# *Maryland Population Research Center*

## Satellites, Self-reports, and Submersion: Exposure to Floods in Bangladesh

Authors:

**Raymond P. Guiteras**
University of Maryland

**Amir Jina**
University of Chicago

**A. Mushfiq Mobarak**
Yale University

# Satellites, Self-reports, and Submersion: Exposure to Floods in Bangladesh

Raymond Guiteras, Amir Jina, and A. Mushfiq Mobarak

January 13, 2015

Abstract:

Exposure to abnormal floods is believed to have negative short- and long-term consequences for welfare and health in poor countries, and such impacts are likely to grow worse with continued anthropogenic climate change. However, two common proxies for flood exposure, self-reported exposure and rainfall, are problematic. This paper describes a method for constructing objective measures of flood exposure using satellite data. Using the case of Bangladesh in the period 2002-2011, we show that (a) self-reported exposure has an important, non-random bias in that it responds much more strongly to actual exposure in areas where floods are relatively rare, and (b) rainfall is at best weakly correlated with floods.

# Satellites, Self-reports, and Submersion:

# Exposure to Floods in Bangladesh

Raymond Guiteras, Amir Jina, and A. Mushfiq Mobarak*

January 13, 2015

A burgeoning "Climate-Economy" literature (Dell, Jones and Olken, 2014) attempts to understand and project the economic impacts of anthropogenic climate change. This literature has largely focused on uncovering the effects of changes in temperature and precipitation on economic activity. Important papers have documented both short-run effects—for example, on agriculture (e.g., Schlenker and Roberts, 2009), health (e.g., Deschênes and Greenstone, 2011), and labor (e.g., Graff Zivin and Neidell, 2013)—and long-run effects—for example, on economic growth (Dell, Jones and Olken, 2012), and education (e.g., Maccini and Yang, 2009). This has been made possible by the availability of temperature and precipitation data with reasonable spatial and temporal resolution.

Global climate change is likely to cause rising sea levels, more powerful cyclones and greater coastal storm surges, and increased frequency and severity of flooding (IPCC, 2014). The economics literature has made less progress in modeling the socio-economic effects of these other phenomena expected to be associated with climate change, and which may have more intense, deleterious effects in the short run. A handful of recent papers have used physical science models to create such data (e.g., Anttila-Hughes and Hsiang, 2012; Hsiang and Jina, 2014). However, no similar effort has been made for flooding, a class of disaster that affects more people than any other (EM-DAT, 2012). We describe the progress we have made in creating a time series of flood exposure derived from a new analysis of satellite data. We focus on the lower Ganges Delta, and the nation of Bangladesh in particular, one of the countries historically most affected by floods, and predicted to experience increasing flood severity due to climate change (Mirza, 2010).
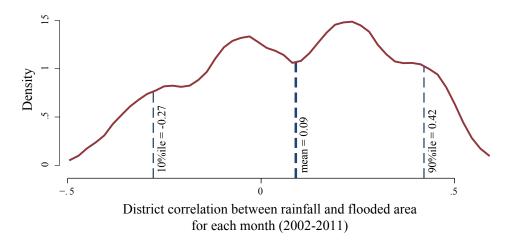
Figure 1: Distribution of correlation coefficients, $\rho(flood_i, rain_i)$ for all districts, $i$, in Bangladesh for each month in our dataset.

This paper makes two key contributions. First, we present new, objective long-run time series measures of floods that will allow us to study human behavioral responses to changes in the distribution of disaster events. In particular, the unexpected nature of the change may itself have productivity consequences separate from the occurrence of a disaster event. The socio-economic consequences of the disaster may therefore depend on the novelty factor, i.e. how much experience people already had in dealing with similar events in the past. This is a dimension of adaptation that is possible to study only with rich data variation in the background frequency of exposure at the locations where those events occur. Our dataset does exactly this, giving accurate measures of both the long-term average and the short term variation in exposure required to study adaptation.[1]

Second, we show that rainfall and self-reported exposure are weak proxies for true flood exposure. The most damaging floods are caused by rivers bursting their banks, generally caused by rainfall occurring over an entire river basin[2] and not just directly above where flooding occurs. We demonstrate that flooding in districts in Bangladesh is not directly correlated with rainfall at those specific locations. Floods are a result of complex hydrology and this lack of correlation will likely hold in many locations around the world. We also show that self-reported exposure is a weak proxy for objective exposure, and that measurement error is likely to be correlated with important determinants of socio-economic outcomes, in particular mean exposure to floods.

---

[1] In an analogous approach, Hsiang and Jina (2014) use a physical model to examine responses to tropical cyclones in groups of countries with a range of average exposures, from low to high. They find significantly larger marginal effects in the "naive", or less frequently exposed, countries.

[2] See online appendix fig. 1 for a map of the entire scale of the river system of which Bangladesh is part.

# 1 Measuring floods and their impacts

In the climate impacts literature, we wish to estimate the following equation:

$$y = f(E) + \varepsilon \tag{1}$$

where $y$ represents an outcome of interest and $E$ represents environmental exposure. Unbiased, precise estimation of $f$ requires accurate data on both outcomes and exposures. For example, we could model agricultural yields, $y$, as a function of temperature and precipitation, $f(T, P)$. In contrast, much of the work on the impact of extreme events has relied on self-reported survey data or nationally reported disaster statistics for both the left-hand side variable $y$ (e.g. damage, losses) and the right-hand side variable $E$ (e.g. subject states her household was affected by flooding). This is equivalent to modeling:

$$y_1 = f(y_2) + u \tag{2}$$

with $y_1$ and $y_2$ both being outcomes of some underlying environmental exposure, leading to compounding of errors:

$$y_1 = f(E) + \varepsilon + u \tag{3}$$

In agriculture, this would be similar to estimating the effect of climate on income by regressing income on agricultural yields, which clearly would provide little information about the effect of interest. Additionaly, errors in measurement may be correlated with $\varepsilon$, which represents other determinants of outcomes. For example, poorer households might be more exposed but less able to assess damages accurately.

To estimate the impacts of floods we face a singular problem: no comprehensive database of flood exposure through time exists for Bangladesh. We derive flood extent for each union[3] in Bangladesh using remote sensing data collected by the NASA Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS is an array of satellites that scan the Earth's surface every two days, recording reflectance values over 36 bands in the visible and infra-red spectra. As clouds are opaque to visible and infra-red light, cloud cover will restrict the use of images for detecting surface properties. Due to this, data are processed into cloud-free composites of 16 days. Composite data are available for the period between 2000-2013 at 250m×250m resolution. This results in a total of 3,159×2,482 pixels for each of 253 time periods. These $1.98 \times 10^9$ pixels are used to derive flood extents for the whole period.

We follow the framework of Sakamoto, Phung and Nhan (2009) and adapt it based on extensive fieldwork and observation in Bangladesh in 2012. The intuition behind the method is to construct two measures, one of which is sensitive to surface water and the other to surface vegetation (or greenness). If the value of the index for water surpasses

---

[3]Administrative units have the following hierarchy: Division ⊃ District (Zila) ⊃ Sub-district (Upazila) ⊃ Union. Unions have an average size of approximately 10-20km$^2$.

Table 1: Comparison of self-reported flood effects and objectively measured flooded area.

| Survey question | Answered "YES" to survey question | | Answered "NO" to survey question | |
|---|---|---|---|---|
| | # HH answered "YES" | Satellite-derived flood proportion average | # HH answered "NO" | Satellite-derived flood proportion average |
| **Survey question** | | | | |
| Affected by July '04 floods | 900 | 0.118 | 892 | 0.065 |
| House damaged/lost | 534 | 0.155 | 366 | 0.063 |
| Latrine damaged/lost | 307 | 0.152 | 593 | 0.100 |
| Water source damaged/lost | 142 | 0.169 | 758 | 0.108 |
| Food stocks damaged/lost | 64 | 0.248 | 836 | 0.108 |
| Crops damaged/lost | 667 | 0.114 | 233 | 0.130 |
| Farm destroyed | 197 | 0.076 | 703 | 0.129 |
| Livestock died | 67 | 0.169 | 833 | 0.114 |
| HH members sick | 49 | 0.239 | 851 | 0.111 |
| HH members died | 3 | 0.094 | 897 | 0.118 |
| Lost employment/inc. source | 102 | 0.096 | 798 | 0.121 |

Values in the "YES" and "NO" columns represent the objectively measured flood extent in July 2004 as a proportion of total sub-district area, averaged over the number of households who reported either an effect or no effect. Each row is a separate question asking about self-reported damages from flooding in July 2004.

that for greenness then we can say that there is overlying surface water. In practice, the algorithm for classifying floods is more complex, though the intuition remains the same.

The land surface at each point in time is classified into three categories: 1) **Non-flood**: Pixels which show no evidence of standing surface water; 2) **Mixed**: Pixels which show a mixture of standing water and vegetation; and 3) **Flood**: Pixels which are unambiguously flooded over their whole extent. We then use the time dimension to distinguish between temporary flooding and permanent water.

# 2 Other Data

*Survey Data:* We use the nationally representative Child and Mother Nutrition Survey of Bangladesh 2005 (BBS/UNICEF, 2007), which focuses on children aged 0-59 months and their mothers. Data were collected throughout 2005. Importantly for the current analysis, questions were asked about the impact of floods during the Monsoon season in the previous year, regarded as a particularly bad flood year. 57% of households report being affected by the 2004 floods.

*Rainfall Data:* Rainfall data for each district in Bangladesh is derived from the Tropical Rainfall Measuring Mission satellite at daily frequency from 1998-2013. This is summed to give monthly totals.
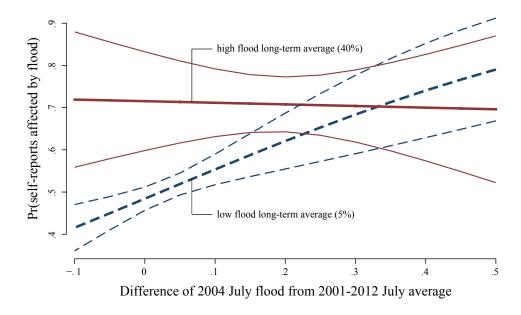
Figure 2: Interaction between average flood exposure and flood exposure in 2004.

*Merging data:* Pixel-level floods data are projected onto union boundaries, obtained from the Government of Bangladesh's Local Government Engineering Division (LGED) and averaged at the union level. Unions are matched to the CMNS via Bangladesh Bureau of Statistics (BBS) geocodes. We exclude urban locations, resulting in 1792 households.

# 3 Results

## 3.1 Rainfall versus flooding

Flooding in Bangladesh results from rainfall accumulating over the entire river basin, and it may be less influenced by its own rainfall than by a complex and broad set of hydrological conditions in an area approximately ten times its size. We estimate the correlation between monthly rainfall and monthly flood extent measured at the Zila (district) level over the period 2002-2011 (figure 1). The overall correlation is positive but modest (0.09), and far from uniform throughout time. In fact, the correlation is negative for approximately 40% of the months between 2002 and 2011, and the 10th to 90th percentiles span -0.27 to 0.42. We conclude that rainfall is a poor proxy for floods. This will be especially true when the exact timing of a flood is important—for example a flood that occurs during a sensitive growing period for crops or during a vulnerable stage of human development may have disproportionately large impacts.

## 3.2 Self-reports versus satellites

Self-reported data are not only subject to recall bias, but also to other forms of cognitive bias like reference dependence. A flood which has a larger effect might have greater pertinence and so be more likely to be reported. Of particular

concern for analysis of flooding is that people may adapt to the average exposure conditions, viewing them as a reference point to judge deviations from that average. This would imply that a household frequently exposed to larger floods and one not frequently exposed may view a flood of the same magnitude in different ways. We must also be concerned that this difference will result from adaptation. This could be positive—a household invests in protecting vulnerable productive assets and property when flood impacts are understood—or negative—a household ceases to invest in vulnerable assets, accepting some level of productivity loss in the process.

In table 1 we divide households into those that reported they were affected by the 2004 floods, and those who reported no effect. We present the number of households answering "Yes" and "No" in each case. We then estimate the average flood exposure across all households in each category. We see that households reporting "Affected by July '04 floods" did experience a higher objective flood exposure (11.8% inundation compared to 6.5% for those answering "No"). However, if we look at the exposure of those who lost their farms, we see the opposite pattern. Households in riskier areas could have changed farming practices, and so those losing farms may have been affected by a smaller, but unexpected flood.

We then examine the response of households at low exposure levels to the deviation of the 2004 flood from their local average, and compare this to the response of households at higher average levels. We run a logit regression to determine the probability of reporting being "affected" as a function of average exposure, the deviation from average in July 2004, and their interaction.[4] Fig. 2 shows that low exposure households are more likely to report being affected if they experience a larger flood (dashed line). In contrast, high exposure households reporting is comparatively inelastic to flood size in 2004. Households in each category appear to perceive exposure relative to their average environment. This renders self-reports of little value, and points to the need for objective measures of exposure.

## 4    Conclusions and Future Directions

People appear to be adapted to their average environment, and to experience seemingly similar shocks differently. This limits the usefulness of self-reported data in understanding the impacts of an extreme event like flooding. Moreover, without knowledge of average levels of exposure, we are unable to understand what this adaptation might entail. This is crucial when trying to understand the impacts of climate change, as people will not only experience new exposures, but also experience them differently. Future work will aim to identify these differential responses, and to characterize adaptive investments and behaviors.

---

[4]Results shown in online appendix table 1.

# References

**Anttila-Hughes, Jesse K., and Solomon M. Hsiang.** 2012. "Destruction, Disinvestment, and Death: Economic and Human Losses Following Environmental Disaster." *isites.harvard.edu*.

**BBS/UNICEF.** 2007. *Child and Mother Nutrition Survey 2005.* Bangladesh Bureau of Statistics and UNICEF, Dhaka.

**Dell, Melissa, Benjamin F Jones, and Benjamin a Olken.** 2012. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century." *American Economic Journal: Macroeconomics*, 4(3): 66–95.

**Dell, Melissa, BF Jones, and BA Olken.** 2014. "What do we learn from the weather? The new climate-economy literature." *Journal of Economic Literature*, 52(3): 740–798.

**Deschênes, Olivier, and Michael Greenstone.** 2011. "Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US." *American Economic Journal: Applied Economics*, 3(October): 152–185.

**EM-DAT.** 2012. *The OFDA/CRED International Disaster Database.* Universit Catholique de Louvain, Brussels (Belgium).

**Graff Zivin, Joshua, and Matthew Neidell.** 2013. "Environment, Health, and Human Capital." *Journal of Economic Literature*, 51(2012): 689–730.

**Hsiang, Solomon M., and Amir S. Jina.** 2014. "The Causal Effect of Environmental Catastrophe on Long-run Economic Growth: Evidence from 6,700 cyclones." *NBER Working paper no. 20352*.

**IPCC.** 2014. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects.* Cambridge University Press.

**Maccini, Sharon, and Dean Yang.** 2009. "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall." *American Economic Review*, 99(3): 1006–1026.

**Mirza, M. Monirul Qader.** 2010. "Climate change, flooding in South Asia and implications." *Regional Environmental Change*, 11(S1): 95–107.

**Sakamoto, Toshihiro, VC Phung, and VN Nhan.** 2009. "Agro-ecological interpretation of rice cropping systems in flood-prone areas using MODIS imagery." *Photogrammatic Engineering and Remote Sensing*, 75(4): 413–424.

**Schlenker, Wolfram, and Michael J Roberts.** 2009. "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change." *Proceedings of the National Academy of Sciences of the United States of America*, 106(37): 15594–8.
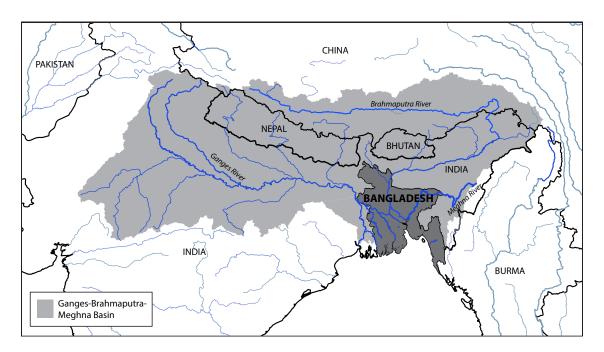
# A   Appendix tables and figures



Figure 3: Ganges-Brahmaputra-Meghna basin. Bangladesh drains water from an area almost 10 times larger than its own land area. This implies that rainfall over the entire region matters for flooding, and not rainfall directly over Bangladesh.

Table 2: Does objective exposure predict survey response?

| | (1) | (2) | (3) |
|---|---|---|---|
| **Dependent variable** | Self-reported flood effect | | |
| **Model** | July 2004 | July Average | interaction |
| $\text{Flood}_{July04}$ | 1.884*** | | |
| | (0.314) | | |
| $\text{Flood}_{July01-11}$ | | 2.389*** | 2.813*** |
| | | (0.452) | (0.774) |
| $\Delta_{July04}$ | | | 3.204*** |
| | | | (0.865) |
| $\text{Flood}_{July04} \times \Delta_{July04}$ | | | -8.472** |
| | | | (3.620) |
| Constant | -0.156*** | -0.116*** | -0.204*** |
| | (0.0540) | (0.0523) | (0.0576) |
| Observations | 1792 | 1792 | 1792 |
| Pseudo $R^2$ | 0.0168 | 0.0133 | 0.0192 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$