Supplementary Information for

Rainfall anomalies are a significant driver of cropland expansion

Esha Zaveri, Jason Russ, Richard Damania

Email: ezaveri@worldbank.org, jruss@worldbank.org, rdamania@worldbank.org

This includes:

Supplementary text
Figs. S1 to S4
Tables S1 to S17
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1910719117

Supplementary Information

S1 Examining impact of irrigation dams

This section describes the construction of a global dataset of access to irrigation from irrigation dams. For each gridcell, we estimate the total number of large, upstream dams within specific distance thresholds from the gridcell centroids. The universe of such facilities come from the Global Reservoir and Dam (GRanD) v1 dataset, from SEDAC (Lehner *et al.* 2011a and Lehner *et al*. 2011b). GRanD contains all reservoirs with storage capacities greater than 0.1km^3, with 6,862 in total. However, we only include those whose main, major, or secondary purpose is irrigation, which leaves 2,039 such facilities (this is after removing 142 for which no construction year is available). Fig. S3 shows their locations.

We first estimate the "command" gridcells of each dam – defined as potential recipients of irrigation services (Duflo and Pande, 2007). To be considered part of a dam's command area, a gridcell must satisfy three criteria. First, the gridcell centroid must be within the same river sub-basin as the dam, thus ensuring a hydrological link. We use the *World map of major hydrological basins (*derived from *HydroSHEDs)* from FAO (2015) to determine river-basin boundaries. Second, the gridcell centroid must be at a lower elevation than the dam reservoir. This ensures that the irrigation canals can flow into the gridcell by force of gravity. Finally, the gridcell must be within a certain distance threshold from the reservoir. The literature does not provide much guidance on how far a command area might extend. For instance, Duflo and Pande (2007) assume that the district in which the dam was built encompasses both the catchment and command area of the dam (and therefore has ambiguous impacts) and that the neighboring downstream district represents the command area.[1] Following the existing literature we assume that areas close to any storage facility are part of the catchment area. Hence, gridcells which are less than 25km from the facility—are deemed to be part of the catchment area. We then empirically establish the spatial threshold over which agricultural benefits accrue. After experimentation with different spatial thresholds, we find that at distances of 25-50km downstream the impact on agricultural productivity is largest. These criteria are shown diagrammatically in Fig. S4.

[1] This threshold is non-constant and will change based on the size of the downstream district, which can be as small as 9 km^2 (Mahe district in the state of Puducherry) or as large as 45,632 km^2 (Kutch district in the state of Gujarat). In other papers, this distance threshold can vary widely. For instance, You *et al.* (2011) used 150km as the maximum distance for a dams command area, although they admit that this was arbitrarily chosen.

From the GRanD dataset we take the year following construction as the year a facility becomes operational, allowing for adjustments by farmers. We also note that the majority of these were constructed long before our samples begin, with only 40% of them constructed after 1970, 10 years before the beginning year of our cropland dataset, and less than 1% constructed after 2000, the beginning year of our NPP dataset. We must therefore rely more on within country spatial variation, than variation over time, for identification.

Identification strategy

To examine how access to irrigation infrastructure can mitigate or exacerbate weather anomalies, we estimate the following equations for the intensive and extensive margins are shown in equations (1) and (2):

(1):

$$\Delta \log(NPP_{it}) = \alpha_1 + \alpha_2 * upirr_{it} + \alpha_3 Prec_{it}^{-} + \alpha_4 Prec_{it}^{-} * upirr_{it} + \alpha_5 Prec_{it}^{+} + \alpha_6 Prec_{it}^{+} * upirr_{it} + X_{it}'\lambda + f_c(t) + \theta_t + \gamma_c + \varepsilon_{it};$$

(2):

$$\Delta \log(Crop_{it}) = \alpha_1 + \alpha_2 * upirr_{it} + \alpha_2 Prec10_{it}^{-} + \alpha_2 Prec10_{it}^{-} * upirr_{it} + \alpha_3 Prec10_{it}^{+} + \alpha_3 Prec10_{it}^{+} * upirr_{it} + X_{it}'\lambda + f_c(t) + \theta_t + \gamma_c + \varepsilon_{it};$$

where $upirr_{it}$ measures the number of upstream irrigation facilities from gridcell *i* in year *t*. $upirr_{it}$ is therefore an estimate of the number of dams whose command areas impact gridcell *i*.

For the cropland analysis, $Prec10_{it}^{-}$ and $Prec10_{it}^{+}$ are converted to binary variables. Rather than measuring the number of years with anomalies in the past decade like in the baseline model, they are instead transformed into binary variables, to facilitate interpretation. We show results for 2+, 3+, and 4+ years of anomalies, out of the last 10 in the main text.[2]

A well-known problem that arises when estimating equations such as (1) and (2) is the non-random placement of irrigation infrastructure- these are likely built in areas where they will have the largest impact on agricultural production. Therefore, estimating equations (1) and (2) via OLS will lead to biased estimates.

To account for inherent placement biases, we adopt an instrumental variable strategy that exploits geographical characteristics of a region, a technique commonly followed in the literature (Duflo and Pande, 2007; Strobl and Strobl, 2011). The instruments we use are intended to predict the suitability for irrigation. We calculate three variables: the total length of rivers within the 25-50 km buffer around each gridcell, the share of these rivers with a slope suitable for irrigation, and a national propensity for irrigation infrastructure construction following Duflo and Pande (2007).

The first instrument measures the total length of rivers which reside within the 25-50km distance buffer (the same 25-50km distance buffer used with our irrigation measure). We use the USGS Hydrosheds River Database to obtain a global shapefile of

[2] These counts were chosen because they are common enough to make statistical inferences on the difference between regions with upstream irrigation facilities and without. When moving beyond 4+ years, the sample that received these anomalies becomes too small to detect differences between irrigated regions and non-irrigated regions and coefficients become noisy.

rivers.[3] River access is a necessary component of storage and irrigation and so they are an important predictor of irrigation infrastructure. Since the 25-50km buffer resides outside of the gridcell of interest, it should be uncorrelated with agricultural potential within the gridcell to the extent that spatial spillovers are negligible, and therefore satisfy the exclusion restriction.

The second instrument measures the slopes of the rivers which reside within the same 25-50km distance buffer. Specifically, it is the share of rivers which have a gentle slope gradient, between 1.5% and 3%. According to the engineering literature, irrigation requires a gentle river slope in order to create a long reservoir and to allow the water to reach the irrigated area via gravity. If the river gradient is too steep, the flow of water can erode the canals that transport water to the command area (Cech, 2010). This is also the slope gradient that Duflo and Pande (2007) and Strobl and Strobl (2011) found to best predict irrigation dam construction. Likewise, we tested other river gradient slopes and found 1.5-3% to be the best fit.

Finally, while the first two instruments measure the suitability of irrigation infrastructure at the gridcell level, the final instrument captures ex-ante variation in irrigation infrastructure allocation at the country level, which serves as a proxy for the experience or propensity for construction in a country. Following Duflo and Pande (2007) and Strobl and Strobl (2011), we construct $\overline{D}_{ct} = \left(\frac{g_{c,s-10}}{\sum_c g_{c,s-10}}\right) * g_t$, where $g_{c,s-10}/\sum_c g_{c,s-10}$ is the ratio of large dams in each country to total large dams in year *s*-10, where *s* is the starting year of the dataset, and g_t is the total number of global dams in each year t. Taking the country-to-global ratio 10 years prior to the start of our dataset ensures its exogeneity. We consider country-level variation in irrigation infrastructure allocation propensity, while Duflo and Pande (2007) use state-level, and Strobl and Strobl(2011) use river-basin level variation, reflecting differences in geographic scope and data availability.[4] Moreover, to ensure that our results are robust to differences in growth patterns across gridcells with different river gradients and river lengths in different countries we control for country-year time trends.

The first stage for equation (1) which estimates the impact on NPP, and equation (2) which estimates the impact on cropland, are given in equations (3) and (4), respectively:

$$Y_{it} = \beta_1 + \beta_2(RG_i * \overline{D}_{ct}) + \beta_3(RL_i * \overline{D}_{ct}) + \beta_4\left(RG_{i,} * \overline{D}_{ct} * Prec_{it}^{-,+}\right) + \eta_{it}, \quad (3)$$

$$Y_{it} = \beta_1 + \beta_2(RG_i * \overline{D}_{ct}) + \beta_3(RL_i * \overline{D}_{ct}) + \beta_4\left(RG_{i,} * \overline{D}_{ct} * Prec10_{it}^{-}\right) + \eta_{it}; \quad (4)$$

where $Y_{it} = upirr_{it}$ or $upirr_{it} * Precip_{it}^{-,+}$, RG_i is the share of river gradients between 1.5-3% within the 25-50km buffer around gridcell *i*, $\overline{D}_{ct}$ is the national propensity for dam construction in country *c* at year *t*, and RL_i is the total length of rivers within the 25-50km buffer around gridcell *i*. We therefore have three endogenous variables in each

---

[3] In some limited cases, particularly in extreme latitudes, the Hydrosheds River database does not cover the entire landmass. In these instances, we supplemented the dataset by calculating a synthetic river network using a Drainage Line Processing tool in QGIS.

[4] Duflo and Pande's analysis was limited to India and Strobl and Strobl's analysis was focused on Africa while our study is global.

regression ($upirr_{it}, upirr_{it} * Precip^{-}_{it}, upirr_{it} * Prec10^{+}_{it}$), and four instrumental variables.

As stated above, our instrumental variable and identification strategy closely follows the literature on the impact of dams. Nevertheless, instruments that work in one setting and context may not hold when applied to another. We therefore verify that our instruments matter in terms of predicting dam construction and present results from a reduced first-stage in table S16. This differs from the standard first-stage from a two-stage least square (2SLS) model in that it does not include the interaction terms between the rainfall anomalies and the instruments, which makes interpretation challenging.

Formally, we estimate the following equation for both crop productivity and cropland data sets:

$$(upirr_{it}) = \beta_1 + \beta_2(RG_i * \overline{D}_{ct}) + \beta_3(RL_i * \overline{D}_{ct}) + \theta_{it}'\gamma + \eta_{it}; \quad (5)$$

where θ_{it} includes all exogenous covariates from equations (3) and (4).

Results from estimating equation (5) are shown in table S16. Columns 1 and 2 show results from the NPP sample, and columns 3 and 4 show results over a different period used in the cropland sample. We find evidence for the importance of a gentle river gradient and river length in increasing dam construction. Note that in all specifications, dam construction is increasing in river length, in the share of river slopes between 1.5% and 3%, and their interaction with a country's propensity for dam construction, as theory and the literature predict. Further, these two variables are always jointly significant, according to an F-test.

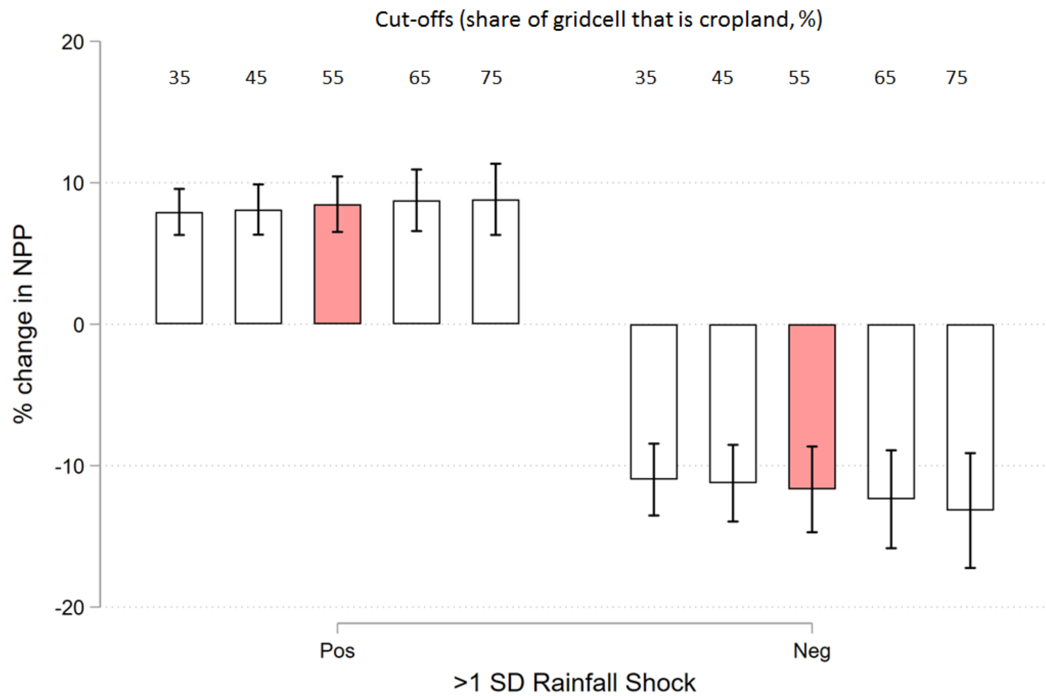

Fig. S1. Impact of rainfall anomalies on agricultural productivity, different ESA cropland thresholds. This figure shows percent change in NPP and 95 percent confidence intervals from coefficients obtained for 5 different thresholds. Standard errors are clustered at the province-year level. Each bar represents percent change in NPP as a result of a wet or dry rainfall anomalies (shocks) when using samples corresponding to 5 different cropland thresholds. Negative and positive coefficients are estimated jointly using separate regressions for each threshold. All regressions include grid and year fixed effects, country-specific trends and population

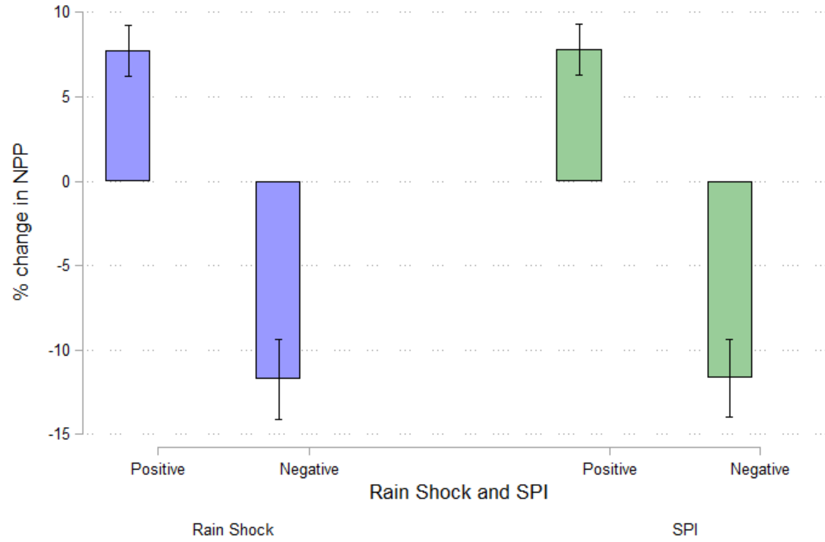

Fig. S2. Impact of rainfall anomalies on agricultural productivity, standardized precipitation index. This figure shows percent change in NPP and 95 percent confidence intervals. Standard errors are clustered at the province-year level. Each bar represents percent change in NPP as a result of a wet or dry rainfall anomalies (shocks) or SPI anomalies (shocks). Negative and positive coefficients are estimated jointly using separate regressions for each threshold. Bars in green show rainfall anomalies calculated using the Standardized Precipitation Index. All regressions include grid and year fixed effects, country-specific trends and population.

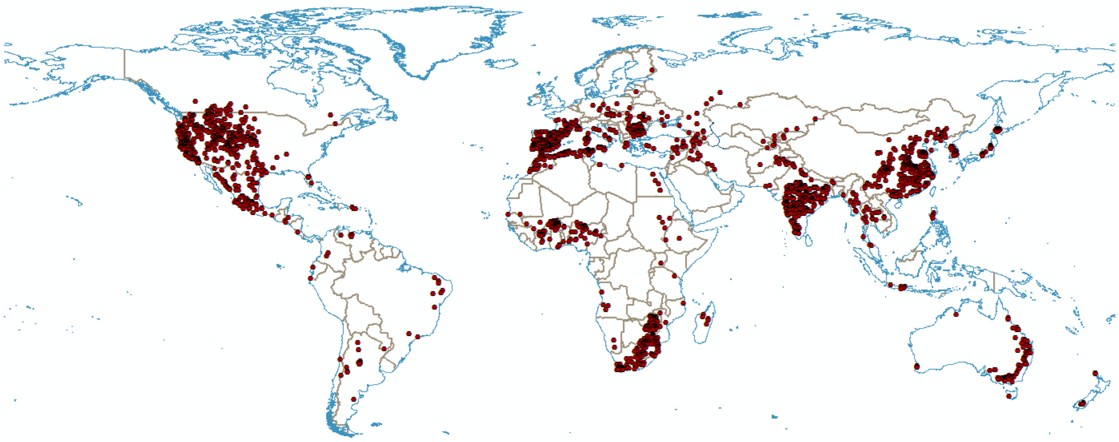

Fig. S3. GRanD large irrigation dam locations

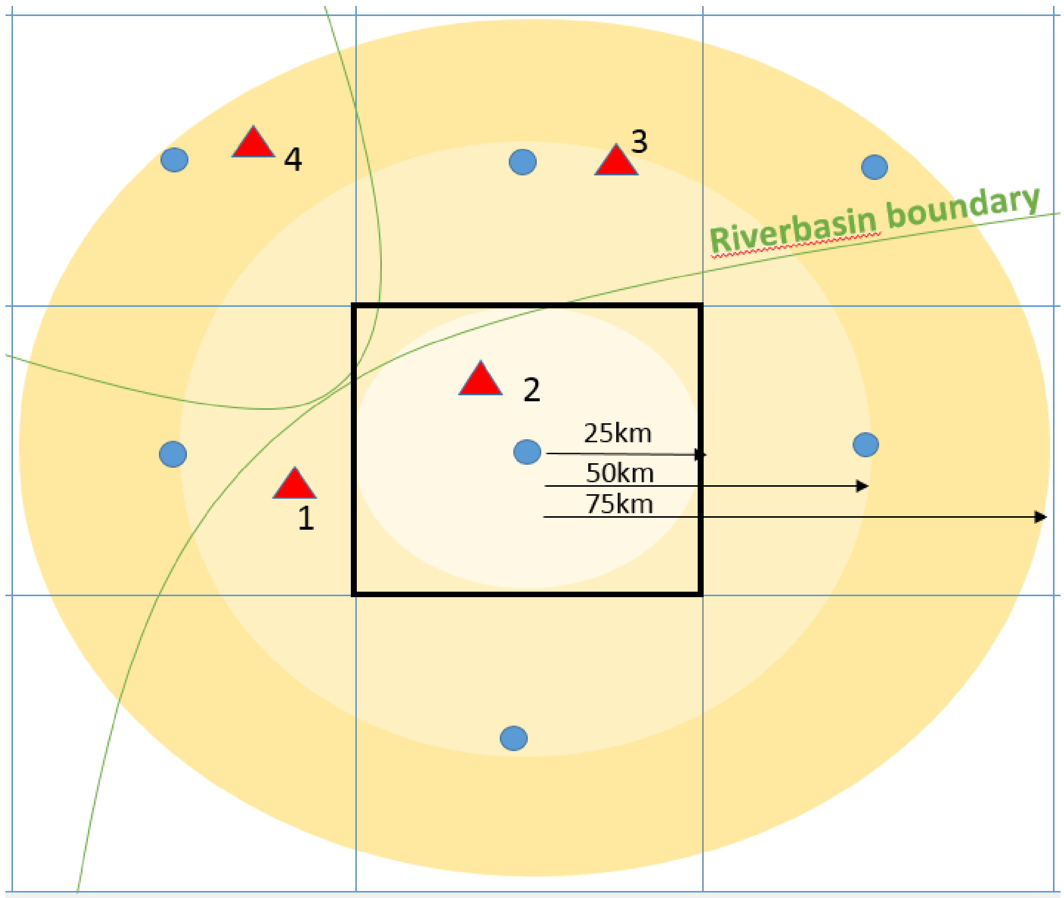

Fig. S4. Defining the command area of dams. In this figure, the gridcells represent 0.5 x 0.5 degree grids (approximately 50kms x 50 kms at the equator) used in our sample. Gridcell centroids are marked by circles, and dams are labeled by triangles. Concentric circles around the centroid of the selected gridcell show distance thresholds of 25, 50, and 75km. There are 4 dams within the vicinity of the selected gridcell. Dams 3 and 4 cannot have command areas which impact the highlighted gridcell because they are in separate riverbasins. Because Dam 2 falls within the same riverbasin as the selected gridcell, but since it is within 25km of the gridcell centroid, it is also excluded. Dam 1, which falls within the same riverbasin as the selected gridcell, and is within 25-50km of the gridcell centroid, will be considered to have a command area which impacts that gridcell as long as it is at a higher elevation than the gridcell centroid.

Table S1: Rainfall anomalies

Distribution of rainfall anomalies in each gridcell over 10 years

# Anomalies-Years	Wet	Dry
	# Gridcell-Year observations impacted	
0	282,005	345,119
1	292,103	300,792
2	219,921	195,069
3	130,924	100,935
4	62,095	43,743
5	24,368	19,617
6	7,709	9,638
7	2,314	4,235
8	564	2,154
9	346	892
10	108	263

Notes: Table shows the distribution of the number of wet (positive) and dry (negative) anomalies experienced by gridcells in our cropland sample over the 10 year period prior to observation. Anomalies are defined as deviations from the long run average by using the z-score. Dry anomalies are defined as z-scores below -1, and wet anomalies are defined as z-scores above +1. The sample includes 58,009 gridcells from 1992-2015.

Number of anomalies by income group and agroecological zone

Average Number of Anomalies in 10 year period	Full Sample	High Income	Developing	Arid	Humid
Dry Anomalies	1.373	1.408	1.369	1.231	1.728
Wet Anomalies	1.495	1.676	1.398	1.553	1.354
Net Anomalies (Wet - Dry)	0.122	0.267	0.0293	0.321	-0.374

Table S2: Impact of rainfall anomalies on crop productivity

Dependent variable:	Agricultural Productivity				
	(1)	(2)	(3)	(4)	(5)
Wet Rainfall Anomalies	**0.0863*****	**0.0775*****	**0.0773*****	**0.0785*****	**0.0756*****
	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
Dry Rainfall Anomalies	**-0.1303*****	**-0.1175*****	**-0.1174*****	**-0.1190*****	**-0.1207*****
	(0.013)	(0.012)	(0.012)	(0.011)	(0.012)
Average Temperature (C)		**-0.0322***	**-0.0321***	**-0.0019**	**-0.0024**
		(0.016)	(0.016)	(0.003)	(0.003)
Average Temperature (C) ^2		**-0.0011**	**-0.0011**	**0.0000**	**0.0001**
		(0.001)	(0.001)	(0.000)	(0.000)
Year fixed effects	Yes	Yes	Yes	Yes	No
Gridcell fixed effects	Yes	Yes	Yes	No	No
Country fixed effects	No	No	No	Yes	Yes
Country x time trend	Yes	Yes	Yes	Yes	Yes
Population	Yes	No	Yes	Yes	Yes
Gridcell characteristics	No	No	No	Yes	Yes
N	87958	87958	87958	87932	87932
R-sq	0.089	0.0998	0.0999	0.08	0.0465

Notes: Dependent variable is the change in log NPP in a cropland gridcell. Each column presents the regression coefficients from a separate regression. Standard errors reported in parentheses are clustered at the province-year level. Wet (Dry) rainfall anomalies is a time-varying dummy indicating if annual precipitation in a year is at least 1 standard deviation higher (lower) than the long run mean of a gridcell. Columns (4) and (5) also control for time-invariant gridcell characteristics related to presence of unconsolidated aquifer, CTI index and standard deviation of elevation in place of gridcell fixed effects. Statistical significance is given by + $p<0.10$ * $p<0.05$ ** $p<0.01$ ***$p < 0.001$

Table S3: High-income and developing countries, World Bank income group classifications

Developing			High Income	
Afghanistan	Guatemala	North Korea	Australia	Spain
Albania	Guinea	Pakistan	Austria	Sweden
Algeria	Guinea-Bissau	Palestina	Bahamas	Switzerland
Angola	Guyana	Panama	Belgium	Taiwan
Armenia	Haiti	Papua New Guinea	Brunei	Trinidad and Tobago
Azerbaijan	Honduras	Paraguay	Canada	United Arab Emirates
Bangladesh	India	Peru	Chile	United Kingdom
Belarus	Indonesia	Philippines	Croatia	United States
Belize	Iran	Romania	Cyprus	Uruguay
Benin	Iraq	Russia	Czech Republic	
Bhutan	Ivory Coast	Rwanda	Denmark	
Bolivia	Jamaica	Senegal	Estonia	
Bosnia and Herzegovina	Jordan	Serbia	Finland	
Botswana	Kazakhstan	Sierra Leone	France	
Brazil	Kenya	Solomon Islands	Germany	
Bulgaria	Kosovo	Somalia	Greece	
Burkina Faso	Kyrgyzstan	South Africa	Hungary	
Burundi	Laos	South Sudan	Iceland	
Cambodia	Lebanon	Sri Lanka	Ireland	
Cameroon	Lesotho	Sudan	Israel	
Central African Republic	Liberia	Suriname	Italy	
Chad	Libya	Swaziland	Japan	
China	Macedonia	Syria	Kuwait	
Colombia	Madagascar	Tajikistan	Latvia	
Congo, Dem. Republic	Malawi	Tanzania	Lithuania	
Congo, Republic	Malaysia	Thailand	Luxembourg	
Costa Rica	Mali	Timor-Leste	Netherlands	
Cuba	Mauritania	Togo	New Caledonia	
Djibouti	Mauritius	Tunisia	New Zealand	
Dominican Republic	Mexico	Turkey	Northern Cyprus	
Ecuador	Moldova	Turkmenistan	Norway	
Egypt	Mongolia	Uganda	Oman	
El Salvador	Montenegro	Ukraine	Poland	
Equatorial Guinea	Morocco	Uzbekistan	Portugal	
Eritrea	Mozambique	Vanuatu	Puerto Rico	
Ethiopia	Myanmar	Venezuela	Qatar	
Fiji	Namibia	Vietnam	Saudi Arabia	
Gabon	Nepal	Yemen	Slovakia	
Gambia	Nicaragua	Zambia	Slovenia	
Georgia	Niger	Zimbabwe	South Korea	
Ghana	Nigeria			

Table S4: Impact of rainfall anomalies on cropland expansion in developing countries, alternative cropland datasets

	ESA Developing	MODIS 1 Developing	MODIS 2 Developing
Dependent variable:		Cropland Area	
	(1)	(2)	(3)
No. Wet Anomalies	0.000079	-0.000036	-0.000035
	(0.000)	(0.000)	(0.000)
No. Dry Anomalies	0.000321*	0.000772***	0.000687***
	(0.000)	(0.000)	(0.000)
Year Fixed Effects	Yes	Yes	Yes
Cell Fixed Effects	Yes	Yes	Yes
Country x time trend	Yes	Yes	Yes
Control for population	Yes	Yes	Yes
N	704664	508924	508924
R-sq	0.338	0.075	0.069

Notes: Dependent variable is the 5 year average annual change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression based on gridcells in developing countries as per the World Bank classification. Standard errors are clustered at the province-year level. Number of wet (dry) anomalies in 10 years is a time-varying count variable indicating the number of years, of the prior ten years, for which annual precipitation in the gridcell was at least 1 standard deviation higher (lower) than the long run mean of the gridcell. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001.

Table S5: Impact of rainfall anomalies on cropland expansion in developing countries, alternative weather dataset from CRU

	Full Sample	High Income	Developing
Dependent variable:	Cropland Area		
	(1)	(2)	(3)
No. Wet Anomalies	0.000075	0.000168	0.000020
	(0.000)	(0.000)	(0.000)
No. Dry Anomalies	0.000482***	0.000252	0.000513**
	(0.000)	(0.000)	(0.000)
Year Fixed Effects	Yes	Yes	Yes
Cell Fixed Effects	Yes	Yes	Yes
Country x time trend	Yes	Yes	Yes
Control for population	Yes	Yes	Yes
N	784252	226184	540764
R-sq	0.394	0.376	0.398

Notes: Dependent variable is the 5 year average annual change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression. Standard errors clustered at the province-year level are reported in parentheses. Rainfall anomalies are calculated using CRU data. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001

Table S6: Impact of rainfall anomalies on cropland expansion in developing countries, Standardized Precipitation Index

	Rainfall anomalies	SPI
Dependent variable:	Cropland Area	
	(3)	(4)
No. Wet Anomalies	0.000079	-0.000299*
	(0.000)	(0.000)
No. Dry Anomalies	0.000321*	0.000247*
	(0.000)	(0.000)
Year Fixed Effects	Yes	Yes
Cell Fixed Effects	Yes	Yes
Country x time trend	Yes	Yes
Control for population	Yes	Yes
N	704664	861256
R-sq	0.338	0.079

Notes: Dependent variable is the 5 year average annual change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression. Standard errors clustered at the gridcell level are reported in parentheses. Column (1) shows the main baseline results. Rainfall anomalies are calculated using the Standardized Precipitation Index in column 2. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001

Table S7: Impact of rainfall anomalies on cropland expansion, changes over different time periods

	(1)	(2)	(3)	(4)	(5)
No. wet anomalies in 10 years	-0.000265*	-0.000242+	-0.000168	-0.000020	0.000079
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
No. dry anomalies in 10 years	0.000273*	0.000385**	0.000337**	0.000369**	0.000321*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of years for cropland average change	1	2	3	4	5
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Cell Fixed Effects	Yes	No	Yes	No	Yes
Country x time trend	Yes	Yes	Yes	Yes	Yes
Control for population, contemporaneous temperature, precipitation	Yes	Yes	Yes	Yes	Yes
N	861256	822108	782960	743812	704664
R-sq	0.079	0.145	0.210	0.274	0.338

Notes: Dependent variable is the 5 year average annual change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression based on number of years for cropland average change. Standard errors reported in parentheses are clustered at the gridcell level. Number of wet (dry) anomalies in 10 years is a time-varying count variable indicating the number years, of the prior ten years, for which annual precipitation in the gridcell was at least 1 standard deviation higher (lower) than the long run mean of the gridcell.
Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001.

Table S8: Impact of rainfall anomalies on cropland expansion, using growing season data from Sub-Saharan Africa

Dependent variable:	Cropland Area
No. Wet Anomalies (growing season)	0.000087
	(0.000207)
No. Dry Anomalies (growing season)	0.000227+
	(0.000124)
Year Fixed Effects	Yes
Cell Fixed Effects	Yes
Country x time trend	Yes
Control for population	Yes
N	69513
R-sq	0.261

Notes: Dependent variable is the 5 year average annual change in log Cropland in a gridcell. Standard errors are clustered at the gridcell level. Number of wet (dry) anomalies in 10 years is a time-varying count variable indicating the number of years, of the prior ten years, for which annual precipitation in the gridcell was at least 1 standard deviation higher (lower) than the long run mean of the gridcell. Statistical significance is given by + $p<0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$.

Table S9: Impact of rainfall anomalies on cropland expansion, controlling for distance to city

	Full Sample	High Income	Developing
Dependent variable:		Cropland Area	
	(1)	(2)	(3)
No. Wet Anomalies	0.000096	-0.000059	0.000267+
	(0.000110)	(0.000072)	(0.000157)
No. Dry Anomalies	0.000189+	-0.000047	0.000472**
	(0.000100)	(0.000076)	(0.000154)
Distance to nearest City	0.000260+	-0.001060***	0.000398*
	(0.000148)	(0.000131)	(0.000177)
Year Fixed Effects	Yes	Yes	Yes
Cell Fixed Effects	Yes	Yes	Yes
Country x time trend	Yes	Yes	Yes
Additional Controls		log(population), temperature, precipitation, CTI elevation	
N	1021917	295545	704232
R-sq	0.032	0.017	0.037

Notes: Dependent variable is the 5 year average annual change in log Cropland in a gridcell. Standard errors are clustered at the province-year level. Number of wet (dry) anomalies in 10 years is a time-varying count variable indicating the number of years, of the prior ten years, for which annual precipitation in the gridcell was at least 1 standard deviation higher (lower) than the long run mean of the gridcell. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001.

Table S10: Impact of rainfall anomalies on cropland expansion, accounting for commercial-scale agriculture

	(1) Developing	(2) Developing
Dependent variable:	Cropland Area	
No. Wet Anomalies	0.000104	0.000104
	(0.000)	(0.000)
No. Wet Anomalies * Soy/Palm Gridcell	-0.000830	-0.000830
	(0.001)	(0.001)
No. Dry Anomalies	0.000323+	0.000323*
	(0.000)	(0.000)
No. Dry Anomalies * Soy/Palm Gridcell	-0.000113	-0.000113
	(0.001)	(0.001)
Year Fixed Effects	Yes	Yes
Cell Fixed Effects	Yes	Yes
Country x time trend	Yes	Yes
Control for population	Yes	Yes
N	704664	704664
R-sq	0.338	0.338

Notes: Dependent variable is the 5 year average annual change in log Cropland in a gridcell. Standard errors are clustered at the province-year level in column (1) and gridcell level in column (2). Soy/Palm Gridcell is an indicator for whether soy/palm occupies the largest amount of harvested area in the gridcell derived from Monfreda et al. (2008). Number of wet (dry) anomalies in 10 years is a time-varying count variable indicating the number of years, of the prior ten years, for which annual precipitation in the gridcell was at least 1 standard deviation higher (lower) than the long run mean of the gridcell. Statistical significance is given by + $p<0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$.

Table S11: Impact of cropland expansion on forest area

Column 1 shows estimates from a simple univariate regression which predicts changes in forest cover based on changes in cropland. Column 2 adds gridcell and year fixed effects. Column 3 adds contemporaneous temperature and rainfall, population, and country time trends. Finally, column 4 adds the indicators of the number of rainfall anomalies in the past 10 years, making it fully consistent with the deforestation regression estimated in Table 1, column 4

	(1)	(2)	(3)	(4)
Dependent variable:	Forest Area			
Δ log (Cropland)	-0.101053***	-0.075690***	-0.071550***	-0.071427***
	(0.009)	(0.007)	(0.007)	(0.007)
No. Wet anomalies				0.000816+
				(0.000)
No. Dry anomalies				-0.001078**
				(0.000)
Annual Precipitation			0.000002	0.000002
			(0.000)	(0.000)
Annual Avg Temperature			-0.000432	-0.000466
			(0.001)	(0.001)
Log Population			-0.000742	-0.000005
			(0.004)	(0.004)
Year Fixed Effects	No	Yes	Yes	Yes
Cell Fixed Effects	No	Yes	Yes	Yes
Country x time trend	No	No	Yes	Yes
Control for population	No	No	Yes	Yes
N	744819	744819	704664	704664
R-sq	0.001	0.307	0.322	0.322

Notes: Dependent variable is the 5 year average annual change in log forest area in a gridcell. No. wet anomalies (dry anomalies) denotes number of wet (dry) anomalies in 10 years indicating the number years, of the prior ten years, for which annual precipitation in the gridcell was at least 1 standard deviation higher (lower) than the long run mean of the gridcell. Standard errors are clustered at the province-year level. Statistical significance is given by + $p<0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$.

Table S12: Impact of irrigation infrastructure on agricultural productivity and cropland expansion, OLS results

Dependent variable:	Agricultural Productivity	Cropland Area		
		X=2	X=3	X=4
	(1)	(2)	(3)	(4)
No. Upstream Dams	0.0006	-0.000469***	-0.00065***	-0.00074***
	(0.002)	(0.000)	(0.000)	(0.000)
Wet Anomalies	0.0854***			
	(0.003)			
Wet Anomalies x No.Upstream Dams	-0.0131*			
	(0.005)			
X+ Wet Anomalies		0.000296*	0.000497**	0.001035***
		(0.000)	(0.000)	(0.000)
X+ Wet Anomalies x No. Upstream Dams		-0.000635***	-0.000365*	-0.000531*
		(0.000)	(0.000)	(0.000)
Dry Anomalies	-0.1174***			
	(0.004)			
Dry Anomalies x No.Upstream Dams	0.0241***			
	(0.006)			
X+ Dry Anomalies		0.000731***	0.001095***	0.001390***
		(0.000)	(0.000)	(0.000)
X+ Dry Anomalies x No. Upstream Dams		-0.000028	-0.000099	0.000272
		(0.000)	(0.000)	(0.000)
N	64922	704232	704232	704232

Notes: Dependent variable in column 1 is the change in log NPP in a gridcell. Dependent variable in columns 2-4 is the 5-year average annual change in log Cropland area in a gridcell All regression models include controls for year and country fixed effects as well as temperature, population, country-year trends, terrain roughness, aquifer presence and compound topographic index CTI. Columns 2-4 also control for annual precipitation. X+ Wet (Dry) rainfall anomalies is a time-varying dummy variable indicating if rainfall was at least 1 standard deviation higher (lower) than the long run mean of the gridcell for X or more years out of the prior ten years, where X is given in the 2nd row. Upstream dams is a count variable which indicates the number of upstream dams from the gridcell. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001.

Table S13: Cropland Classes

Data Source	Cropland Classes
European Space Agency (ESA)	Rainfed cropland [10, 11, 12] Irrigated cropland [20] Cropland and natural vegetation mosaic [30]
MODIS (MCD12Q1) IGBP scheme	Croplands [12] Cropland and natural vegetation mosaic [14]
University of Maryland scheme (Hansen et al. 2000)	Croplands [12] Cropland and natural vegetation mosaic [14]

Table S14: Impact of rainfall anomalies on cropland expansion in developing countries, with alternative controls for population

	(1)	(2)	(3)	(4)
Dependent variable:	Cropland Area			
No. Wet anomalies	0.000079	0.000093	-0.000240	-0.000168
	(0.000146)	(0.000142)	(0.000171)	(0.000166)
No. Dry anomalies	0.000321*	0.000297+	0.000721***	0.000671***
	(0.000163)	(0.000162)	(0.000189)	(0.000188)
log(Population)	-0.006962***		-0.012410***	
GPW version 3	(0.001916)		(0.002670)	
log(Population)				-0.012633***
GPW version 4				(0.002297)
Time Period	1992-2015	1992-2015	2000-2015	2000-2015
Year Fixed Effects	Yes	Yes	Yes	Yes
Cell Fixed Effects	Yes	Yes	Yes	Yes
Country x time trend	Yes	Yes	Yes	Yes
N	704664	705618	587220	580575
R-sq	0.338	0.338	0.382	0.383

Notes: Dependent variable is the 5 year average annual change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression for the developing country sample. Standard errors clustered at the province-year level are reported in parentheses. Columns (1) shows the main result from table 1 which employs GPW version 3 as the population variable. Column (2) removes the population variable. Column (3) restricts the sample to the time period of 2000-2014, which are the years for which GPW version 4 is available, for comparison purposes. Column (4) then repeats the specification in Column (3) but replaces the population variable with GPW version 4. Statistical significance is given by + $p<0.10$ * $p<0.05$ ** $p<0.01$ ***$p < 0.001$

Table S15: Impact of net rainfall anomalies on cropland expansion

	Full Sample	High Income	Developing
Dependent variable:		Cropland Area	
	(1)	(2)	(3)
No. Positive Anomalies - No. Negative Anomalies	-0.000071	-0.000011	-0.000120
	(0.000084)	(0.000062)	(0.000129)
Year Fixed Effects	Yes	Yes	Yes
Cell Fixed Effects	Yes	Yes	Yes
Country x time trend	Yes	Yes	Yes
Control for population	Yes	Yes	Yes
N	1022457	295545	704664
R-sq	0.332	0.297	0.338

Notes: Dependent variable is the 5 year average annual change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression based on gridcells in developing countries as per the World Bank classification. Standard errors are clustered at the province-year level. Number of wet (dry) anomalies in 10 years is a time-varying count variable indicating the number of years, of the prior ten years, for which annual precipitation in the gridcell was at least 1 standard deviation higher (lower) than the long run mean of the gridcell. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001.

Table S16: Summary Statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
Intensive Margin					
Δ Log NPP	88,231	0.008791	0.283744	-7.33378	7.62405
Wet rainfall anomalies	95,018	0.165884	0.371979	0	1
Dry rainfall anomalies	95,018	0.138511	0.345437	0	1
Upstream irrigation dams (developing world)	69,944	0.091402	0.406766	0	6
Extensive Margin					
Δ Log Cropland	1,022,457	0.0045893	0.0472116	-1.5986	1.55424
Number of wet anomalies in 10 years	1,022,457	1.351798	1.433918	0	10
Number of dry anomalies in 10 years	1,022,457	1.531698	1.405528	0	10
log(Population)	1,022,457	8.126324	3.716066	-14.289	17.377
Contemporaneous Annual Rainfall (mm)	1,022,457	703.7321	688.5652	0	13437.4
Contemporaneous Mean temperature (C)	1,022,457	10.70241	13.53565	-28.05	37.883
Upstream irrigation dams (developing world)	704,232	0.0188106	0.1814095	0	6

Table S17: Geography and upstream dams

	(1)	(2)	(3)	(4)
	2000-2013	2000-2013	1992-2014	1992-2014
River length * $\bar{D}_{ct}$	0.000177***	0.000162***	0.000008***	0.000014***
	(0.000)	(0.000)	(0.000)	(0.000)
River slope share *$\bar{D}_{ct}$	0.000209***	0.000264***	0.000029**	0.000057***
	(0.000)	(0.000)	(0.000)	(0.000)
Wet rainfall anomalies		-0.004145		
		(0.005)		
Dry rainfall anomalies		0.006199		
		(0.006)		
Average Temp(C)		0.006069***		
		(0.001)		
Average Temp(C) Sq.		0.000244***		
		(0.000)		
4+ Wet rainfall anomalies				-0.002383***
				(0.000)
4+ Dry rainfall anomalies				-0.003014***
				(0.001)
Contemporaneous Annual Rainfall (mm/year)				-0.000005***
				(0.000)
Contemporaneous Mean temperature (C)				0.000917***
				(0.000)
Log(Population)		0.032536***		0.005498***
		(0.003)		(0.000)
Year fixed effects	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Country specific trends	Yes	Yes	Yes	Yes
F-test for river length & slope	67.122	64.677	13.26	23.10
N	92428	92428	1388100	1306519
R-sq	0.094	0.108	0.058	0.066

Notes: Dependent variable is the number of upstream dams in a gridcell. Cluster-robust standard errors are reported in parentheses. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001

References

Duflo E, Pande R (2007) Dams. *Quarterly Journal of Economics* 122(2): 601-646.

Cech, Thomas V. 2010. *Principles of water resources: history, development, management, and policy*. New Jersey: John Wiley & Sons

Siebert S, Kummu M, Porkka M, Döll P, Ramankutty N, Scanlon BR (2015) A global data set of the extent of irrigated land from 1900 to 2005. *Hydrology and Earth System Sciences* 19(3): 1521-1545

Strobl E, Strobl RO (2011) The distributional impact of large dams: Evidence from cropland productivity in Africa. *Journal of Development Economics* 96(2): 432-450.

Lehner B et al. (2011a) Global Reservoir and Dam Database, Version 1 (GRanDv1): Dams, Revision 01.

Lehner B et al. (2011b) High-Resolution Mapping of the World's Reservoirs and Dams for Sustainable River-Flow Management. *Frontiers in Ecology and the Environment* 9 (9): 494-502. http://dx.doi.org/10.1890/100125.

You L et al. (2011) What is the irrigation potential for Africa? A combined biophysical and socioeconomic approach. *Food Policy* 36(6): 770-782.