes: $N =$ 1,931,954 for full sample and 809,362 for HRD (high river density) sample. Del. = University of Delaware. Includes controls for household demographics, wealth, degree days, historical climate conditions, and year fixed effects. Standard errors clustered at primary sampling unit and presented in parentheses.

two of the three rainfall proxies but not for the satellite-based measure. In these areas, households appear to be responding to rainfall and flooding in very different ways. One possible explanation is that, because much of the country experiences annual flooding, households may not perceive or respond to these events as shocks. To explore this possibility, we repeat the analysis with the satellite-based measure normalized by the subdistrict-specific mean and standard deviation over all years and report the results in column 9.[4] The point estimates remain quite similar but precision increases substantially. We now observe a significant negative effect for the first quintile. This suggests that, in high river density areas, households have already adapted to annual flooding patterns, so the raw measures exhibit little correlation with out-migration. However, relatively small deviations from the norm do still produce a migration response. The fact that large flooding shocks have no effect for high river density areas but significant negative effects for the country as a whole perhaps suggests that households facing recurrent floods have few remaining options for adaptation.

IV. Discussion

Using nationally representative data on migration in Bangladesh, we find a significant negative effect of extreme flooding on the probability of a household sending out at least one migrant in the previous year. Individuals may be more likely to be trapped than internally displaced by floods. An alternative explanation is that the broader benefits from extreme flooding outweigh the short term costs. Flooding can improve overall soil quality and yields in subsequent crop cycles (Banerjee 2010), potentially increasing the opportunity cost of an absent family member. We show that results using proxies for flooding from gridded datasets are qualitatively similar to those using satellite-based measures when focusing on the top quintiles. However, the coarseness of these proxies may be masking other sources of variation. Only the satellite-based indicator captures the effects of localized floods (represented by the lower quintiles), which are driven by proximity to rivers, topography, and other conditions unrelated to rainfall.

Specifications using satellite-based indicators alone convey a non-monotonic relationship between migration and flooding. Even modest flooding (second and third quintiles, 3–17 percent of the subdistrict) significantly deters migration, but there is a markedly larger effect in the fourth and particularly fifth quintiles. Broader exposure to flooding throughout a subdistrict can reduce opportunities to access credit and/or utilize risk pooling mechanisms to finance migration. However, correlations between flooding and migration appear to be quite fragile and vary substantially across areas. In areas with high river density, we continue to observe a negative relationship between rainfall and migration, but this is evident for

[4] We cannot control for the historical average, as we do for rainfall, because MODIS satellites were not launched until 2002.

our satellite-based flooding measure only after standardizing within subdistricts and in different segments of the distribution. Despite substantial correlation between rainfall and flooding, our results suggest that households experience these two phenomena quite differently, in part because many may have already adapted to annual flooding patterns.

REFERENCES

Auffhammer, Maximilian, Solomon M. Hsiang, Wolfram Schlenker, and Adam Sobel. 2013. "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change." *Review of Environmental Economics and Policy* 7 (2): 181–98.

Banerjee, Lopamudra. 2010. "Effects of Flood on Agricultural Productivity in Bangladesh." *Oxford Development Studies* 38 (3): 339–56.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Under-investment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh." *Econometrica* 82 (5): 1671–1748.

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30 (4): 171–98.

Gray, C. L., and V. Mueller. 2012. "Natural Disasters and Population Mobility in Bangladesh." *Proceedings of the National Academy of Sciences* 109 (16): 6000–6005.

Guiteras, Raymond, Amir Jina, and A. Mushfiq Mobarak. 2015. "Satellites, Self-Reports, and Submersion: Exposure to Floods in Bangladesh." *American Economic Review* 105 (5): 232–36.

Islam, Nazrul, and Hiroshi Uyeda. 2007. "Use of TRMM in Determining the Climatic Characteristics of Rainfall over Bangladesh." *Remote Sensing of Environment* 108 (3): 264–76.

Ji, Lei, Li Zhang, and Bruce Wylie. 2009. "Analysis of Dynamic Thresholds for the Nomalized Difference Water Index." *Photogrammetric Engineering and Remote Sensing* 75 (11): 1307–17.

Lu, Xin, David J. Wrathall, Pal Roe Sundsoy, M. Nadiruzzaman, Erik Wetter, Asif Iqbal, Taimur Qureshi, et al. 2016. "Unveiling Hidden Migration and Mobility Patterns in Climate Stressed Regions: A Longitudinal Study of Six Million Anonymous Mobile Phone Users in Bangladesh." *Global Environmental Change* 38: 1–7.

Mueller, V., C. Gray, and K. Kosec. 2014. "Heat Stress Increases Long-Term Human Migration in Rural Pakistan." *Nature Climate Change* 4: 182–85.

National Research Council of the National Academies. 2007. *NOAA's Role in Space-Based Global Precipitation Estimation and Application.* Washington, DC: National Academies Press.

Willmott, C. J., and K. Matsuura. 2001. "Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950–1999)." Center for Climate Research.

Xiao, Xiangming, Stephen Boles, Steve Frolking, Changsheng Li, Jagadeesh Y. Babu, William Salas, and Berrien Moore. 2006. "Mapping Paddy Rice Agriculture in South and Southeast Asia using Multi-Temporal MODIS Images." *Remote Sensing of Environment* 100 (1): 95–113.

Xu, Hanqiu. 2006. "Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery." *International Journal of Remote Sensing* 27 (14): 3025–33.