Flooded Cities
Online appendix

Adriana Kocornik-Mina, Thomas K.J. McDermott,
Guy Michaels, Ferdinand Rauch

March 14, 2019

This is the web appendix for the paper Kocornik-Mina et al (2019).

1 Online Appendix A: Theory

To frame our empirical investigation, we outline a simple framework that allows us to consider how individuals may respond to a large flood. We consider a discrete-time model, where a person has to choose between two locations, one to which we refer as "Risky" (indexed by R) and another which we will for simplicity consider "Safe" (indexed by S).

The person in question resides initially in the risky location, and considers whether to relocate to the safe location. The period utility of the person from the risky location is

$$U_R = V_R - P_F(D_F - T_F) + u(c), \tag{1}$$

where V_R is the utility from residing in the risky location; P_F is the assessed probability of a flood, which we discuss below; D_F and T_F are the damage from a flood and the transfers received in the aftermath of a flood, expressed in utility terms; and $u(c)$ is the utility from non-residential consumption. We drop time subscripts to increase simplicity. The period utility from the safe location is

$$U_S = V_S + u(c), \tag{2}$$

where V_S is the utility from residing in the safe location. But in order to move the person has to pay relocation costs M, which capture the cost of moving. We also assume that once a flood has hit the person has to pay the cost M regardless of whether they move or stay, since the flood implies paying costs of renovating over and above those captured by D_F. The point of this simplifying assumption is that when a flood hits, the cost of moving (compared to staying) is lower than in the absence of the flood. We assume that non-residential consumption is a numeraire good, whose price is normalized to one. The budget constraint is therefore:

$$I = p(L) + c + M \times 1_{move},$$

where I is income (which we assume to be constant);[1] $p(L)$ is the rental price of residing in location $L \in \{R, S\}$, which is paid to absentee landlords; and 1_{move} is an indicator for moving. The choice over relocation represents an infinite horizon problem, with discount rate θ. Given the simple structure of the model, and holding prices fixed, our individual relocates from the risky to the safe location if $V_S - V_R + P_F(D_F - T_F)$ is sufficiently large. An important factor in this model is how the person assesses the probability of a flood. Following Turner (2012) we model flooding through a Beta-Bournoulli Bayesian learning model.[2] We assume that the risk of a flood (by which we mean a large flood) in a given year is x. Our resident's prior is that x is

[1] We could assume that income depends on location, but this would not substantively change the model.

[2] Gallagher (2014) provides evidence of Bayesian learning in the context of floods. As we explain below this is a simplification, since this probability can rise with climate change, or decline with public investment to address climate change.

distributed according to a Beta distribution: $x \sim \beta(\alpha, \beta)$. The probability distribution function is:

$$f(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, x \in [0,1], \alpha > 0, \beta > 0, \tag{3}$$

where the normalization constant is the Beta function:

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx. \tag{4}$$

The prior probability of a flood is therefore

$$P_F = E[x] = \frac{\alpha}{\alpha+\beta}. \tag{5}$$

After observing t years, during which a flood has occurred S_t times, the updated posterior is:

$$E[x|t, S_t] = \frac{\alpha + S_t}{\alpha+\beta+t}. \tag{6}$$

In other words, for an individual who has information on flood events in the past t years, the expected probability of a flood next year increases by $1/(t+\alpha+\beta)$ if a flood took place in year t compared to the case where it did not. As t approaches infinity there is no updating. The model captures the intuition of Bayesian learning: as t approaches infinity there is no more updating, since the degree of risk is known.

If as a result of the flood people update the risk of flooding, some of them may want to relocate from the risky area to the safe one. Similarly, if there is no flood and people update, there may be movement in towards the risky areas. But without loss of generality we will focus on the case where a flood does occur.

If people do update and some want to move from the risky area to the safe one, then the population or the prices in the safe area (or both) will increase relative to the risky area. If housing supply is inelastic, then the price of housing in the safe areas will increase relative to the risky areas, but there will be no change in the population ratio between the two locations and we might not detect any change in night time light activity. But if housing supply is not completely fixed, then we expect both the price of housing and the population of the safe area to increase relative to the risky area so that updating will result in changes in economic activity as reflected by night time lights.

This simple model guides our empirical investigation in the following ways. First, we investigate the link between risk and low elevation locations. Anticipating and quantifying flood risk in the real world is a complicated endeavor, but we ask specifically how much more susceptible to large scale flooding are low elevation locations, compared to high elevation ones. This informs us about the approximate magnitude of P_F.

Second, we ask whether people generally reside in riskier low elevation urban areas. In the model, the benefits to living in risky areas (if $V_R > V_S$), or moving costs, M, might make it prohibitively expensive to relocate. One set of advantages for risky areas could be that living near coasts or rivers makes seaborne activities, such as trade and fishing, less costly. At the same time, living in flood prone areas may be the legacy of historical lock-in (Bleakly and Lin 2012; Michaels and Rauch 2013).

Third, floods may cause people to leave the riskier areas because of either Bayesian updating, or because floods reduce the cost of moving to safer areas (relative to staying in the riskier ones). Our paper examines the extent to which large floods move economic activity away from risky areas towards safer ones.

Fourth, because updating decreases in t, we expect that there will be more updating in newly populated urban areas. In the empirical analysis we examine whether there is more relocation from riskier to safer areas in the aftermath of a flood in urban areas that concentrated no (measurable) economic activity until recently.

Fifth, we examine whether the presence of higher risk of flooding due to climatic factors shifts people towards safer areas. In our model, an increase in P_F holding all else constant, shifts people away from risky low elevation areas.

In addition to the issues raised by our model, an important question is whether rising sea levels and a changing climate will affect the aggregate global concentration of economic activity in flood prone areas. Of course, rising sea levels may increase the risk of flooding both in low elevation areas and in other areas that are currently safer. But it seems plausible to assume that at least in the near future, it is in the low elevation areas that rising sea levels will have a greater effect. In our analysis we will shed some light on the aggregate concentration of urban economic activity in low elevation areas over a longer period of time.

Our analysis also touches upon a number of normative considerations. As Kydland and Prescott (1977) note, flood protection may exacerbate the moral hazard problem of living on the flood plains. By spending public money to reduce the risk borne by those living in flood prone areas, such flood protection involves a cost. At the same time, as our paper shows, people may be reluctant to relocate away from risky areas. As sea levels rise and the world becomes richer, the tradeoffs between flood protection and the relocation of economic activity to safer areas are likely to become an important issue for public debate (see Strauss, Kulp and Levermann, 2015).

Another normative issue is how much ex-post transfers should victims receive, and in what form. In the model, a larger value of T_F makes movement away from risky areas less likely. From the perspective of a donor, if a property is frequently flooded, the costs of repeatedly paying compensation might be high. In developing countries where institutions are weak, finding private flood insurance may be a difficult challenge, especially for the poor. Ex-post disaster relief, including from large scale floods, is therefore a task that governments and non-government organizations around the world engage in from time to time. The main policy issue that we raise is whether it should be possible, in certain circumstances, to concentrate public reconstruction efforts towards safer areas, in order to avoid the high risk of recurrent disasters.

2 Online Appendix B: What do variations in night lights capture?

Our aim in this paper is to examine how prevalent it is for economic activity to concentrate in flood-prone areas, and whether cities adapt to major floods by relocating economic activity to safer areas. In particular, we ask does economic activity within cities readjust in response to major shocks, which are potentially recurrent, and which disproportionately threaten specific neighborhoods? This research question requires data on economic activity (or a reasonable proxy thereof), with temporal variation, high spatial resolution, and global coverage. In our analysis we observe variations in night light intensity on a 1km grid, in more than 1800 cities in 40 countries around the world in response to large-scale flooding (or extreme precipitation) events. We undertake the usual data cleaning exercises as described in the data section.

As Donaldson and Storeygard (2016) point out, the correlation between lights and economic activity in the cross section has long been noted (see for example Croft 1973; Doll, Muller, and Morley 2006), while Henderson et al. (2012) were perhaps the first to formally test the relationship between changes in lights and economic growth, using GDP data at the national level. Since then numerous studies have used the lights data as a proxy for economic activity or prosperity at a local (sub-national) level. An example closely related to our paper, in terms of the challenge of finding income data at the city level or finer spatial resolution is Storeygard (2016) who uses lights to proxy for economic activity for a sample of cities in Africa. Lights data offer the opportunity to study variation in economic activity where traditional economic data (especially on income and especially in a panel) are generally not available - as is the case at the city level in Africa.

Ghosh et al. (2013) provide a relevant summary of the various ways in which night time light data have been used to measure human wellbeing at the subnational level, while Donaldson and Storeygard (2016) highlight a number of novel uses of night lights data as a proxy for economic activity within small geographic units. They note that "[lights data] can plausibly be used as a proxy for economic activity under the assumption that lighting is a normal good" (p.183).

In spite of the increasing prevalence of studies exploiting the night lights data as a proxy for local economic activity, an important question is how lights respond to changes in economic activity over various spatial and temporal scales.

Using annual data for a panel of countries from 1992 to 2008, Henderson, Storeygard and Weil (2012) find evidence of a linear relationship between lights and GDP, with an estimated lights-GDP elasticity of around 0.3. The estimated relationship is of similar magnitude for a restricted sample of low and middle-income countries. At national level, some light emission growth variables correlate stronger with growth of GDP, non-agricultural GDP and manufacturing value-added (Addison and Stewart 2015).

More relevant to us are studies that examine the relationship between lights and economic activity at subnational scales. An early example is Sutton, Elvidge and Ghosh (2007) who find that "night lights track economic output" at the state or province level, for four countries (China, India, Turkey and the USA).

Taking a global perspective, Chen and Nordhaus (2011) compare night lights to economic output measured on a 1-degree grid (approx. 100km by 100km). They find a strong positive relationship between luminosity and output at the grid cell level globally. They note that the relationship is weak at the low end of the output/luminosity spectrum, specifically for log output density of less than -9, which is equivalent to about $100,000 per sq km. They conclude that estimates of income are only substantially improved for countries or regions within countries with relatively weak or non-existent traditional economic statistics.

Hodler and Raschky (2014) also estimate the relationship between lights and GDP at the regional level using a broader panel dataset of regional GDP (assembled by Gennaioli et al. 2013), which includes data for 1,503 regions in 82 countries. Similarly to Henderson et al. (2012) they find a lights-GDP elasticity of around 0.3, with a slightly larger estimate (0.386) for the short-run relationship than for the long-run relationship (0.227). Their evidence also points to a roughly linear relationship between lights and GDP at the regional level. Hodler and Raschky (2014) conclude "the relationship between night-time light and GDP is linear and thereby similar across regions with different nighttime light intensity and income levels" (p. 1030).

As mentioned above, a more closely related study, in terms of the data challenges that we face, is Storeygard (2016) who uses lights data to proxy for city-level income for a sample of cities in Africa, where income data are unavailable. As a verification exercise, Storeygard tests the relationship between lights and city (or prefecture) level GDP using Chinese data, finding that the elasticity of GDP with respect to light is significant and positive for a long difference specification (from 1990/92 - 2005). The point estimate (of around 0.25) is very similar (using either the city or prefecture data) to that found for the global sample at the country level.

As part of our own robustness tests, we estimated a lights-GDP elasticity of 0.2 for an annual panel of Indian districts over our study period of 2003-2008, which is again quite similar to the findings in Henderson et al. (2012), Hodler and Raschky (2014) and Storeygard (2016). The consistency of this finding is encouraging.

Mellander et al. (2015) examine the strength of the relationship between night time lights and economic activity using fine-grained official socio-economic data on individuals and establishments in Sweden (on a 250m grid for urban areas and 1000m for rural areas). They find that night time light has a relatively weak relationship with economic activity as measured by people's wages (consumption) or wages by establishment (production), but a relatively strong relation with population density, at this fine spatial scale. Their findings clearly indicate the limitations of the lights data for analyses in high income countries, where top-coding in the lights data becomes an important constraint. As most of our sample is in low to middle income countries, the share of top coded cells in our dataset is small (see discussion in the data section).

The relationship between lights and income or wealth has also been tested at the micro level for developing countries, for example using data from the Demographic and Health Survey (DHS), e.g. Weidmann and Schutte 2017 (see also Michalopoulos and Papaioannou 2013). Weidmann and Schutte (2017) use the DHS data to compare lights to a wealth index constructed at the household level (based on assets) for a sample of 34,047 clusters (typically a village or neighborhood within a city) from 56 surveys in 39 countries for the years 2003 and 2012. They find a correlation between lights and the wealth index that averages 0.73 across all 56 surveys included. While some of this correlation is accounted for by the differences across rural and urban locations, they also estimate the relationship for rural and urban locations separately, finding that the average correlation between lights and wealth is slightly higher for urban clusters (0.62) compared to rural clusters (0.42). They also test the relationship between predictive accuracy and the size of the buffer around the cluster location point used to measure the light intensity associated with that point; the minimum is 2km for urban clusters and 5km for rural clusters, reflecting the random artificial error introduced to the location information in the DHS data (for the sake of preserving anonymity of survey respondents). Weidmann and Schutte also experiment with buffer radii of 5km, 10km and 20km. They find that buffer size matters a lot - as they increase the minimum radius there is an increase in prediction error, which they interpret as "a clear indication that the local levels of night lights - and not the emissions across a region - seem to matter for prediction [of wealth]" (p.131).

3 Online Appendix C: The relationship between night lights, GDP, and floods in India

The evidence that we present in the paper focuses on the effect of large floods on night time lights. As we discuss in the literature survey (in Part B of this Online Appendix), a recent literature confirms that there is a strong relationship between night lights and economic activity. This makes night time lights a useful measure of GDP, since it is available at the local-annual level around the globe, even in locations where local GDP is missing or mis-measured. This strongly suggests that our results reflect the effect of floods on economic activity around the world's flooded cities.

In this section we further explore the relationship between night time lights, GDP, and floods, using data from one particular country - India. We focus on India because its cities were affected by large floods covered in our dataset in five of the six years from 2003-2008. No other country in our dataset of global cities was affected for more than three years. In addition, over the period of our study, large floods affected districts in roughly three quarters of India's states. This means that India, in addition to its size and population, exhibited a fair degree of temporal and spatial variation in the occurrence of large floods, making it an interesting case study.

Data

We obtain GDP data on Indian districts from the Indian government's planning commission (`http://planningcommission.nic.in/plans/stateplan/index.php?state=ssphdbody.htm`). We complement these with administrative boundary shapefiles from the GADM database of Global Administrative Areas (`http://www.gadm.org`).

Starting with the GDP data, we note that there were changes in districts in some states over time. To harmonize district definitions we made the following changes, which involved merging districts (listed below by state): Assam: Bongaigaon includes Chirang; Barpet includes Baksa (an imperfect match); Darrang includes Udalguri. Haryana: Gurgaon includes Mewat (an imperfect match). Jharkhand: Dumka includes Jamtara; Gumla includes Simdega; Palamu includes Latehar; Singhbhum includes Saraykela Kharsawa. West Bengal: Midnapore covers Midnapore East and Midnapore West.

Next, we merge the GDP data to the administrative boundary shapefile. This involved the following steps. We start with district-level outcomes on 24 states. One state's name was changed from Orissa to Odisha. Another state, Telangana, was later created out of Andhra Pradesh. We are missing district GDP data for some states, most of which have relatively small population (by India's standards). The states and Union Territories (UTs) for which we miss district data are, in alphabetical order: Chhattisgarh, Dadra and Nagar Haveli, Daman and Diu, Goa, Gujarat, Jammu and Kashmir, Lakshadweep, Nagaland, NCT of Delhi, Puducherry, and Tripura. Nevertheless, the states that we do have information on cover 87.5% of India's population, 71% of Indian GDP and 89.7% of its area in 2000.

After dropping the 11 states and UTs for which we have no GDP data, the remaining 25 states and UTs contain a total of 581 districts, according to the GADM administrative boundaries. The GDP data for the same 25 states and UTs contain a total of 522 districts.

An initial computer match of the GADM and GDP data, based on matching district and state names (using the reclink command in Stata), returned 403 exact matches, and a further 59 matches (using a minscore=0.95). A further iteration of this process (dropping the minscore to the default 0.6) returned a further 22 matches. These computer matches were checked manually line-by-line to ensure they were accurate and no false positives were kept.

The remaining districts were matched manually - many were identifiable from variations in the district name, while others (some 44) involved districts that had been newly created in the years after the end of our short time series of GDP data (which covers the period 2000-2008). Only two districts in the GADM data remained unmatched: Balod and Surajpur (combined population of approx. 1.5m), both in the state of Chhattisgarh.

Having matched the district-year GDP data to the GIS map, we now proceed to match in the night lights and flooding data. We define a flood indicator to take a value of 1 if any point in a district and year was flooded. We take the mean light intensity value of all the pixels in a district as the light variable. The accounting year for population and GDP does not overlap with calendar years. For this reason, and following the accounting year dates, we compute a weighted average that gives three quarter weight to the current year population, and one quarter to last year's. GDP and population data are provided at district level. We also have a district map that we use to get flood and light data via GIS. We can match the districts on the map with the districts in the census data with a few corrections of abbreviated or slightly differently spelled names.

We combine all these data to construct a panel of district-year observations. We have a balanced panel for 519 districts in the years 2000-2004 inclusive. Because of missing data this number becomes 452 in 2005, 196 in 2006 and 105 in 2007. For 2008 we only have 13 observations in the dataset.

Results

We begin to use our dataset by examining the relationship between night lights and GDP in Indian districts as reported in Appendix Table A9. In Column (1) we regress the logarithm of night lights on the logarithm of GDP at the district level.[3] The regression estimates for the first year of our study, 2003, give a precisely estimated coefficient of 0.44, with a standard error of about 0.04. This coefficient is quite stable when we add state fixed effects (Column (2)), and is broadly similar for other years. When we reverse the order of the left hand side and right hand side variables of interest, regressing the logarithm of GDP on the logarithm of light, we get a coefficient of about 1.11 (s.e. 0.12) without state fixed effects and 0.93 (s.e. 0.1) with state fixed effects. All these regressions suggest a strong correlation between GDP and night light activity at the district level in India.

We next examine the same relationship using panel data, and controlling for district and year fixed effects. The panel of lights and GDP data covers the period 2000-2008 (Columns 5-6). The panel of lights, GDP and floods covers the period 2003-2007 (Columns 7-10). The regression coefficient from this regression is around 0.19, with a standard error of around 0.05. When we now reverse the order of the variables of interest, however, we get an imprecise estimate of 0.17 with a standard error of 0.12 (Columns (3) and (4)). Taken together these results still suggest that local GDP and night lights, which are both measured with error, do tend to co-move in

[3]While in the main part of the paper we cluster the standard errors by country, here we have only one country, so we cluster the standard errors by state.

panel data, but the relationship between them is not as precise. It seems plausible that the elasticity of GDP with respect to night lights in India is around 1, but there is quite a bit of uncertainty about this magnitude.

When we then use the Indian district-level panel to regress the logarithm of night lights on the flood indicator, again controlling for district and year fixed effects and clustering by state, the coefficient estimate is around -0.02 with a standard error of around 0.014. In other words, the point estimate is quite similar to what we find in our main regressions, but it is less precise. We think that this lower precision is reasonable, since we are only using a fraction of the data that we use in the main analysis. When we repeat the regression with lagged floods we get an imprecisely estimated coefficient of -0.02. When we repeat the analysis using the logarithm of GDP instead of the logarithm of lights as the dependent variable, the coefficients are -0.005 and 0.015 for the flood indicator and the lagged flood indicator, respectively.

Taken together, these results suggest that in India night lights and GDP tend to co-move, as the literature finds. But using variation across districts in India alone, we do not have enough power to detect effects of floods on night lights or local GDP. This is perhaps unsurprising, since our regression estimates using all the world's cities fall well within the standard confidence intervals estimated using the Indian data.

References

[1] Addison, D. and B. Stewart (2015) "Nighttime Lights Revisited: The Use of Nighttime Lights Data as a Proxy for Economic Variables", Policy Research Working Paper 7496, World Bank Group, November 2015.

[2] Bleakley, H. and J. Lin (2012), "Portage and path dependence," The Quarterly Journal of Economics, 127(2), 587.

[3] Chen, Xi, and William D. Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." PNAS 108(21): 8589-94.?

[4] Croft, Thomas A. 1973. "Burning Waste Gas in Oil Fields." Nature 245: 375-76

[5] Doll, Christopher N. H., J.-P. Muller, and Jeremy G. Morley. 2006. "Mapping Regional Economic Activity from Night-time Light Satellite Imagery." Ecological Economics 57(1):75-92.

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

[7] Gallagher, J. (2014), "Learning about an infrequent event: Evidence from flood insurance take-up in the United States," American Economic Journal: Applied Economics 6(3): 206-233.

[8] Gennaioli, N. R. La Porta, F. Lopez-de-Silanes and A. Shleifer (2013), "Growth of Regions", NBER Working Paper 18937.

[9] Ghosh, T., S. J. Anderson, C. D. Elvidge and P. C. Sutton (2013), "Using nighttime satellite imagery as a proxy measure of human well-being", sustainability, 5: 4688-4719.

[10] Henderson, J.V., A. Storeygard and D.N. Weil (2012), "Measuring economic growth from outer space." American Economic Review 102.2: 994-1028.

[11] Hodler, R. and P. A. Raschky (2014), "Regional Favoritism", The Quarterly Journal of Economics : 995-1033.

[12] Kocornik-Mina Adriana, Thomas K. J. McDermott, Guy Michaels and Ferdinand Rauch, (2019), "Flooded Cities", American Economic Journal: Applied Economics, Forthcoming.

[13] Kydland, F.E., and E.C. Prescott (1977), "Rules rather than discretion: The inconsistency of optimal plans," The Journal of Political Economy, pp. 473-491.

[14] Mellander, C., J. Lobo, K. Stolarick, Z. Matheson (2015), "Night-time light data: A good proxy measure for economic activity?", PLoS ONE 10 (10): e0139779

[15] Michaels, G. and F. Rauch (2013), "Resetting the urban network: 117-2012." Bureau for Research and Economic Analysis of Development (BREAD).

[16] Michalopoulos, S. and E. Papaioannou (2013), "Pre-colonial ethnic institutions and contemporary African development", Econometrica 81 (1): 113-152.

[17] Storeygard, A. (2016), "Farther on down the Road: Transport costs, trade and urban growth in Sub-Saharan Africa", Review of Economic Studies 83; 1263-1295.

[18] Strauss, B. H., S. Kulp and A. Levermann (2015), "Carbon choices determine US cities committed to futures below sea level," PNAS, www.pnas.org/cgi/doi/10.1073/pnas.1511186112.

[19] Sutton, Elvidge and Ghosh (2007), "Estimation of gross domestic product at sub-national scales using nighttime satellite imagery," International Journal of Ecological Economics and Statistics 8 (S07): 5-21.

[20] Turner, G. (2012), "Natural disasters, risk expectations, and population growth. A Study of US counties since 1850," working paper, available from http://www.webmeets.com/files/papers/EAERE/2012/941/NatRiskLocation_12.01.31.pdf.

[21] Weidmann, N. B. and S. Schutte (2017), "Using night light emissions for the prediction of local wealth", Journal of Peace Research 54 (2): 125-140.

4 Appendix Tables

(not necessarily for publication)

Table A1: Light intensity by elevation, democracy and income levels

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$ln(Y_{ik})$	$ln(Y_{ik})$	$ln(Y_{ik})$	$ln(Y_{ik})$	$ln(Y_{ik})$	$ln(Y_{ik})$	$ln(Y_{ik})$	$ln(Y_{ik})$	$ln(Y_{ik})$
$Elev < 10m_i$	0.182	0.309	0.272	0.053	0.390	0.364	-0.028	0.004	0.024
	(0.037)	(0.060)	(0.056)	(0.012)	(0.389)	(0.410)	(0.267)	(0.270)	(0.209)
$Elev < 10m_i \times DemocracyIndicator_k$		-0.175	-0.191	0.007				0.009	-0.225
		(0.067)	(0.076)	(0.037)				(0.029)	(0.078)
$Elev < 10m_i \times ln(GDPpercapita)_k$					-0.021	-0.023	0.009	0.005	0.028
					(0.037)	(0.040)	(0.029)	(0.029)	(0.023)
Observations	3,642,083	3,610,249	3,610,249	3,610,249	3,562,613	3,562,613	3,562,613	3,543,409	3,543,409
Country FE	Yes	Yes	Yes	No	Yes	Yes	No	No	Yes
City FE	No	No	No	Yes	No	No	Yes	Yes	No
River and Coast FE	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes

Notes: The regressions reported in this Table are variations on Equation 2 and include the full global sample of all urban areas.
The dependent variable in all regressions $ln(Y_{ik})$ is the natural log of mean light intensity (measured in 2012) at each gridpoint i (located in in country k).
$Elev < 10m_i$ is a dummy variable for locations that are less than 10m above sea level.
$ln(GDPpercapita)_k$ is the natural log of GDP per capita (in 2011) in country k (data are from the Penn World Tables v8).
$DemocracyIndicator_k$ is a dummy for countries with a Polity IV score (in 2008) greater than or equal to 5.
Regressions with river and coast controls include dummies for locations within 10km of the nearest river or coast.
Robust standard errors, clustered by country, in parentheses.
Robust standard errors are clustered by country.

Table A2: Main effects of flood on light, city-year panel

	(1)	(2)	(3)	(4)	(5)	(6)
	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$
$Flood_{jt}$	-0.017			-0.019		
	(0.007)			(0.006)		
$Precip > 500mm_{lt}$		-0.039			-0.040	
		(0.011)			(0.015)	
$Precip > 1000mm_{lt}$			-0.057			-0.058
			(0.014)			(0.014)
$ln(light_{t-1})$				Yes	Yes	Yes
Observations	10,363	10,363	10,363	9,878	9,878	9,878
No. of urban areas	1,817	1,817	1,817	1,702	1,702	1,702

Notes: The results presented in this Table correspond to Equation 4 and use the sample of cities affected by at least one of the large flood events in our data. The dependent variable in all regressions $ln(Y_{jkt})$ is the natural log of mean light intensity for each city j (located in country k) in year t.
$Flood_{jt}$ is a dummy indicating whether or not city j was hit by a large flood in year t.
$Precip > 1000mm_{lt}$ ($> 500mm_{lt}$) indicates locations that experienced monthly precipitation of 1000mm (500mm) or more in year t.
All regressions include year fixed effects, city fixed effects and country-specific trends.
Columns (4) to (6) corrected by the Arellano-Bond methodology using $ln(lights_{t-2})$ as an instrument for $ln(lights_{t-1})$.
Robust standard errors, clustered by country, in parentheses.

Table A3: Recovery, gridpoint year panel, extreme precipitation (500mm) indicator

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$
$Precip > 500mm_{jt}$	-0.025					-0.027				
	(0.008)					(0.008)				
$Precip > 500mm_{jt-1}$		0.004					0.003			
		(0.012)					(0.011)			
$Precip > 500mm_{jt-2}$			0.002					0.003		
			(0.006)					(0.007)		
$Precip > 500mm_{jt-3}$				-0.013					-0.012	
				(0.012)					(0.012)	
$Precip > 500mm_{jt-4}$					-0.009					-0.009
					(0.010)					(0.010)
$ln(light_{t-1})$						Yes	Yes	Yes	Yes	Yes
Observations	1,422,018	1,422,018	1,422,018	1,422,018	1,422,018	1,392,501	1,392,501	1,392,501	1,392,501	1,392,501
No. of gridpoints	243,303	243,303	243,303	243,303	243,303	235,460	235,460	235,460	235,460	235,460

Notes: The results in this Table are variations on Equation 3 and use the sample of cities affected by at least one of the large flood events in our data.
The dependent variable in all regressions $ln(Y_{ijkt})$ is the natural log of mean light intensity at each gridpoint i (located in city j in country k) in year t.
$Flood_{jt+s}$ is a dummy indicating whether or not city j was hit by a large flood in year $t+s$.
$Precip > 500mm_{jt+s}$ indicates locations that experienced monthly precipitation of 500mm or more in year $t+s$.
All regressions include year fixed effects, gridpoint fixed effects and country-specific trends.
Columns (6) to (10) corrected by the Arellano-Bond methodology using $ln(lights_{t-2})$ as an instrument for $ln(lights_{t-1})$.
Robust standard errors, clustered by country, in parentheses.

Table A4: Recovery, gridpoint year panel, extreme precipitation (1000mm) indicator

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$	$ln(Y_{ijkt})$
$Precip > 1000mm_{jt}$	-0.080					-0.083				
	(0.018)					(0.018)				
$Precip > 1000mm_{jt-1}$		0.054					0.053			
		(0.033)					(0.034)			
$Precip > 1000mm_{jt-2}$			0.004					0.002		
			(0.020)					(0.019)		
$Precip > 1000mm_{jt-3}$				0.002					0.004	
				(0.013)					(0.012)	
$Precip > 1000mm_{jt-4}$					0.001					0.003
					(0.029)					(0.028)
$ln(light_{t-1})$						Yes	Yes	Yes	Yes	Yes
Observations	1,422,018	1,422,018	1,422,018	1,422,018	1,422,018	1,392,501	1,392,501	1,392,501	1,392,501	1,392,501
No. of gridpoints	243,303	243,303	243,303	243,303	243,303	235,460	235,460	235,460	235,460	235,460

Notes: The results in this Table are variations on Equation 3 and use the sample of cities affected by at least one of the large flood events in our data.
The dependent variable in all regressions $ln(Y_{ijkt})$ is the natural log of mean light intensity at each gridpoint i (located in city j in country k) in year t.
$Flood_{jt+s}$ is a dummy indicating whether or not city j was hit by a large flood in year $t+s$.
$Precip > 1000mm_{jt+s}$ indicates locations that experienced monthly precipitation of 1000mm or more in year $t+s$.
All regressions include year fixed effects, gridpoint fixed effects and country-specific trends.
Columns (6) to (10) corrected by the Arellano-Bond methodology using $ln(lights_{t-2})$ as an instrument for $ln(lights_{t-1})$.
Robust standard errors, clustered by country, in parentheses.

Table A5: Recovery, city-year panel

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$	$ln(Y_{jkt})$
$Flood_{jt}$	-0.017					-0.019				
	(0.007)					(0.006)				
$Flood_{jt-1}$		-0.003					-0.008			
		(0.014)					(0.019)			
$Flood_{jt-2}$			0.017					0.017		
			(0.016)					(0.015)		
$Flood_{jt-3}$				0.004					-0.008	
				(0.009)					(0.021)	
$Flood_{jt-4}$					0.014					0.014
					(0.009)					(0.011)
$ln(light_{t-1})$						Yes	Yes	Yes	Yes	Yes
Observations	10,363	10,281	10,315	10,352	10,338	9,878	9,869	9,833	9,785	9,796
No. of urban areas	1,817	1,814	1,820	1,818	1,819	1,702	1,707	1,707	1,703	1,712

Notes: The results presented in this Table correspond to Equation 4 and use the sample of cities affected by at least one of the large flood events in our data.
The dependent variable in all regressions $ln(Y_{jkt})$ is the natural log of mean light intensity in each city j (located in country k) in year t.
$Flood_{jt+s}$ is a dummy indicating whether or not city j was hit by a large flood in year $t+s$.
All regressions include year fixed effects, city fixed effects and country-specific trends.
Columns (6) to (10) corrected by the Arellano-Bond methodology using $ln(lights_{t-2})$ as an instrument for $ln(lights_{t-1})$.
Robust standard errors, clustered by country, in parentheses.

Table A6: Interactions with elevation, extreme precipitation (1000mm) indicator

	(1) $ln(Y_{ijkt})$	(2) $ln(Y_{ijkt})$	(3) $ln(Y_{ijkt})$	(4) $ln(Y_{ijkt})$	(5) $ln(Y_{ijkt})$	(6) $ln(Y_{ijkt})$	(7) $ln(Y_{ijkt})$	(8) $ln(Y_{ijkt})$	(9) $ln(Y_{ijkt})$
$Precip > 1000mm_{jt} \times elev_{<10i}$	-0.120			-0.122			-0.100		
	(0.019)			(0.020)			(0.016)		
$Precip > 1000mm_{jt} \times elev_{10+i}$	-0.052			-0.056			-0.031		
	(0.022)			(0.021)			(0.011)		
$Precip > 1000mm_{jt-1} \times elev_{<10i}$		0.111			0.111			0.117	
		(0.015)			(0.016)			(0.017)	
$Precip > 1000mm_{jt-1} \times elev_{10+i}$		0.020			0.018			0.033	
		(0.035)			(0.035)			(0.030)	
$Precip > 1000mm_{jt-2} \times elev_{<10i}$			0.006			0.007			-0.022
			(0.009)			(0.014)			(0.011)
$Precip > 1000mm_{jt-2} \times elev_{10+i}$			0.003			0.001			-0.005
			(0.028)			(0.026)			(0.023)
$ln(light_{t-1})$				Yes	Yes	Yes			
Observations	1,422,018	1,422,018	1,422,018	1,392,501	1,392,501	1,392,501	1,422,018	1,422,018	1,422,018
No. of gridpoints	243,303	243,303	243,303	235,460	235,460	235,460	243,303	243,303	243,303

Notes: The results presented in this Table correspond to Equation 5 and use the sample of cities affected by at least one of the large flood events in our data.
The dependent variable in all regressions $ln(Y_{ijkt})$ is the natural log of mean light intensity at each gridpoint i (located in city j in country k) in year t.
$Precip > 1000mm_{lt+s}$ indicates locations that experienced monthly precipitation of 1000mm or more in year $t+s$.
$Elevation_h$ is a dummy for elevation band h, where h is either less than 10m above sea level, or 10m or more above sea level.
All regressions include year fixed effects and gridpoint fixed effects.
Columns (1) to (6) include country-specific trends. Columns (7) to (9) include city-specific trends.
Columns (4) to (6) corrected by the Arellano-Bond methodology using $ln(lights_{t-2})$ as an instrument for $ln(lights_{t-1})$.
Robust standard errors, clustered by country, in parentheses.

Table A7: Interactions with elevation, excluding locations within 10km of rivers and coasts

	(1) $ln(Y_{ijkt})$	(2) $ln(Y_{ijkt})$	(3) $ln(Y_{ijkt})$	(4) $ln(Y_{ijkt})$	(5) $ln(Y_{ijkt})$	(6) $ln(Y_{ijkt})$
$Flood_{jt} \times elev_{<10i}$	-0.030			-0.032		
	(0.007)			(0.006)		
$Flood_{jt} \times elev_{10+i}$	-0.020			-0.021		
	(0.013)			(0.014)		
$Flood_{jt-1} \times elev_{<10i}$		0.015			-0.009	
		(0.011)			(0.011)	
$Flood_{jt-1} \times elev_{10+i}$		-0.001			-0.016	
		(0.012)			(0.017)	
$Flood_{jt-2} \times elev_{<10i}$			0.037			0.038
			(0.017)			(0.017)
$Flood_{jt-2} \times elev_{10+i}$			0.017			0.017
			(0.014)			(0.014)
$ln(light_{t-1})$				Yes	Yes	Yes
Observations	814,294	810,524	812,476	795,536	792,801	790,097
No. of gridpoints	139,712	139,683	140,298	134,640	134,600	134,474

Notes: The results presented in this Table correspond to Equation 5 and use the sample of cities affected by at least one of the large flood events in our data, restricted to exclude gridpoints within 10km of the nearest river or coast.
The dependent variable in all regressions $ln(Y_{ijkt})$ is the natural log of mean light intensity at each gridpoint i (located in city j in country k) in year t.
$Flood_{jt+s}$ is a dummy indicating whether or not city j was hit by a large flood in year $t+s$.
$Elevation_h$ is a dummy for elevation band h, where h is either less than 10m above sea level, or 10m or more above sea level.
All regressions include year fixed effects and gridpoint fixed effects and country-specific trends.
Columns (4) to (6) corrected by the Arellano-Bond methodology using $ln(lights_{t-2})$ as an instrument for $ln(lights_{t-1})$.
Robust standard errors, clustered by country, in parentheses.

Table A8: Interactions with elevation, excluding cities entirely less than 10m above sea level

	(1) $ln(Y_{ijkt})$	(2) $ln(Y_{ijkt})$	(3) $ln(Y_{ijkt})$	(4) $ln(Y_{ijkt})$	(5) $ln(Y_{ijkt})$	(6) $ln(Y_{ijkt})$
$Flood_{jt} \times elev_{<10i}$	-0.021			-0.022		
	(0.006)			(0.007)		
$Flood_{jt} \times elev_{10+i}$	-0.019			-0.020		
	(0.012)			(0.012)		
$Flood_{jt-1} \times elev_{<10i}$		0.012			-0.006	
		(0.008)			(0.010)	
$Flood_{jt-1} \times elev_{10+i}$		-0.007			-0.022	
		(0.014)			(0.018)	
$Flood_{jt-2} \times elev_{<10i}$			0.046			0.045
			(0.016)			(0.016)
$Flood_{jt-2} \times elev_{10+i}$			0.007			0.007
			(0.011)			(0.011)
$ln(light_{t-1})$				Yes	Yes	Yes
Observations	1,379,280	1,372,088	1,375,024	1,351,484	1,345,342	1,339,592
No. of gridpoints	235,874	235,861	236,793	228,408	228,358	228,261

Notes: The results presented in this Table correspond to Equation 5 and use the sample of cities affected by at least one of the large flood events in our data, restricted to exclude cities that are entirely less than 10m above sea level.
The dependent variable in all regressions $ln(Y_{ijkt})$ is the natural log of mean light intensity at each gridpoint i (located in city j in country k) in year t.
$Flood_{jt+s}$ is a dummy indicating whether or not city j was hit by a large flood in year $t + s$.
$Elevation_h$ is a dummy for elevation band h, where h is either less than 10m above sea level, or 10m or more above sea level.
All regressions include year fixed effects and gridpoint fixed effects and country-specific trends.
Columns (4) to (6) corrected by the Arellano-Bond methodology using $ln(lights_{t-2})$ as an instrument for $ln(lights_{t-1})$.
Robust standard errors, clustered by country, in parentheses.

Table A9: Lights, GDP and floods for a panel of Indian districts

	(1) Ln light	(2) Ln light	(3) Ln GDP	(4) Ln GDP	(5) Ln light	(6) Ln GDP	(7) Ln light	(8) Ln light	(9) Ln GDP	(10) Ln GDP
Ln GDP	0.441	0.481			0.189					
	(0.039)	(0.027)			(0.051)					
Ln light			1.108	0.932		0.168				
			(0.117)	(0.104)		(0.121)				
Flood_t							-0.020		-0.005	
							(0.014)		(0.013)	
$\text{Flood}_t - 1$								-0.023		0.015
								(0.027)		(0.017)
Year 2003 only	Yes	Yes	Yes	Yes						
State f.e.		Yes		Yes						
District f.e.					Yes	Yes	Yes	Yes	Yes	Yes
Year f.e.					Yes	Yes	Yes	Yes	Yes	Yes
Observations	491	491	491	491	3,157	3,157	1,683	1,192	1,804	1,285

Notes: This table shows the correlation between log lights and log GDP in the year 2003 (Columns 1-4), and for an annual panel (2003-2008, columns 5-6) at the level of districts in India. The Table also reports results of regressions of log light and log GDP on a flood indicator using the district level panel data for India (Columns 7-10).
The light variable gives the mean light measure for each district. Flood is a dummy variable indicating if any gridpoint in the district was subject to one of the floods in our dataset in each year.
Robust standard errors are clustered at the level of state throughout the table.

Figure A1: This figure shows the coefficients from Columns 1-5 from Table 5, and their 95 percent confidence intervals. The figure includes additional years before the flood not shown in the table. Year 0 indicates the year of the flood.

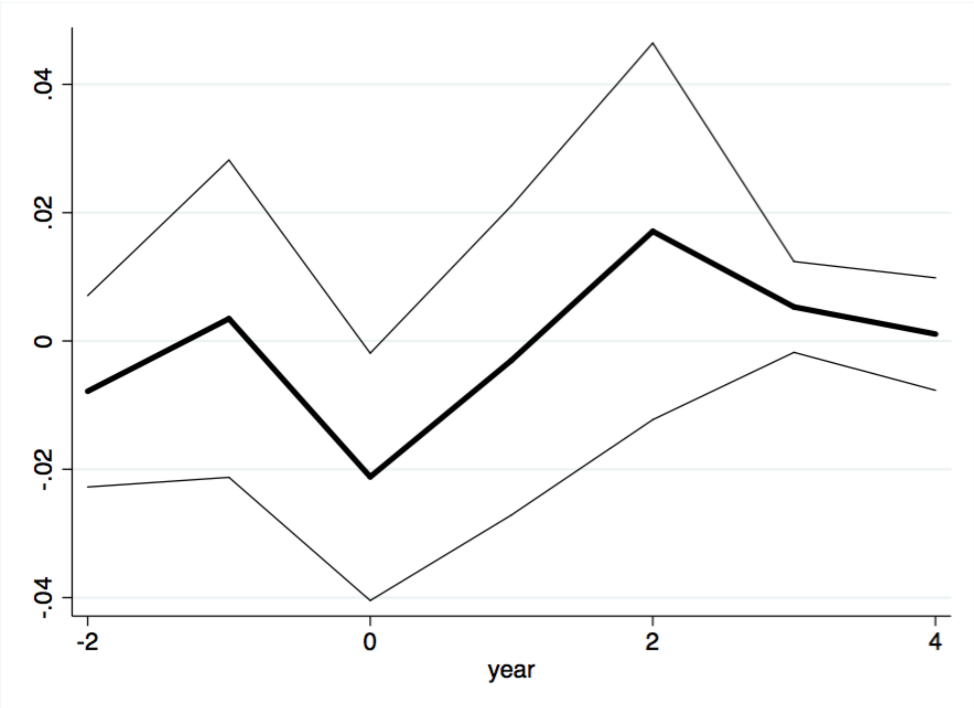

Figure A2: This figure shows the impact of a flood observed at time 0 separately for urban areas in our sample that we define as New, and poorly lit ("Poorlit") as follows: "New" areas are defined as having light = 0 in the year 1992. "Poorlit" is defined as areas with light > 0 and <= some cutoff in the year 1992. In Panels A and B we chose arbitrary, round numbers as light cutoffs (10 and 20) for poorly lit areas. In Panel C we chose the cutoff such that the resulting number of observations is close to the number of observations in the "new" category. This turns out to be at <= 8. Median light intensity in our main dataset of urban areas is 14, the 25th percentile value is 7 and the 75th percentile 31. The regression specification to which these coefficients correspond is similar to the specification used for Table 7. The figure includes additional years before the flood not shown in the table.