Validating Migration Responses to Flooding Using Satellite and Vital Registration Data

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

Quantifying flood-driven migration is crucial for local 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 harboring messages of mass exodus (Meyers, 2002; Gemmene, 2011) 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 greater role for drought-related crop failure (than flooding) on permanent migration. However, the study is limited to selected sites, while environmental exposure and migration will vary 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 national call data records. They find that population flows are largely unchanged by this event. But, lacking knowledge of who possesses the SIM card, 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 sub-district (*upazila*) of origin, we use remote sensing techniques as in Guiteras, Jina and Mobarak (2015). Use of satellite data to measure environmental exposure across space and time has gained traction (Donaldson and Storeygard, 2016). Typical proxies of flood exposure are rainfall extremes, measured by converting raw precipitation data into anomaly or percentile variables (Mueller, Gray and Kosec, 2014; Gray and Wise, 2016). 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 2002-2010 Sample Vital Registration System (SVRS), an annual survey of over 200,000 households conducted by the Bangladesh Bureau of Statistics. Samples

* Chen: The Ohio State University, 324 Ag. Admin., 2120 Fyffe Road, Columbus, OH 42310, chen.1276@osu.edu. Mueller: International Food Policy Research Institute, 2033 K Street, NW, Washington, DC 20006, v.mueller@cgiar.org. Jia: The Ohio State University, 275 Mendenhall Lab, 125 South Oval Mall, Columbus OH, 43210, jia.102@osu.edu. Tseng: Center for Space and Remote Sensing Research, National Central University, No.300, Jhongda Rd., Jhongli City, Taoyuan, 32001, Taiwan, khtseng@csrsr.ncu.edu.tw. 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.

are nationally representative, in order to provide inter-censal 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 understates 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. Sub-district 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.

For satellite data, differences in spectral bands and spatial and temporal resolution can also lead to different results. In the case of TRMM, only moderate to high rainfall rates can be detected, due to sensitivity limitations (National Academies Press, 2007). And, 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 three flood proxies using weather station data and two commonly-used gridded datasets.

Epanechnikov kernel densities of the sub-district correlations of monsoon precipitation levels across data products show positive correlations of the TRMM measure with both the weather station and Delaware data for the majority of the distributions (see Appendix). The average correlations between the TRMM, weather station and Delaware products are 0.74 and 0.66, respectively, with a more modest correlation of 0.54 between weather station and Delaware precipitation variables. We examine how the discrepancies in data products might translate into different predictions for flooding displacement using the model described below.

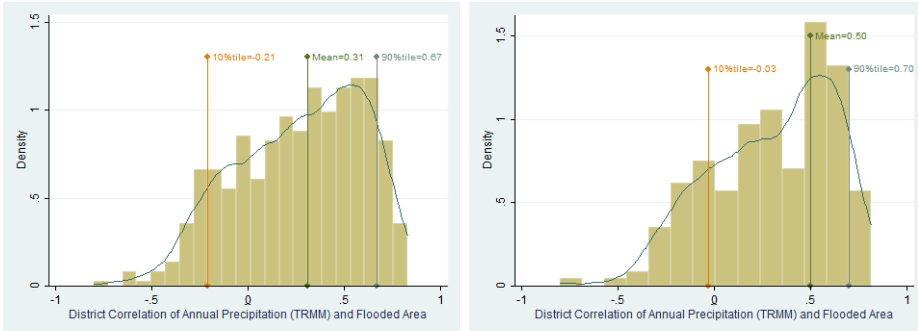
Flooding Extent Measure. Data are drawn from NASA Moderate Resolution Imaging Spectroradiometer (MODIS) satellites at 500m resolution. We construct the Modified Normalized Difference Water Index (MNDWI) (Xu, 2006), which differentiates water and non-water features based on surface reflectance.[1] A pixel is defined as water if MNDWI>0.1.[2] Upazila-level measures are based on the maximum percentage of water pixels over all 8-day composites in the period. Finally, to differentiate water bodies from flooding, we look at the difference in water coverage between the monsoon (July-Dec) and dry (Jan-Mar) 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. However, the correlation is much stronger in areas where flooding is due primarily to river over-topping, as in Panel (b). This figure restricts the sample to areas with high river density, defined as the top 40% of sub-districts with respect to river length as a proportion of total area.[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

[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]Derived from Global Lakes and Wetland Database.



(a) All Areas

(b) High River Density, Top 40%

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

Appendix). In the absence of remote sensing measures, this suggests rainfall proxies may suffice in areas experiencing primarily river flooding.

We also examine whether monsoon precipitation may be a better proxy for flooding, given that this season accounts for well over half of 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 Delaware product (see Appendix). Longer-term precipitation measures better reflect overall water balance, and the limitations of satellite products in detecting rainfall across seasons may make annual precipitation 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)

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

We adopt the convention of looking at quintiles, 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 Appendix), climate (lagged quintile categorical variables for growing degree days over the growing season, and 30-year running averages for degree days and annual precipitation); as well as time invariant regional characteristics (whether the household is located in the coastal zone or in the northwest region). We also control for competing time-specific influences on migration by including a time fixed effect γ_t. Standard errors are clustered at the *upazila* level to allow for sub-district correlation in unobserved factors influencing migration.

III. Results

Table 1 displays the point estimates from (1) when including the flood proxies (precipitation quintiles derived from weather stations and gridded data products) and the preferred remote sensing measure of flooding extent. Looking at the full sample, we find significant negative associations between the fourth and fifth quintile precipitation variables and migration across all data products. 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 declines by 0.6 to 1.8 percentage points. The remote sensing measure, however, reveals effects of localized flooding as well, with significant negative effects observed at lower quintiles, albeit smaller in magnitude.

Table 1—Migration-Flood Relationships

	Station	Station	Del.	Del.	TRMM	TRMM	Flood	Flood
Quintile 2	0.002	0.009	0.001	0.002	-0.002	0.000	-0.004	0.003
	(1.21)	(3.37)	(0.44)	(0.85)	(1.14)	(0.15)	(2.07)	(0.96)
Quintile 3	0.001	-0.002	-0.002	0.004	-0.001	-0.001	-0.003	-0.000
	(0.30)	(0.68)	(1.18)	(1.04)	(0.57)	(0.41)	(1.77)	(0.02)
Quintile 4	-0.004	-0.005	-0.007	-0.004	-0.009	-0.009	-0.006	-0.001
	(1.84)	(1.59)	(2.97)	(1.11)	(3.76)	(2.21)	(2.57)	(0.36)
Quintile 5	-0.006	-0.003	-0.016	-0.011	-0.018	-0.014	-0.008	0.004
	(2.39)	(0.78)	(5.22)	(2.36)	(6.40)	(2.86)	(3.28)	(0.88)
Sample	Full	HRD	Full	HRD	Full	HRD	Full	HRD

Note: 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, and year fixed effects. Standard errors clustered at upazila level. t-statistics presented in parentheses.

Given stronger correlation between rainfall proxies and satellite-based measures in areas with high river density, we also utilize this sample restriction in our regressions. The significant negative effects in the fifth quintile are still evident for two of the three rainfall proxies but, surprisingly, vanish for the satellite-based measure. This suggests that correlations between flooding and migration are quite fragile and vary substantially across areas. Moreover, our estimates suggest that, while rainfall and flooding exhibit reasonably high correlation, they may represent very different challenges for vulnerable households.

IV. Discussion

Using nationally representative data on migration in Bangladesh, we find a modest negative effect of flooding on the probability that a household sent out at least one migrant in the previous year. Individuals may be more likely to be trapped (Black et al., 2011) than internally displaced by floods. An alternative explanation is 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 from gridded datasets are qualitatively similar to those using remote sensing measures when focusing on the top quintile. However, the coarseness of these proxies may be masking other sources of variation. This becomes apparent when examining the point estimates on the remaining percentiles of the flooding measures across specifications. Only remote sensing indicators capture the effects of localized floods (represented by the lower quintiles), which are driven by proximity to rivers, topography, and other conditions unrelated to local rainfall.

Specifications using remote sensing indicators alone convey a non-monotonic relationship between migration and flooding. Even modest flooding (2nd and 3rd quintiles, 3-17% of the sub-district) significantly deters migration, but there is a markedly larger effect in the 4th and particularly 5th quintiles. Broader exposure to flooding throughout a sub-district can reduce opportunities to access credit and/or utilize risk pooling mechanisms to finance migration. However, our findings are not robust across regions. In areas with high river density, we continue to observed a negative relationship between rainfall and migration, but we do not find any significant relationship between migration and our satellite-based flooding measure. Thus, despite high correlation between rainfall and flooding, our results suggest that households experience these two phenomena quite differently.

REFERENCES

Auffhammer, M., S. Hsiang, W. Schlenker, and A. Sobel. 2013. "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change." *Review of Evironmental Economics and Policy*, 7: 181–198.

Banerjee, L. 2010. "Effects of Flood on Agricultural Productivity in Bangladesh." *Oxford Development Studies*, 38(3): 339–356.

Black, R., S. Bennett, S. Thomas, and J. Beddington. 2011. "Climate Change: Migration as Adaptation." *Nature*, 478(7370): 447–449.

Bryan, G., S. Chowdhury, and A. Mobarak. 2014. "Under-Investment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh." *Econometrica*, 82(5): 1671–1748.

Donaldson, D., and A. Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives*, 30(4): 171–198.

Gemmene, F. 2011. "Why the Numbers Don't Add Up: A Review of Estimates and Predictions of People Displaced by Environmental Changes." *Global Environmental Change*, 21: S41–S49.

Gray, C., and E. Wise. 2016. "Country-specific effect of climate variability on human migration." *Climatic Change*, 555–568.

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

Guiteras, R., A. Jina, and M. Mobarak. 2015. "Satellites, Self-reports, and Submersion: Exposure to Floods in Bangladesh." *American Economic Review (Papers and Proceedings)*, 105(5): 232–236.

Islam, M., and H. Uyeda. 2007. "Use of TRMM in Determining the Climatic Characteristics of Rainfall over Bangladesh." *Remote Sensing of Environment*, 108(3): 264–276.

Ji, L., L. Zhang, and B. Wylie. 2009. "Analysis of dynamic thresholds for the normalized difference water index." *Photogrammetric Engineering and Remote Sensing*, 75(11): 1307–1317.

Lu, X., D. Wrathall, P. Sundsy, M. Nadiruzzaman, E. Wetter, A. Iqbal, T. Qureshi, A. Tatem, G. Canright, K. Eng-Mensen, and L. Bengtsson. 2016. "Unveiling Hidden Migration and Mobility Patterns in Climate Stress Regions: A Longitudinal Study of Six Million Anonymous Anonymous Mobile Phone Users in Bangladesh." *Global Environmental Change.*

Meyers, N. 2002. "Environmental Refugees: A Growing Phenomenon of the 21st Century." *Philosophical Transactions of the Royal Society of London B Biological Science*, 357: 609–613.

Mueller, V., C. Gray, and K. Kosec. 2014. "Heat Stress Increases Long-Term Human Migration in Rural Pakistan." *Nature Climate Change*, 4: 182–185.

National Academies Press. 2007. *NOAAs Role in Space-Based Global Precipitation Estimation and Application.* Washington, D.C.

Willmott, C. J., and K. Matsuura. 2012. "Terrestrial Precipitation: 1900-2010 Gridded Monthly Time Series."

Xiao, X., S. Boles, S. Frolking, C. Li, J.Y. Bau, and W. Salas. 2006. "Mapping paddy rice agriculture in South and Southeast Asia using multitemporal MODIS images." *Remote Sensing of Environment*, 100(1): 95–113.

Xu, H. 2006. "Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery." *International Journal of Remote Sensing*, 27(14): 3025–3033.

Appendix

Table A1—Summary Statistics

	Mean	Std.
Migration	0.05	0.22
Monsoon season precipitation, station	1,528.32	516.39
Monsoon season precipitation, U. Del.	1,673.06	797.11
Monsoon season precipitation, TRMM	1,605.05	480.70
Flood	0.17	0.18
Degree days	5,874.80	216.58
30-year average degree days	5,713.01	151.92
30-year average precipitation	2,037.44	151.92
Head is male	0.89	0.31
Head's age	45.22	13.76
Head is literate	0.50	0.50
Head is muslim	0.88	0.32
Head is hindu	0.10	0.30
No. of hh members	4.85	2.20
Male hh members 0-5 years old	0.31	0.56
Male hh members 6-16 years old	0.65	0.84
Male hh members 17-54 years old	1.25	0.87
Male hh members greater than 54 years old	0.23	0.43
Female hh members 0-5 years old	0.30	0.56
Female hh members 6-16 years old	0.61	0.83
Female hh members 17-54 years old	1.29	0.76
Female hh members greater than 54 years old	0.20	0.41
Primary water source comes from tap	0.08	0.28
Primary water source comes from well	0.90	0.30
Secondary water source comes from tap	0.09	0.28
Secondary water source comes from well	0.48	0.50
Has own water source	0.53	0.50
Has kerosene as source of light	0.46	0.50
Has electricity as source of light	0.53	0.50
Has kerosene as source of fuel	0.004	0.07
Has electricity as source of fuel	0.006	0.08
Has gas as source of fuel	0.09	0.29
Has modern or sanitary latrine	0.58	0.49

Note: N=1,931,954. U. Del.=University of Delaware.

Table A2—Out-Migration Rates

	Rural destination	Urban destination	Abroad
Rural origin	33.99	11.61	9.39
Urban origin	13.12	28.85	3.03

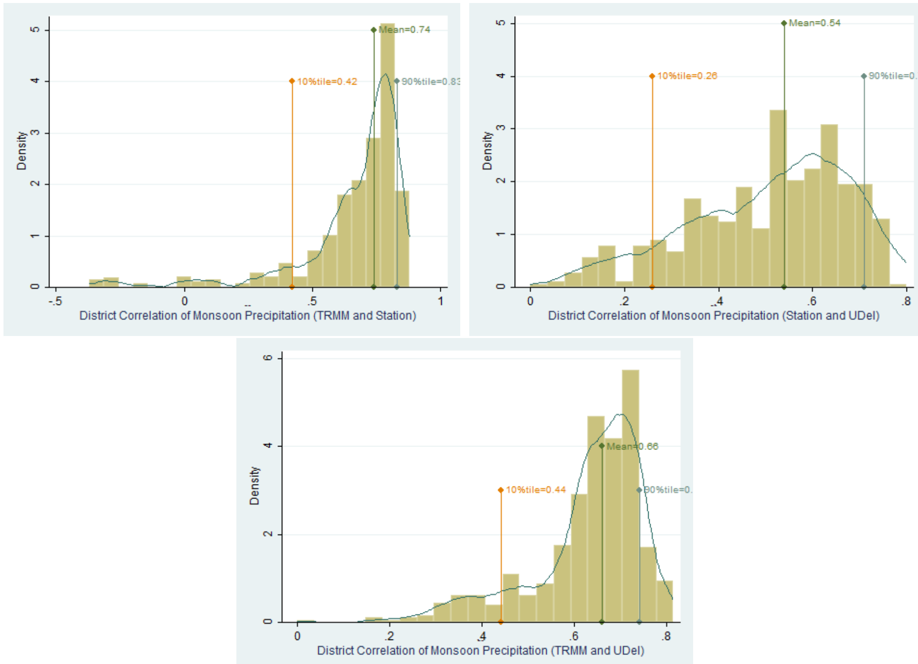


Figure A1. Epanechnikov Kernel Densities of Correlations between Data Products

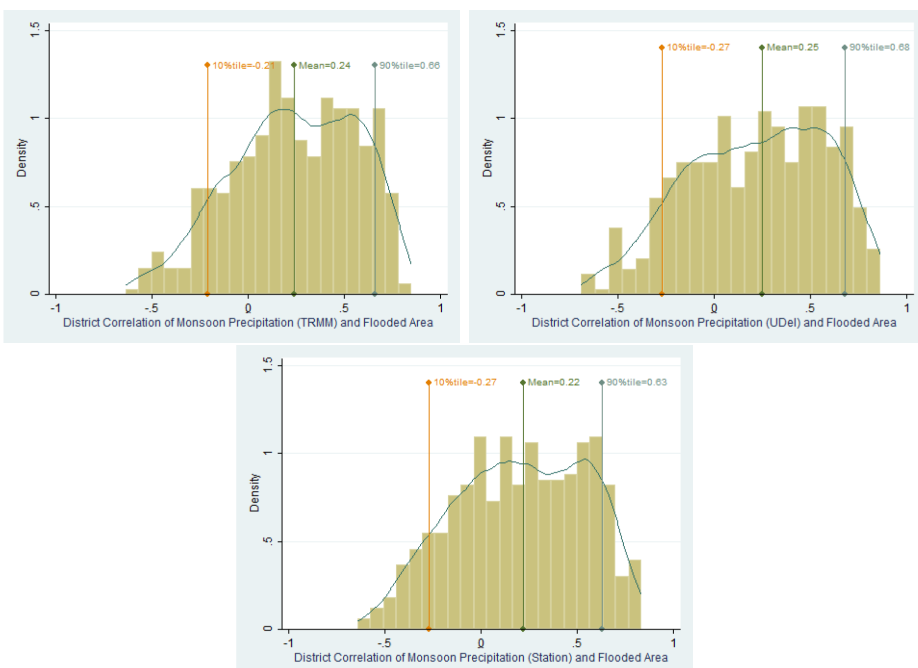
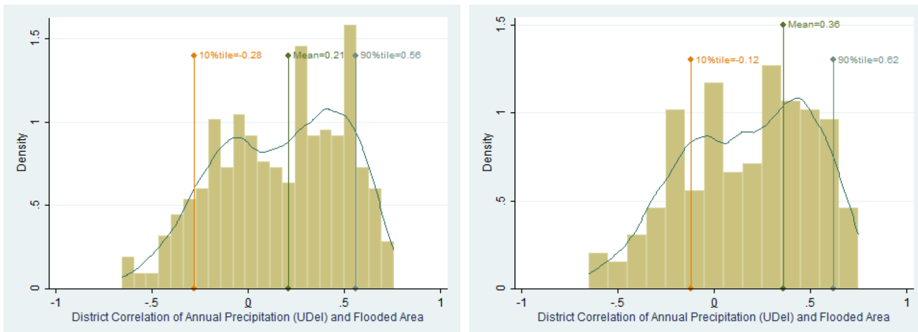
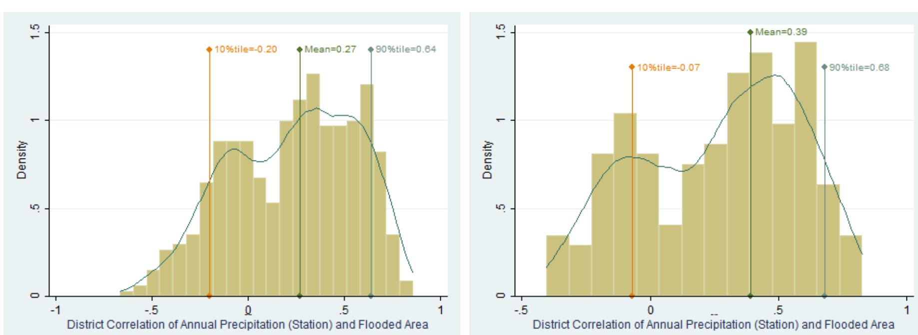


Figure A2. Epanechnikov Kernel Densities of Correlations between Monsoon Precipitation and Flood Measures



(a) All Areas (b) High River Density (top 40%)

Figure A3. Epanechnikov Kernel Densities of Correlations between UDel Precipitation and Flood Measures



(a) All Areas (b) High River Density (top 40%)

Figure A4. Epanechnikov Kernel Densities of Correlations between Station Precipitation and Flood Measures