Drenched fields and parched farms: Evidence along the extensive and intensive margins

Esha Zaveri[1] Jason Russ[2] Richard Damania[3]

[PRELIMINARY DRAFT. COMMENTS WELCOME]

Abstract: Climate models predict an increase in rainfall variability, higher frequency of extreme water events, and deeper water scarcity over the coming decades. These changes could have profound impacts on the global agricultural sector. In this paper, we explore how rainfall variability impacts agricultural production along two margins: the intensity of output (yields) and the extensiveness of production, and how water infrastructure influences this relationship. Using global, gridded datasets on net primary productivity (NPP), land cover, and weather, we find that, on average, contemporaneous wet shocks tend to increase agricultural productivity. Contemporaneous dry shocks decrease crop productivity, while repeated dry shocks also tend to increase the rate of cropland expansion, perhaps as an adaptation technique to compensate for lower yields. We argue that the theoretical underpinnings for these results can be found in the "safety-first" model, where the priority of the economic agent is to generate a threshold level of income or output. Further, using an instrumental variables based identification strategy for dam construction, we find that the buffering impact of upstream irrigation dams varies by geography, climate, and income levels. Upstream dams, in general, decrease the extent of cropland expansion to persistent dry shocks across different climatic zones and income levels, yet in developing countries with arid climates, they appear to accentuate the adverse effects of dry shocks on agricultural productivity. One possible reason for this relationship which we find evidence for is mal-adaptation, where the presence of dams incentivize farmers to plant water-intensive crops in otherwise unsuitable areas. Other potential reasons which we discuss include the management of dams, particularly if preference is given to hydropower generation during lean seasonal flows, or more generally if water storage is inadequate. We also find that having access to vast groundwater reserves dampens the negative effects of dry shocks on cropland productivity. Differentiating between uses and sources can, therefore, provide key insights into the margins along which irrigation can buffer productivity from rainfall variability.

Keywords: Rainfall shocks, agriculture, adaptation, dams, aquifers, food security

[1] Postdoctoral Fellow, Center on Food Security and the Environment, Stanford University, ezaveri@stanford.edu
[2] Economist, Water Global Practice, World Bank, jruss@worldbank.org
[3] Global Lead Economist, Water Global Practice, World Bank, rdamania@worldbank.org

JEL Codes: O13, Q15, Q25, Q28, Q54

1. Introduction

The variability of rainfall is a recurring challenge in the agricultural sector. For thousands of years, humans have struggled to adapt to the unpredictable nature of climactic variations. Indeed, written accounts of how to manage rainfall variability date back to the ancient treatise *Arthashatra*, written in the 4th century BCE, by the Indian scholar *Kautilya*, who discussed ways to predict a beneficial rainy season, and how this should translate into cropping decisions and expected yields (Kautilya and Rangarajan, 1992).[4] This important issue continues to occupy center stage in policy discussions on food security. Looking into the future, the problem could worsen with the collision of two 21st century transitions – growing populations that propel an increase in the demand for food and water, coupled with a changing climate that renders rainfall more erratic and less predictable. A deeper understanding of the consequences of these trends and the effectiveness of remedies that seek to buffer economies from rainfall variability will be helpful in finding more effective responses to these problems.

These issues are of policy significance for at least three reasons. First, if water scarcity or adverse rainfall shocks have large quantitative impacts that disrupt agriculture, the impacts could cascade to other parts of the economy with wider implications. Second of greater concern, many of the poorest countries of the world, with rapidly expanding populations and elevated levels of water stress, also endure strong inter-annual variability of rainfall which compromises their ability to meet escalating domestic food demand. Third, this has environmental implications. The current extensification path to increasing food supply is rapidly becoming unsustainable as land grows more scarce and the forest, rangeland, and wetlands margins become exhausted. As a consequence, land clearing has emerged as one of the major contributors to climate change and is believed to be responsible for about 6-17% of anthropogenic CO_2 emissions (Baccini et al., 2012).

[4] Quoted in Barnett, C. (2016). *Rain: A Natural and Cultural History*. Broadway Books, New York.

In this paper we conduct a global, geographically disaggregated empirical analysis at the gridcell level covering 175 countries over the years 1980-2013 to investigate some of these issues in greater detail. We examine how plausibly exogenous rainfall shocks impact observed patterns of agricultural productivity (measured by net primary productivity) and cropland expansion. Moreover, to cope with rainfall shocks and variability, investments in water infrastructure remain key to climate change adaptation plans in many countries where agricultural dependence remains high (Narain et al., 2011). Yet, there is limited evidence about the effectiveness of existing irrigation infrastructure in mitigating against these shocks, at a global disaggregated scale. In this paper we also explore whether the presence of infrastructure such as irrigation dams, or deep-water aquifers, affects the impacts of precipitation shocks. By examining the outcomes of interest directly – a physical measure of yields on the intensive margin and changes in cropped area and land use on the extensive margin, we avoid relying on assumptions about mechanisms that might lead to these changes.

There is a vast theoretical literature on farmer responses to various forms of risk, but the models do not offer simple or unambiguous predictions. The response to shocks in most theoretical contexts depends upon attitudes towards risk and assumptions about behavior, all else equal. For instance, standard expected utility theory models predict that there will be a reduction in economic activity when risks rise. Intuitively, since the likelihood of good states of the world occurring declines with greater risk, risk averse agents seek to reduce their exposure to more hazardous outcomes.[5] But this result could be overturned if behavior is better described by a "safety-first" model, where the priority of the economic agent is to generate a minimum (perhaps survival) threshold level of income. In such circumstances, higher risks could induce farmers to extend cultivated areas in an attempt to meet their threshold level of income. The empirical

[5] To see why consider a CRS production function y = $g(l,w,z)$, where l is land, w is normally distributed stochastic rainfall and z any other input. Assume for convenience that l and w are used in fixed proportions but this is not a necessary assumption. Assume too that agents are risk neutral for simplicity. Now define $l^* = \text{Argmax}(Eg(l,w,z) - C)$; where C is a convex and well behaved cost function, and E is the expectation operator. Define r as an index of a mean preserving increase in uncertainty. Then applying Theorem 1 of Diamond and Stiglitz (1974) it follows that $\partial l^*/\partial r < 0$.

literature is uneven too with greater attention paid to impacts on yields on the intensive margin rather than the responses on the extensive margin.

Our main identification strategy is straightforward. It involves examining the consequences of exogenous deviations of rainfall from its mean in each grid cell. Thus, a grid cell observation in a year of "normal" rainfall acts as a control for the same grid cell observed in a year of a "deviation" or "shock". By focusing on exogenous fluctuations in rainfall we circumvent problems of conflating levels of rainfall with other features of an economy that have long run effects. If, for instance, better levels of rainfall attract more migrants, which in turn induces agglomeration effects, correlations between rainfall and economic outcomes would be conflated and biased upwards reflecting the consequences of features such as agglomeration rather than the weather. By using exogenous variations in rainfall our approach allows us to disentangle and identify causal effects.

Consistent with past literature, our results on the intensive margin show that, in general, contemporaneous wet (dry) shocks increase (decrease) crop productivity across regions with different climates and levels of development. The buffering impact of upstream irrigation dams, however, is largely heterogenous, and varies with geography, income levels, and the water-intensity of cropping patterns. Our results show that upstream irrigation dams exacerbate the negative effects of dry shocks in regions with lower levels of development and more arid climates. We find suggestive evidence that this effect is driven by mal-adaptation, and the planting of water-intensive crops in otherwise unsuitable regions. On the other hand, having access to large reserves of groundwater, has the potential to dampen the negative effects of dry shocks on productivity. Distinguishing between different uses and sources is, therefore, key to understanding the differential ability of irrigation to reduce the uncertainty associated with increased rainfall variability.

On the extensive margin, we find evidence that, in all but the wealthiest countries, repeated rainfall shocks do lead to increases in cropland expansion. This is particularly true in the case of dry shocks, and the impact increases when the shock is felt over an extended period of time. This is consistent with the hypothesis of cropland expansion being used as an adaptation technique to decreasing yields, in the face of fixed costs to agricultural expansion. Further, we find evidence that the presence of upstream irrigation dams does tend to mitigate this expansionary behavior, even in more arid and lower developed regions. We find these results to be consistent with both models of 'safety first' behavior, as well as risk aversion typically observed with farmers. This is discussed in greater detail in Section 5.

This paper is related to three prominent strands of literature: the rapidly expanding econometrics literature on climate change impacts on the economy; a somewhat older regional and agronomic literature on the effects of weather on crop yields, land-use, or both; and a recent prominent literature on the economic effects of hydrological infrastructure.

The econometric literature on the effects of climate change is vast (see Dell *et al* 2013 for a survey). Approaches at disentangling climate signals from other confounding factors have varied over time. The earlier literature used cross-sectional approaches and attempted to explore links between variations in climate across countries or sub regions and agricultural outcomes, with controls for factors such as geography and socioeconomic variation. Mendelsohn *et al* (1994) is the seminal paper and as in much of that literature, finds that both temperature and rainfall are important for agriculture.

Concerns about omitted variable bias and other identification problems have led to a greater emphasis on panel regression approaches in recent years (Auffhammer and Schlenker 2014, Dell *et al* 2013, Burke *et al* 2015). In general, much of this literature finds that higher temperatures have lasting and statistically significant negative effects on economic performance, with less clear evidence on rainfall. In many specifications, rainfall has been found to have no discernable robust impact (Dell *op cit,* Burke *et al op cit*).

It is difficult to reconcile this recurring result in the climate change literature, with the findings of crop specific and regional studies which consistently report that rainfall has a significant impact on agricultural outcomes. For instance, Auffhammer *et al* (2006) examine the response of rainfed kharif rice in India and find that both temperature and rainfall have impacts on yields, and that rainfall, but not temperature, affects harvested area. Fishman (2016) finds that increases in intra-seasonal variability of rainfall can have even more harmful effects than temperature increases on yields in India. A strong correlation between accumulated rainfall on the one hand and yields and planted rice areas on the other, is also observed in the Philippines (Koide *et al* 2013), for wet-season rice in Java, rice plantations in Indonesia (Naylor *et al* 2001), and rainfed lowland rice in Thailand (Sawano *et al* 2008). A recent global study by Lesk *et al* (2016) examines responses of yields and land-use patterns to climate shocks. The study finds that production losses due to drought are associated with a reduction in both harvested area and yields, suggesting that there are no compensating responses on the extensive margin.

One possible explanation for the diverging results is that findings may be sensitive to the outcome variable that is used. Much of the climate change literature uses GDP as a proxy for economic activity. But GDP may not be significantly affected by rainfall if the impacts are confined to the agricultural sector and the sector is a small component of the economy. Nevertheless, even when attention is restricted to agricultural GDP the rainfall signal remains muted and is often found to be statistically insignificant (Dell *et al op cit*). This suggests that measurement is not the only factor responsible for the differences. Another possibility is that aggregation at the country level could mask heterogeneity of impacts across regions and crops. For instance, Barrios *et al* (2010) find that in Sub-Saharan Africa higher rainfall is associated with faster growth. This may suggest that the baseline level of rainfall, the distribution and timing of rainfall, and factors such as infrastructure may matter. By this line of reasoning impacts in countries that are wealthy, wet, and well-endowed with hydrological infrastructure may not be the same as in countries that are arid and poor. This remains an area that warrants further research. Our results cast light on this issue by suggesting the need for spatial disaggregation to distinguish between countries and measures that more directly capture changes in agricultural output.

This paper is also related literature on the effects of infrastructure in buffering the impacts on rainfall variability. The seminal contributions are by Dufflo and Pande (2005, 2007) who pioneered the estimation and identification strategies that have been used in nearly all subsequent empirical work. Dufflo and Pande (2005) demonstrate that there are differences in the economic impacts in the control and command areas of dams in India, where the former tend to experience losses while the latter tend to experience gains. Using a more extensive global dataset with nighttime lights as a proxy for economic activity and a similar estimation approach, Olmstead and Sigman (2015) find that extreme droughts tend to dim lights and thus signal a reduction of activity. The OLS estimates indicate that dams buffer these impacts; but when instrumental variables are used to correct for endogeneity bias, it is found that dams worsen the impacts, suggesting maladaptation. Blanc and Strobl (2014) estimate and compare the effects of small and large irrigation dams on cropland productivity in South Africa and find that large dams increase cropland productivity downstream and have a negative effect on cropland within the vicinity of the dam. Small dams generate smaller benefits, but without these confounding negative effects. Our results shed some light on these findings, suggesting that the buffering effects are conditional upon the level of development, climate, and the types of crops that are cultivated.

The remainder of this paper is organized as follows. Section 2 describes the global datasets employed in this analysis. Section 3 describes our empirical strategy. Section 4 presents and discusses our mains results, with extensions and robustness checks provided in Section 5. We conclude with a brief discussion in Section 6.

2. Data

We use multiple georeferenced and gridded datasets related to agricultural productivity, agricultural land cover, weather, dams and aquifers to construct a global dataset encompassing 175 countries over the period 1980-2013. A description of these data is provided below.

2.1 Agricultural Data

2.1.1 Crop Productivity

In order to measure changes in agricultural productivity, and in turn, food security at a global grid-level scale, we require a measure that can provide a common unit of productivity across different crop types to facilitate comparison and aggregation over all types (Hicke, Lobell and Asner, 2004). Following the past literature in economics (Strobl and Strobl, 2011; Blanc and Strobl, 2013; Blanc and Strobl, 2014) and remote sensing (Lobell et al., 2002; Heinsch et al. 2005; Turner et al. 2006; Zhang et al., 2008) we use a satellite- based estimate of net primary production (NPP) as a proxy for crop productivity. NPP is linearly related to the amount of solar energy that plants absorb over a growing season (Running et al. 2004). Therefore, once an area has been identified as cropland, NPP can calculate the rate at which solar energy is converted into chemical energy during photosynthesis and stored as biomass in grams of carbon per square meter on that land. It also provides a measure of nutritional value since the availability of carbon stored in the form of plant material for food consumption can be roughly converted into kilocalories (Blanc and Strobl, 2013).[6] NPP, therefore, serves as the principal energy source for ecosystems and, in turn, for human populations that depend upon them (Abdi et al. 2014). There is also substantial evidence for the strong positive correlation between satellite-derived estimates of NPP and crop yields, providing further credibility to our measure of crop productivity.[7] Several researchers have validated the accuracy of NPP values from satellite data using NPP values derived from actual yield data, finding high correlation between the two (Lobell et al. 2002; Lu and Zhuang, 2010; Tum and Gunther 2011).[8]

Our time-varying NPP data comes from the moderate resolution imaging spectroradiometer (MODIS), whose data starts in 2000. MODIS is particularly suited to measure NPP due to the direct connection between absorbed solar energy and satellite-derived spectral indices of vegetation.[9] In particular, the MOD17 algorithm provides the first continuous, near-real-time measurement of global vegetation productivity from the EOS MODIS sensor (Running et al. 2004). In this study, we use the annual MOD17A3 measures from 2000-2013 generated by the Numerical Terradynamic Simulation Group (NTSG) at the University of Montana (Zhao et al., 2005) which corrects for cloud contamination prevalent in MODIS land products.[10]

---

[6] 1 gram of carbon is roughly equal to 9.33 kcal

[7] A closely related measure to NPP is the normalized difference vegetation index (NDVI). NDVI when combined with growing season data of different crops can also provide a measure of plant health and physical productivity that is directly related to NPP. For instance, MODIS-NPP is determined using NDVI along with other factors (Running et al. 2004), and in general, NDVI is considered a good predictor for NPP. However, without knowing the time-varying distribution of crops underlying the global land cover data, we cannot accurately estimate the corresponding growing season data and therefore, cannot measure the maximum NDVI during the growing season.

[8] Many studies also use actual harvested crop yield data to derive NPP values (Prince et al. 2001; Monfreda et al.2008)

[9] The advanced very high resolution radiometer (AVHRR) is a predecessor to MODIS and provides data going back to 1980. However, MODIS is considered a substantial improvement over AVHRR. It provides a finer instantaneous field of view and is built with seven bands that accurately monitor land cover with more robust spectral signatures (Hansen et al. 2003). It also limits the extent of background scattering from adjacent pixels (Hansen et al. 2003). To construct time-series data, we need a consistent measure of NPP across time. We, therefore, focus our analysis to the 2000-2013 period for which we have spatially detailed MODIS data.

[10] The improved MOD17 by the NTSG is a post-reprocessed dataset that corrects for cloud-contamination in NASA’s MOD17 dataset

Our interest is in estimating NPP from cropland. To facilitate this, we make use of the Global Land Cover 2000 (GLC2000) data to identify cropland areas. GLC2000 classifies land cover across the globe into 22 distinct land cover categories based on images acquired by the SPOT 4 satellite during the year 2000. We use the land cover categories “Cropland”, “Mosaic of Cropland/Shrub or Herbaceous Cover” and “Mosaic of Cropland/Tree Cover/Other Natural Vegetation” to identify agricultural crop areas within our grid cells. We depict the identified cropland areas in figure 1. Our final data measures changes in NPP for each 0.5 degree gridcell that falls in either of the cropland categories.

We use such a time-invariant crop area map from the beginning of our sample period in 2000, instead of also capturing any changes in cropland over the same sample period, to isolate and identify changes in productivity (intensive margin) separately from changes in cropland (extensive margin).[11]

2.1.2 Cropland

Our data on time-varying cropland data is sourced from the Earthstat global cropland data, developed by the Land Use and Global Environment Research Group at McGill University.[12] The data is a revision of an earlier cropland data set developed in Ramankutty and Foley (1999). It uses a cropland map for year 2000 (Ramankutty et al., 2008) that was created using remote-sensing data and land use statistics and combines this with sub-national cropland extent statistics and estimates based on a scaling approach methodology followed by Ramankutty and Foley (1999) to create a time-series of cropland extent at five year intervals. We source our cropland variable over the period 1980-2005[13] to measure changes in the log of cropland in each 0.5 degree gridcell.[14]

---

[11] Future drafts of this paper will also use other land cover maps in the literature such as GlobCover2009, MCDQ1 and land cover map by Ramankutty et al. (2008) as robustness checks.
[12] Several other studies (e.g., Hurtt et al., 2011; Klein Goldewijk et al., 2011; Houghton, 2008) have used different approaches in order to reconstruct gridded data sets of cropland, pastureland, urban land, wood harvest, etc. covering several centuries.
[13] Note that in specifications where we include population as a control variable, our dataset begins in 1990, due to restricted availability of that dataset.
[14] In order to include zero values when taking the log of cropland, we add 0.01 to each value.

2.2 Weather

Our weather data comes from Matsuura and Willmott (2001). This gridded dataset contains monthly observations of precipitation and average temperature at the 0.5 degree gridcell level. We transform this data into average monthly temperature, and total precipitation (mm), per year, for each gridcell. To define precipitation shocks, we calculate the long run mean and standard deviation of annual precipitation for each gridcell from 1900-2014. We then define a positive (negative) shock in a given year/gridcell if annual precipitation in that year/gridcell is at least 1 standard deviation higher (lower) than the long run mean for that gridcell.

2.3 Water infrastructure

2.3.1 Dams

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

Using these dams from GRanD, we attempt to estimate whether each gridcell is a potential beneficiary served by the dam for irrigation. As shown by Duflo and Pande (2005 and 2007), it is regions in the command areas of dams that tend to benefit from their construction. We therefore attempt to estimate which gridcells fall into the command areas of each of the irrigation dams in GRanD. To do so, we use three criteria. First, the gridcell centroid must be within the same river subbasin as the dam, thus ensuring that the dam and the gridcell are hydrologically linked. We use the *World map of major hydrological basins (derived from HydroSHEDs)* from FAO (2015) to determine riverbasin boundaries. Second, the dam must be at a higher elevation than the centroid of the gridcell. This ensures that the dammed river, or irrigation canals, can flow into the gridcell by force of gravity. Finally, the dam must be within a certain distance threshold from the gridcell. The literature does not provide much guidance on how far the command area of a dam can extend. For instance, Duflo and Pande (2005) 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.[15] Following the existing literature we assume that areas close to the dam are part of the catchment area. Hence, gridcells which are less than 25km from the dam—which for the majority of gridcells encompasses dams that fall into the gridcell itself—are part of the catchment area of the dam. 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 of a dam the impact on agricultural productivity is largest, as we show in section 5. These grid cells also satisfy the first two criteria, and are therefore likely to have command areas which fall within this threshold. This is shown diagrammatically in figure 3.

The GRanD dataset also contains the year that the dam was constructed, allowing us to create a panel. We take the year following construction as the year the dam becomes operational, allowing for adjustments by farmers and time needed for filling the dam. We also note that the majority of dams 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.

2.3.2 Aquifers

In addition to calculating the number of upstream irrigation dams in each gridcell, we also use aquifer maps to identify gridcells with naturally occurring large groundwater endowments. Our aquifer maps are sourced from the World-wide Hydrogeological Mapping and Assessment Programme (WHYMAP). Figure 4 maps deep and shallow aquifers along with their respective recharge rates. Unconsolidated aquifers in large sedimentary basins with good conditions for

[15] 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.

groundwater exploitation are depicted in blue while hydrogeological environments of complex structure are depicted in green. These are areas where productive aquifers may occur in close vicinity to non-aquiferous strata, yet have greater reserves of groundwater than shallow aquifers depicted in brown. Both of these types of aquifers, depicted in blue and green, together account for the large and extensive groundwater resources of the world.

2.4 Other data

We complement our dataset with additional gridcell characteristics related to population, water-intensive cropping, moisture regimes, and level of economic development.

Our population data comes from Gridded Population of the World (GPW) (CIESIN 2016). GWP offers population data at the 0.5 degree gridcell level for the entire globe, at 5 year intervals between 1990 and 2015. The data are linearly interpolated when annual observations are required as with the NPP analysis.

To identify gridcells that are water-intensive we use data on the geographical distribution of agricultural crops from Monfreda et al. (2008). They provide a 5 arc minute x 5 arc minute raster dataset encompassing 137 crops. For each cell in the raster, Monfreda et al. report harvested area in hectares. We aggregate the harvested area variable at the lower resolution of our dataset, i.e. 0.5 degree x 0.5 degree, and identify gridcells whose share of harvested area devoted to rice, cotton, or sugarcane—three of the most water-intensive crops according to water need and drought sensitivity (table 1, Brouwer and Heibloem, 1986)—is greater than the global median. These areas are shown in figure 5.

We also make use of a global spatial dataset of 18 agro-ecological zones to identify gridcells as falling into either arid or humid regions.[16] The dataset uses length of growing period (LGP) data from the IIASA/FAO GAEZ database that combines soil, climate and topographic information with a water balance model and knowledge of crop requirements to derive LGP. LGP describes

[16] https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=3184. The dataset combines length of growing periods along with a map of climatic zones, and moisture regimes to construct different agro-ecological zones.

the number of days during a growing season with adequate moisture and temperature to grow crops. Arid regions are areas where LGP is up to 179 days, and humid regions are areas where LGP ranges from 180 to > 300 days.

Finally, we use World Bank Income group classifications to divide the world into low-income, middle-income (which combines lower-middle and upper-middle income countries), and high-income countries. Classifications are based on mean per-capita GNI in 2015 where low income countries have GNI per capita below $1,025, middle income countries are between $4,036 and $12,475, and high income countries are above $12,475.

3. Empirical Strategy

In this section we lay out two distinct empirical strategies; the first is designed to understand the impact of precipitation shocks on the intensification and extensification of rainfall shocks, and the second examines how and whether infrastructure can mitigate these impacts.

3.1 Impact of precipitation shocks

We employ panel regression analysis to understand how variation in precipitation impacts agricultural productivity and cropland expansion. Our strategy relies on the fact that deviations from long-run precipitation are exogenous with respect to agricultural production in the short and medium term after controlling for observed and unobserved characteristics using cell fixed effects, year fixed effects, and a country-year interaction which neutralizes any country-level trends.

In order to estimate the impact of precipitation shocks on intensification, we estimate the following equation:

$$\Delta \log(NPP_{it}) = \alpha_1 + \alpha_2 Prec_{it}^{-} + \alpha_3 Prec_{it}^{+} + +X_{it}'\lambda + f_c(t) + \theta_t + \gamma_i + \varepsilon_{it} \qquad (1)$$

where NPP_{it} is net primary productivity in gridcell *i* in year *t*, $Prec_{it}^{-}$ ($Prec_{it}^{+}$) is a binary variable indicating if rainfall was at least 1 standard deviation below (above) the long run mean in gridcell *i* and year *t*, $f_c(t)$ are country-specific time trends, θ_t are year fixed effects, and γ_i are gridcell fixed effects. X_{it} is a vector of control variables which includes log of population, and a quadratic term for mean annual temperature (°C). Although we do not focus on the impact of temperature, we do control for it, in order to obtain unbiased estimates of the effects of changes in precipitation. This is important since temperature and precipitation are correlated (Auffhammer et al., 2013), and temperature has been shown in many studies to impact crop productivity (Schlenker et al., 2006; Schlenker and Roberts, 2009). α_2 and α_3 are our coefficients of interest and measure how a negative or positive precipitation shock, respectively, can contemporaneously impact the percentage change in NPP. We use cluster-robust standard errors that account for within-gridcell clustering of errors and arbitrary correlation of observations across time. Summary statistics for these variables are given in table 2.

To estimate the impact of precipitation shocks on cropland extensification, we modify this setup slightly to examine medium term outcomes. Expanding cropland, particularly when it requires spreading onto virgin fields or into forested areas, can require a large, upfront fixed cost. For this reason, we expect that contemporaneous impacts of variable rainfall will likely be muted, but repeated shocks over the medium term could induce extensification as an adaptation strategy. We therefore modify our precipitation shock variables in equation (1) to measure the *number* of years in the past 10 years[17] for which precipitation was at least 1 standard deviation above or below the mean. Table 3 gives the distribution of the number of precipitation shocks in the dataset. The distributions of positive and negative shocks are fairly similar, with more than half of the observations receiving zero or one positive or negative precipitation shock in the previous 10 years, and less than 3% of observations receiving 5 or more.

Formally, we estimate the following equation:

$$\Delta \log(Crop_{it}) = \alpha_1 + \alpha_2 Prec10_{it}^{-} + \alpha_3 Prec10_{it}^{+} + +X_{it}'\lambda + f_c(t) + \theta_t + \gamma_i + \varepsilon_{it} \qquad (2)$$

where $\Delta \log(Crop_{it})$ is the annual percent change[18] in cropland in gridcell *i* in year *t*, $Prec10_{it}^{-}(Prec10_{it}^{+})$ is the number of negative (positive) precipitation shocks greater than 1

[17] We chose 10 years as the cutoff as it is a long enough time period to exploit significant variation in the independent variable, and for impacts on the extensive margin to materialize. We verify that our regressions are robust to using an alternate 15-year window.
[18] Note that cropland is observed at 5 year intervals, so this value takes the 5-year percentage change and divides it by 5 to arrive at an annual value.

standard deviation within the last 10 years, and all other variables are the same as defined above. In addition to controlling for contemporaneous mean annual temperature (°C) and log population, as with equation 1, we also control for contemporaneous precipitation (mm/year). This is because sowing decisions could be informed by information about current year rainfall. Again, α_2 and α_3 are our coefficients of interest and measure how repeated negative or positive precipitation shocks, respectively, can impact the percentage change in cropland. Summary statistics for these variables are given in table 4.

3.2 Influence of dams and aquifers

Next, we examine how access to infrastructure impacts the effect of precipitation shocks on the intensive and extensive margins. Specifically, we examine how access to dams, and in the case of the intensive margin, access to groundwater, can mitigate or exacerbate weather shocks. To do so, we use a similar regression strategy as above, but also include the number of upstream irrigation dams in the regression equation, as well as interaction terms between the number of dams and the precipitation shock variables. To estimate the influence of groundwater endowments on mitigating sensitivity of NPP to precipitation shocks, we also include interactions between the presence of naturally occurring thick aquifers and precipitation shocks, Formally, the estimated equations for the intensive and extensive margins are shown in equations 3, 4 and 5:

$$\Delta\log(NPP_{it}) = \alpha_1 + \alpha_2 * updam_{it} + \alpha_3 Prec_{it}^{-} + \alpha_4 Prec_{it}^{-} * updam_{it} + \alpha_5 Prec_{it}^{+} + \alpha_6 Prec_{it}^{+} * updam_{it} + X_{it}'\lambda + f_c(t) + \theta_t + \gamma_c + \varepsilon_{it}; \quad (3)$$

$$\Delta\log(NPP_{it}) = \alpha_1 + \alpha_2 Prec_{it}^{-} + \alpha_3 Prec_{it}^{-} * aquifer_i + \alpha_4 Prec_{it}^{+} + \alpha_5 Prec_{it}^{+} * aquifer_i + X_{it}'\lambda + f_c(t) + \theta_t + \gamma_c + \varepsilon_{it}; \quad (4)$$

$$\Delta\log(Crop_{it}) = \alpha_1 + \alpha_2 * updam_{it} + \alpha_2 Prec10_{it}^{-} + \alpha_2 Prec10_{it}^{-} * updam_{it} + \alpha_3 Prec10_{it}^{+} + \alpha_3 Prec10_{it}^{+} * updam_{it} + X_{it}'\lambda + f_c(t) + \theta_t + \gamma_c + \varepsilon_{it}; \quad (5)$$

where $updam_{it}$ measures the number of upstream irrigation dams from gridcell *i* in year *t*, as described in section 2. $updam_{it}$ is therefore an estimate of the number of dams whose command areas impact gridcell i. $aquifer_i$ is a dummy variable that identifies gridcells that are naturally endowed with large groundwater reserves as explained in Section 2.3.2.

From the cropland analysis, one change that is made when moving from equation (2) to equation (5) is $Prec10^-_{it}$ and $Prec10^+_{it}$ are converted to binary variables. Rather than measuring the number of years with shocks in the past decade, they are instead transformed into binary variables, to facilitate interpretation.[19] In the main results we use 4 years of shocks as the cutoff, and in section 5.3 we show that results from other cutoffs are similar.

3.3 Instrumental variables for the effect of dams

A well-known problem that arises when estimating equations such as (3) and (5) is the non-random placement of dams and the fact that irrigation dams are likely built in areas where they will have the largest impact on agricultural production. Therefore, estimating equations (3) and (5) 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 of constructing dams upstream from each gridcell, while also taking into consideration the national propensity for dam construction. Specifically, 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 dam construction, and a national propensity for dam 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 dams measure). We use the USGS

[19] This is done because the interaction of two continuous variables can be difficult to interpret.

Hydrosheds River Database to obtain a global shapefile of rivers.[20] River access is a necessary component of dam construction and so they are an important predictor of dams. River access could potentially be correlated with agricultural outcomes, which would violate the exclusion restriction, however, since the 25-50km buffer resides outside of the gridcell of interest, it should be uncorrelated with agricultural activity within the gridcell to the extent that spatial spillovers are negligible[21].

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 dams require a gentle river slope in order to create a long reservoir in proportion to the height of the dam 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.[22]

Finally, while the first two instruments measure the suitability of dams at the gridcell level, the final instrument captures ex-ante variation in dam allocation at the country level, a proxy for the experience or propensity for dam construction in a country. Once again we closely follow Duflo and Pande (2007) and Strobl and Strobl (2011) to construct this measure. First we calculate the ratio of dams in each country to total global dams, 10 years prior to the beginning of the dataset (1990 for the NPP analysis, and 1970 for the cropland dataset). We then multiply this ratio by total global dams in each year. Taking the country-to-global dam ratio 10 years prior to the start of our dataset ensures its exogeneity. This instrument is similar to that used by Duflo and Pande (2007) and Strobl and Strobl (2011), but in our case, we consider country-level variation in dam

[20]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.
[21] In Section 5.5 we show that results are robust to clustering errors at larger geographic scales, a preliminary test for the likely negligible degree of spatial correlation. Future drafts will address spatial correlation more directly.
[22] Results for other river gradients omitted for brevity but available upon request.

allocation propensity, while Dufflo and Pande use state-level, and Strobl and Strobl use river-basin level variation, reflecting geographic scope and data availability (Duflo and Pande's analysis was limited to India and Strobl and Strobl's analysis was focused on Africa while our study is global). Moreover, to ensure that our results are robust to nondam-related differences in growth patterns across gridcells with different river gradients and river lengths in different countries we control for country-year time trends. The validity of our instrumental variable identification strategy is borne out by the Hansen J-test for overidentifying restrictions and instrument exogeneity, the Kleibergen-Papp LM test for instrument relevance and the Kleibergen-Papp Wald F-test for weak instruments in the results.

The first stage for equation (3) which estimates the impact on NPP, and equation (5) which estimates the impact on Cropland, are given in equations (6) and (7), respectively:

$$Y_{it} = \beta_1 + \beta_2(RG_i * \overline{D}_{ct}) + \beta_3(RL_i * \overline{D}_{ct}) + \beta_4\left(RG_{i,} * \overline{D}_{ct} * Prec_{it}^{-,+}\right) + \eta_{it}, \quad (6)$$

$$Y_{it} = \beta_1 + \beta_2(RG_i * \overline{D}_{ct}) + \beta_3(RL_i * \overline{D}_{ct}) + \beta_4\left(RG_{i,} * \overline{D}_{ct} * Prec10_{it}^{-}\right) + \eta_{it}; \; (7)$$

where $Y_{it} = updam_{it}$ or $updam_{it} * Precip_{it}^{-,+}$, 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 regression ($updam_{it}$, $updam_{it} * Precip_{it}^{-}, updam_{it} * Prec10_{it}^{+}$), 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 5. This differs from the standard first-stage from a 2SLS in that it does not include the interaction terms between the precipitation shocks and the instruments, which make the results difficult to interpret.

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

$$(updam_{it}) = \beta_1 + \beta_2(RG_i * \overline{D}_{ct}) + \beta_3(RL_i * \overline{D}_{ct}) + \theta_{it}'\gamma + \eta_{it}; \quad (8)$$

where θ_{it} includes all exogenous covariates from equations (3) and (5).

Results from estimating equation 8 are shown in table 5. 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 in the 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.

4. Results

In this section we first present results on the impact of precipitation shocks on agricultural intensification and extensification, from estimating equations (1) and (2). We then present results which estimate how dams can influence these impacts, by estimating equations (3) and (4).

4.1 Impact of Precipitation Shocks on Agricultural Intensification and Extensification

Results from estimating our baseline productivity and cropland models in equations (1) and (2) are presented in tables 6 and 7, respectively. In both tables, we estimate the model for the full global sample (column 1), as well as for gridcells in arid and humid regions separately (columns 2 and 3, respectively), and for high income, middle income and low incomes (columns 4, 5, and 6, respectively). Columns 7-9 are identical to columns 1-3 except they include country fixed effects, rather than gridcell fixed effects.

4.1.1 Crop Productivity

Table 6 shows that crop productivity, on average, increases in response to positive precipitation shocks and decreases with negative precipitation shocks across different climates and income groups. These effects are robust to an alternate specification that uses country fixed effects instead of gridcell fixed effects, with similar patterns observed; that is, the direction and significance of the impact remains unchanged and the magnitudes of the coefficients change only marginally.

The largest impacts of precipitation shocks are seen in arid regions and middle income countries, with negative shocks playing a significantly larger role. If farmers adapt to permanently drier conditions, one might expect that crop productivity is less sensitive to dry shocks in arid areas. However, our results suggest that the effects of below normal rainfall shocks are not just limited to humid regions that experience these shocks less frequently.[23] We find that rainfall more than 1 standard deviation below the long- term mean leads to a nearly 9%[24] decrease in agricultural productivity per year in arid regions (as opposed to a 3.5% decrease in humid regions), and close to a 7% decrease in middle- income countries. Above normal precipitation shocks, on the other hand, increase crop productivity by approximately 7% in arid regions and 5% in middle income countries. Consistent with past literature, we also find that temperature increases tend to reduce NPP non-linearly, particularly in arid regions, high-income and low income countries, but these impacts are not always consistent across different specifications. While we control for population in our regressions, its impact on NPP is imprecise and insignificant for most of the samples.

4.1.2 Cropland Expansion

Our results for the cropland model, displayed in table 7 show that the impact of positive precipitation shocks varies by climate and income group. In arid regions and in low-income countries, each year with a positive precipitation shock leads to a reduction in cropland by approximately 0.02% and 0.05%, respectively. In humid regions, and in middle income countries, positive precipitation shocks actually increase cropland expansion, by 0.2% per year with a shock. However, the coefficient on positive precipitation shocks becomes insignificant in humid areas when we replace gridcell fixed effects (column 3) with country fixed effects (column 7). Generally speaking, the impacts of positive precipitation shocks are unclear and seem highly dependent on the regression specification and sample (we also show in section 5.5

[23] Prior literature that has focused on temperature increases has also found no evidence for hotter regions experiencing systematically different impacts of temperature shocks on economic growth or yields (Schlenker and Roberts, 2009; Dell, Jones and Olken, 2012)

[24] The percentage impact of a dummy variable is given by $100\,[\exp\left(\hat{\beta} - \frac{1}{2}V(\hat{\beta})\right) - 1]$

that the impact of positive precipitation shocks are not robust to clustering standard errors at different levels).

For negative precipitation shocks, however, we see a much more robust impact with each year with a negative precipitation shock consistently leading to expansions in cropland, in both arid and humid regions, low- and middle-income countries, and in the full sample. The direction and significance of the impact remains unchanged when we switch to country fixed effects, and the magnitudes of the coefficients only differ slightly. In general, we find that for each year within the past 10 years which had a negative precipitation shock greater than 1 standard deviation, cropland expanded by approximately 0.06% on average across the globe and in arid areas, and by 0.1% in humid areas. Although these increases in cropland appear to be quite minor, the mean increase in cropland over the sample period was only 0.1%, implying that precipitation shocks may be a significant contributor to global extensification. It is perhaps unsurprising that we do not see an impact in high income countries, where the far majority of agriculture is irrigated, and farmers might have better access to other adaptation methods such as improved seeds.

The coefficients on contemporaneous rainfall and temperature, although sometimes statistically significant, are always very small implying that they have no actual impact on cropland expansion. Log population is only statistically significant in high income countries, or when gridcell fixed effects are removed, and the negative coefficient is likely because the more urban a gridcell becomes, the less room there is for cropland expansion.

We next allow for non-linear impacts of multiple precipitation shocks. To do so, we replace the variables which count the number of years with positive and negative precipitation shocks with binary variables which indicate if there were greater than X years with precipitation shocks within the past 10 years, with $1 \le X \le 8$. As shown in table 3, less than 1% of observations experienced 6 or more years with negative shocks, and only slightly more than 1% experienced 6 or more years with positive shocks. We therefore present estimates for up to 8 years with shocks, but only cautiously interpret impacts beyond 5 years.

We estimate equation (2) for each binary variable, separately, and plot the coefficients along with the 95% confidence intervals in figure 6. Interpreting first the negative shock coefficients we see that for all three samples, cropland expansion tends to increase approximately linearly with each additional year of shocks until about 5 years, at which point the impact tends to decline. Again, however, with so few observations beyond 5 years of shocks, the coefficients on 7+ and 8+ years of shocks become very noisy. Nevertheless, this is consistent with cropland expansion occurring as an adaptation technique against negative rainfall shocks. Given that there is a fixed cost to expanding cropland, it is unlikely to occur in large amounts after only one or two years of negative shocks. However, as they compound, and expectations about future rainfall are updated, the benefits to expansion increase and extensification becomes necessary to achieve a subsistence level of production.

The pattern for positive shocks is, again, less straight-forward. The full sample shows very modest increases in cropland expansion due to positive precipitation shocks, which do not increase significantly with the number of years for which the shock occurs. However, there is significant heterogeneity between arid and humid areas. Arid areas see virtually no impact on cropland expansion from positive shocks, with very precisely measured, small coefficients for nearly all years of shocks. In humid areas, however, the coefficient on positive shocks looks very much like the pattern we saw from the coefficient on negative shocks. In this case, expansion may be due to the fact that agricultural land becomes significantly more productive and it is profitable to extensify. Again, however, we do not find the impact of positive shocks to be robust, and therefore only interpret these results with caution.

To summarize, the most consistent and strongest results are obtained for negative precipitation shocks. These shocks reduce yields unambiguously. In both arid and humid regions of low-middle income countries, these negative shocks lead to compensating expansion of cultivated area, with a quantitative impact that is large – and of the same order of magnitude as current rates of deforestation. Theoretically such behavior is consistent with a “safety first” response which is most often associated with poor farmers who seek to achieve a certain target level of income. Results for positive precipitation are less robust. In general positive rainfall episodes boost yields, but the consequences for extensification are less well determined.

4.2 Influence of Irrigation Dams on Precipitation Impacts

In the following subsections, we present results from our agricultural productivity and cropland models which include the impact of dams and groundwater.

4.2.1 Influence of Irrigation Dams on the Relationship between Precipitation and Crop Productivity

Estimates of the influence of dams on the precipitation impacts on crop productivity are presented in table 8. For the full sample, and arid and humid regions separately, we show three different specifications. First, we show OLS with gridcell fixed effects (equation 3), clustering at the gridcell level, analogous to columns 1-3 in tables 6 and 7. Next, we replace gridcell fixed effects with country fixed effects, and use robust standard errors, similar to columns 7-9 in tables 5 and 6. Finally, we estimate equations (3) and (6), via 2SLS, following the same specification as columns 7-9 (country fixed effects). The reason for replacing gridcell fixed effects with country fixed effects in the 2SLS specification is two-fold. First, there is little variation over time in our $updam_{it}$ variable. As stated in section 2, the large majority of dams were built before our datasets begin. This would make gridcell fixed effects highly collinear with $updam_{it}$. The second reason deals with the fact that our instruments for $updam_{it}$ only vary over time at the country-level, and therefore our estimation strategy uses *within* country geographic variation to compare gridcells within countries. Nevertheless, our OLS regressions with gridcell fixed effects as well as regressions with country fixed effects, both including and excluding dams demonstrate that results are very similar, and so we do not expect this change to impact our 2SLS estimates.

OLS estimates in table 8 show that specifications with gridcell fixed effects, and those with country fixed effects are largely consistent and similar in pattern (with the sole exception of coefficients estimating the direct impact of dams). In OLS specifications with country fixed effects, the direct impact of upstream irrigation dams on NPP remains muted and insignificant. However, beneficial impacts emerge when regions downstream experience above or below normal rainfall shocks. Columns (2), (5) and (8) show that having a dam upstream decreases the adverse effects of a negative shock in the full sample as well as in arid and humid regions. In addition, upstream irrigation dams also tend to decrease the sensitivity of NPP to positive rainfall shocks in the full sample and in arid regions (that is, the coefficients on the interaction term of positive shocks and dam and the coefficient of the positive rain shock variable are of opposite sign). Overall, the OLS results suggest that dams smooth rainfall variability, and that areas with access to upstream irrigation dams may experience less severe effects of precipitation shocks on productivity.

The instrumental variable estimates, however, show different estimated effects of dams.[25] The direct impact of dams is now positive in all three samples, although it is only statistically significant in arid regions ($p<0.05$), which reap significant direct benefits in the form of boosted productivity. As with the OLS results, we see that when controlling for dams, positive shocks tend to increase NPP in the full, arid, and humid samples. The presence of dams significantly reduces the sensitivity of NPP to these positive shocks across all three samples, and this decrease in sensitivity is particularly striking in the arid regions. However, dams do not appear to reduce the sensitivity of NPP to negative rainfall shocks in the full sample and in humid regions. In contrast to the OLS estimates, we now find that dams amplify the adverse effects of below normal rainfall shocks in arid regions: both the coefficients on the interaction between negative rainfall shocks and dams, and on the negative rainfall shock variable are of the same sign, and are statistically significant. The IV estimates thus suggest that dams may make arid regions more vulnerable to dry shocks.

Why might this be so? We demonstrate that this result is driven by regions where the irrigation benefits of dams may have encouraged water-intensive cropping practices in otherwise unsuitable areas over time.[26] The implication is that the presence of irrigation may have encouraged maladaptation to dry conditions. We begin by illustrating this descriptively. Table 9 provides summary statistics for the harvested area of three water intensive crops—rice, cotton, and sugarcane—for two distinct samples of gridcells in arid regions: one sample includes all cropland gridcells with access to upstream irrigation dams, and the other includes all cropland gridcells without access to dams. It is striking that the mean values for rice harvest area, and sugarcane harvest area are substantially larger in areas with access to upstream irrigation dams (however, there is no discernable difference for cotton). This provides at least partial motivating

---

[25] In general, our 2SLS regressions pass key diagnostic checks, namely the Hansen J test for overidentifying restrictions (the p values are large), the Kleibergen-Papp LM test for instrument relevance and the Kleibergen-Papp Wald test for weak instruments (table 8) providing reassuring evidence about the validity of our instrumental variables.

[26] A similar logic is found in relation to groundwater. Hornbeck and Keskin (2014) find that sudden increased access to the Ogallala aquifer via improved pumps and center pivot irrigation technology increased drought sensitivity over time because farmers shifted to more water-intensive crops.

and anecdotal evidence showing that the presence of dams may have induced more water intensive forms of cultivation.

Taking our descriptive analysis one step further, we identify gridcells whose harvested area for any of the three water-intensive crops is greater than the global median. We remove these gridcells from the sample, and re-estimate equations (3) and (6) for the arid sample. Column (2) of table 10 presents these results and we see that, unlike in column (1) which includes the full arid sample, dams no longer exacerbate the negative impact of a below normal rainfall shock. The interaction effect between the negative rainfall shock and dam is positive but insignificant.[27]

Apart from cropping practices, it is also possible that during periods of below normal rainfall, stored water is prioritized for uses other than irrigation. This can be especially true for multipurpose dams, where lean season flows are diverted for hydropower generation which is often the more lucrative alternative (Strobl and Strobl, 2011). Recent work by Zeng et al. (2016) highlights these competing tradeoffs in multi-purpose dams and finds that as much as 54% of the global installed hydropower capacity competes with irrigation needs, and can significantly reduce water availability for crop production. Therefore, a variety of behavioral, technological, and management factors could be driving the results.

Finally, we also estimate the effects of dams by income group in columns (3), (4) and (5) in table 10. Dams directly increase agricultural productivity in middle-income and low-income countries. They also decrease the sensitivity of NPP to positive shocks in middle-income countries. When assessing their effects in the presence of dry shocks, we find that dams are unable to mitigate the negative effects of below normal rainfall shocks in middle and low-income countries, but in contrast, are able to buffer against these shocks in high-income countries. The latter result, though statistically significant, is imprecise given that we fail to pass instrumental variable

[27] The direct beneficial effect of dams, and its effect on positive shocks is lost as well. Therefore, it is possible that dams impact arid areas only in so far as those areas also grow water-intensive crops. Future drafts will explore the robustness of this result using different thresholds of water-intensity.

diagnostic checks.[28] Nevertheless, this result is suggestive of the importance of dam management practices in determining outcomes.

4.2.2 Influence of Irrigation Dams on the Relationship between Precipitation and Cropland Expansion

Next, we turn to the impact of dams on cropland expansion. We follow the same methodology as the agricultural productivity results, where we first estimate equation (5) via OLS, with gridcell fixed effects, and then switch to country fixed effects. Finally, we estimate equations (5) and (7) via 2SLS. Results are shown in table 11. Our OLS results show that there is little difference between the specifications with gridcell fixed effects, and those with country fixed effects. The direct impact of dams only appears significant for the full sample in column 2, where upstream irrigation dams tend to increase cropland expansion. As in the baseline results without dams, the effects of negative shocks are consistent and strong across all regions. Having four or more years with negative precipitation shocks tends to increase cropland expansion in the full, arid and humid samples. Likewise in the OLS regressions, having four or more years with positive precipitation shocks tends to increase cropland expansion in the full sample and in humid regions, but decreases it very slightly in arid regions. The presence of upstream irrigation dams mitigates the expansive impact of rainfall shocks in most specifications and samples.

Turning to the 2SLS results, we see a more nuanced result. The presence of upstream irrigation dams now tends to decrease cropland expansion in arid areas, but increases it in humid areas. This may be evidence of the fact that farmers in arid areas switch to more intensive agriculture when they gain access to irrigation from dam construction. We find that the impact from positive precipitation shocks become less significant, and small in magnitude. Conversely, negative precipitation shocks remain very significant, both statistically and in magnitude, with 4+ years of negative rainfall shocks leading to increases in cropland expansion in all regions, suggesting a much more robust impact than that of positive shocks. Further, we find that the presence of

[28] The p-value of the Hansen test of over-identifying restrictions is very low, and we fail to pass the test.

upstream irrigation dams mitigates this impact. In fact, even one upstream irrigation dam can entirely eliminate the expansion of cropland observed from multiple negative precipitation shocks.[29]

In section 5.3 we show these results are robust to replacing the 4+ year precipitation shock dummy variable with 2+, 3+, and 5+ year precipitation shock dummy variables.

4.3 Influence of Groundwater on Precipitation Impacts

In order to assess differential impacts of irrigation infrastructure, we also study the effects of groundwater. In general, surface water storage requires large amounts of capital while groundwater storage is free, although users face pumping costs and challenges associated with common-pool resources and the non-renewable nature of some groundwater systems. In addition, surface water availability is itself sensitive to rainfall variability, while deep groundwater reserves are less sensitive to climatic fluctuations.

To evaluate the effect of groundwater on the relationship between precipitation shocks and crop productivity, we identify gridcells with naturally occurring large groundwater endowments using prehistoric hydrogeological characteristics explained in section 2.3.2. Such characteristics capture long term geological potential, and do not reflect the current water table or fluctuations in the annual water depth within an aquifer. Therefore, such a measure of groundwater availability at the outset of the period is exogenous to land use decisions over time.

Table 12 shows that access to larger groundwater reserves reduces the adverse effect of a negative rainfall shock in the full and arid samples, highlighting the potentially larger buffering capability groundwater is able to provide. Since time-varying estimates of groundwater withdrawals and use are not available at a global scale, the $aquifer_i$ variable in equation (4) also includes gridcells where current reserves of groundwater are likely depleting. Our results

[29] Again note that in the first stage, the instruments pass the three important diagnostic tests; the Hansen J test fails to reject that overidentification restrictions are valid, the Kleibergen-Papp LM test rejects that excluded instruments are not correlated with the endogenous variables, and the Kleibergen-Papp Wald test shows that the instruments are relatively strong (particularly in the full sample regressions).

therefore reflect lower-bound estimates of the buffering capacity of groundwater[30]. In general, our results show that understanding effects of different types of irrigation can provide valuable insights into debates surrounding investments in water infrastructure for adapting to climate change.

5. Extensions and Robustness Checks

We perform several extensions and robustness checks to our main results. First, we test the impact on rainfall shocks on NPP across different precipitation quantiles. Second, we demonstrate how dams impact NPP at different distances other than the 25-50km threshold we use in our main specifications. Third, we show extensification and dam results using different number of shock-year combinations, different specifications, and time periods. Finally, we perform preliminary tests to check if spatial correlation may be biasing our results.

5.1 Precipitation Shocks, Agricultural Productivity, and Average Annual Rainfall

In this section, we test if the impact of precipitation shocks varies based on the long-run average rainfall of the region. To do so, we split the sample according to whether each gridcell's long-run average annual rainfall falls in a range of six quantile bins, Q1 to Q6. Q1 represents gridcells with lowest amounts of average rainfall, and Q6 represents those with the highest. Our estimation strategy follows our baseline model in equation (1). Consistent with our earlier results for arid and humid regions, we find that gridcells that receive the highest amounts of rainfall on average are also those least impacted by positive and negative precipitation shocks (table 13, figure 7). Figure 7 plots the coefficients on positive and negative precipitation shocks from each of the six regressions. We see that there is an incremental decrease in magnitude on both coefficients as we move from gridcells falling in the Q1 (driest) bin to those in the Q6 (wettest) bin. Similar to our baseline results in table 6 for the arid sample, we find that gridcells that receive the least amount of rainfall on average (those that fall in the Q1 and Q2 bins) are

[30] In addition to depletion, groundwater quality is an emerging concern. Accounting for this is beyond the scope of the paper. However, so far as locations with lower groundwater quality also lie in the same regions where groundwater is being depleted (as demonstrated in recent work by MacDonald el al., 2016), our estimates continue to provide a lower bound to the buffering capacity of groundwater.

adversely affected by below normal precipitation shocks and the impacts are large (worse) than those estimated for gridcells in the Q3-Q6 bins. For instance rainfall more than 1 standard deviation below the long- term mean leads to a nearly 10% and 11% decrease in agricultural productivity per year for gridcells in the Q1 and Q2 bins, respectively (as compared to decreases in the range of 0.3% to 8% for the other bins). Similarly, positive precipitation shocks also have the largest beneficial impact in these driest sextiles, implying that these regions are the most at risk to climate shocks.

5.2 The Impact of Dams on Agricultural Productivity at Varying Distances

Our empirical strategy focused on dams present between the 25-50 km threshold from the centroid of a grid cell. In table 14, we extend the analysis to other distance thresholds, namely 0-25 km, and 50-75 km. In general, we see similar patterns in the sign and significance of coefficients across the three distances for the full, arid, and humid samples. It is important to note that since we use 0.5 degree x 0.5 degree gridcells (approximately 50kms x 50kms at the equator), as depicted in figure 3, estimating the effect of upstream dams that lie in the 0-25km threshold is likely capturing the impacts in both the command and catchments areas, confounding the effects seen in command areas. In addition, as seen in columns (1), (4) and (7) of table 14 the model using a 0-25km distance threshold does not pass the IV diagnostic check for weak instruments as well as the model that uses the 25-50km threshold (the Kleibergen- Papp F statistics are lower). For the model using a threshold ranging from 50-75km, we find that the coefficients on the dam variables and its interaction with precipitation shocks are lower in magnitude, than those using our preferred threshold of 25-50km. Therefore, compared to all these distances, dams at 25-50km have the most credible and largest impact on agricultural productivity.

5.3 Cropland Expansion, Dams, and Precipitation Shocks at Different Shock-Year Thresholds

When examining the impact of dams on the relationship between precipitation shocks and cropland extensification, we use 4+ years as the cutoff for the number of years with precipitation shocks in our main results. This number was chosen because it offered a middle ground between showing a significant impact on extensification (in Figure 6), and having a large enough share of observations experiencing it (table 3). We now present the results from using different cutoffs for the number of years with precipitation shocks, namely 2+, 3+ and 5+. As in the baseline model with 4+years as the cutoff, we construct dummy variables which are equal to 1 if there are 2+, 3+, or 5+ years with 1+ SD precipitation shocks in the past decade.

Results are shown in table 15. The full sample is shown in columns 1-4, with columns 1, 2, 3 and 4 showing results for 2+, 3+, 4+, and 5+ years of shocks, respectively. Similarly, columns 5-8 and 9-12 show results for the arid and humid samples, respectively. Generally speaking, the results hold for all 4 different thresholds. Dams reduce cropland expansion in arid areas, and increase it in humid areas, although the coefficient is insignificant with the 2+ year threshold in the latter region. The impact of positive rainfall shocks remains weak and noisy, but negative rainfall shocks are shown to robustly increase extensification, with dams mitigating that impact. It is also interesting to note that, consistent with figure 6, the impact of rainfall shock increases with the cutoff, as does the mitigating impact of dams. This confirms that our results shown in section 4.2.1 are not simply a product of selecting a particular cutoff for the number of shocks.

5.4 Cropland Expansion, Dams, and Precipitation Shocks; Different Specifications and Time Periods

The results on precipitation shocks and cropland expansion involve a data sample that is slightly restricted. This is because the cropland dataset is available over the period 1980-2005, but the population data is only available from 1990. By including population in the regression, we therefore lose two time periods (since the dataset is observed only every five years). In this section, we present results where log(population) is excluded from the regression so the sample period can be extended. To do so, we estimate equation 2 using three different time periods and specifications: 1) time period 1990-2005, with log(population) included (the original results); 2) time period 1990-2005 without log(population) included; and 3) time period 1980-2005 without log(population included).

Results which look strictly at the impact of precipitation shocks on extensification (without dams) are presented in table 16, and those that include dams are presented in table 17. Columns 1-3 focus on the full sample. Column 1 includes log(population) with time period from 1990-2005 like in the baseline results. Column 2 excludes log(population) with time period from 1990-

2005, and column 3 excludes log(population) with the extended time period from 1980-2005. Columns 4-6 and 7-9 are analogous for the arid and humid samples, respectively. The sign and statistical significance of the coefficients does not change from one specification to the other, and the magnitudes change only slightly (and the change is statistically insignificant), showing that the results are robust to changes in specification and time period.

5.5 Testing for Spatial Auto-Correlation

In this final sub-section, we perform preliminary tests for potential bias arising from spatial autocorrelation in our globally gridded datasets. In order to do so, we re-estimate equations (1) and (2), but adjust standard errors for possible serial and spatial correlation by clustering errors ε_{it} from all observations (across gridcells and years) at four different geographic levels: country, country-year, province, and province-year. [31] Clustering at a larger geographic level has generally been found to provide equivalent results to directly accounting for spatial autocorrelation[32].

Regressions results for the NPP model are reported in table 18. Column 1 is our baseline model with robust standard errors, and columns 2-5 cluster standard errors at the country, country-year, province, and province-year level, respectively. The coefficients on positive and negative precipitation shocks remain significant ($p<0.001$) across different levels of clustering. Similarly, table 19 reports results for the cropland regressions. Like in the baseline model shown in column 1, the coefficient on positive precipitation shocks remains insignificant for all levels of clustering. The coefficient on negative precipitation shocks is significant across different levels of clustering except for the most restrictive case in column 2 when standard errors are clustered at the country level.

[31] Here we define 'province' as the largest sub-national political unit of the country, according to *GADM database of Global Administrative Areas* (http://gadm.org/). Gridcells are attributed to the administrative unit which their centroid falls in.
[32] Fisher et al. (2012) find that clustering at the state level in the U.S gives equivalent results to directly accounting for spatial correlation using the Conley (1999) corrected standard errors.

While these tests are not conclusive, they do give us confidence that spatial autocorrelation, if present, is not significantly biasing the results.

6. Conclusions

As global climate change is expected to increase rainfall variability in the future, our results show that this will lead to increased stresses on food production and security, and pressures for land use-change. Our analysis on dams also provides a cautionary tale of the potential for mal-adaptation, where important, well-intended infrastructure can be damaging if the consequences are not properly managed.

6.1 Discussion

In this paper, we document a robust global impact of precipitation shocks on agricultural production across both the intensive and extensive margins. We find that across several climactic classifications and all income groups, positive precipitation shocks are associated with increases in agricultural productivity, as measured by net primary productivity, and negative precipitation shocks are associated with reductions in agricultural productivity. Further, using instrumental variables to treat the endogenous placement of dams, we find that areas with dams show improvements in NPP. However, when faced with negative precipitation shocks, regions with dams, particularly those in arid regions and developing countries, experience compounding harm. A variety of factors related to the behavior of end-users of water, and dam operation could influence such a negative effect on agricultural productivity. We provide suggestive evidence for likely mal-adaptation in arid regions, resulting in the planting of water-intensive crops. In addition, we show that groundwater irrigation, especially in areas with deeper reserves, might have a larger buffer capacity in mediating negative precipitation shocks.

We also find that rainfall variability impacts agricultural extensification, with repeated, negative precipitation shocks leading to an increase in cultivated land. The presence of an upstream irrigation dam, however, counteracts that impact. Although these regressions are reduced form and do not allow us to fully understand the mechanism behind this impact, it is consistent with extensification being used as an adaptation technique to buffer against reduced yields.

Some of these results, however, may appear to contradict each other. If negative shocks lower agricultural productivity, and in turn reduce farm incomes, ceteris paribus, then both expected utility theory and loss aversion theory would predict that risk averse farmers lower their exposure to rainfall risks by *not* expanding cropland, especially if risks are higher on rainfed farms. This contradicts our empirical results where we find that farmers respond to negative precipitation shocks by expanding cropland. Moreover, we find that in the presence of dams, this result is reversed: even though the adverse effects of negative precipitation shocks on agricultural productivity are accentuated by dams there is a contraction in cropland in the presence of a dam.

We argue that these seemingly internally inconsistent results can be explained and reconciled by a variant of the safety-first model (Roy 1952). Versions of the safety-first model have been widely used to explain responses to risk in both the finance and agriculture literatures (Moscardi and de Janvry, 1977; Shahabuddin, Mestelman and Feeny, 1986; Alderman and Paxson, 1994; Wale and Yalew, 2007; Sattinger, 2013). In the context of agriculture such behavior would be consistent with farmers seeking a certain level of income, or policies of food self-sufficiency that promote target output levels on farms. We demonstrate this effect using a simple model.

Consider a farmer with a fixed endowment of irrigated land termed *A* that is located downstream of a dam. Let ε be the available unirrigated land where extensification may occur. Stochastic rainfall on the irrigated and unirrigated plots are denoted θ_i and θ_u respectively. Let L_i, L_u be farm inputs such as say fertilizer, or labor. Payoffs from production are simply defined as:

$y_i = \theta_i f(A, L_i) - C(L_i)$ and $y_u = \theta_u g(\varepsilon, L_u) - C(L_u, \varepsilon)$;

where *f* and *g* are strictly concave and increasing in their arguments and *C* is increasing in its arguments and strictly convex. Not unreasonably assume that expected yields on irrigated land is higher than on rainfed land so that $E(y_i/A) > E(y_u/\varepsilon)$. Define expected payoffs of cultivating both plots:

$$\mu = E(\theta_i)Q_i + E(\theta_u)Q_u - C(L_i) - C(L_u, \varepsilon) \quad (9)$$

The variance of payoffs can be shown to be given by the formula:

$$\sigma^2 = \sigma_i^2 Q_i^2 + \sigma_u^2 Q_u^2 + 2\sigma_{iu} Q_i Q_u \quad (10)$$

Where $Q_i = f(A, L_i)$ and $Q_u = g(\varepsilon, L_u)$, and $\sigma_v^2 = E(v^2) - (E(v))^2$ for v = y_i, or y_u.

The Roy safety-first rule hypothesizes that agents seek to minimize the probability α that their expected income (defined in I) falls below some threshold level d. That is, Min α = Prob(μ < d). Then by the Chebychev Inequality this can be recast as the following problem:

$$\min_{L_i, L_u, \epsilon} Z = \frac{d-\mu}{\sigma^2}. \quad (11)$$

Substituting (9) and (10) into (11) yields the following first-order-conditions:

$$\frac{\partial\mu}{\partial\varepsilon} + Z\frac{\partial\sigma^2}{\partial\varepsilon} = 0; \quad \text{(I2a)}$$

$$\frac{\partial\mu}{\partial L_i} + Z\frac{\partial\sigma^2}{\partial L_i} = 0; \quad \text{(I2b)}$$

$$\frac{\partial\mu}{\partial L_u} + Z\frac{\partial\sigma^2}{\partial L_u} = 0. \quad \text{(I2c)}$$

Consider the case of a farmer on unirrigated land (L_i = 0) and assume that increasing planted area increases exposure to rainfall risk $\frac{\partial\sigma^2}{\partial\varepsilon} > 0$. Let mean income be such that $d > \mu$. Then in (11) we have Z >0. It follows that the second term in (12a) is positive and this implies that increases in variance (variability of rainfall) will increase the optimum planting area. Thus for farmers with μ < d higher risks induce a gamble in an attempt to achieve a threshold level of income.

Consider next the case of farmers with irrigated land. Recall from table 8 that the presence of a dam raises average incomes all else equal, which is reflected in our assumption that E$(y_i/A) > E(y_u/\varepsilon)$. The likelihood of such farmers having incomes above the critical threshold is therefore greater. Hence for sufficiently productive irrigated agriculture there will be some farms where μ > d and hence Z < 0. In this case, we therefore have the second term in (12a) which is negative and this implies that as exposure to risk increases (i.e. all else equal σ^2 rises) then ε (unirrigated planted area) declines.

In sum, in this model, responses to risk are given by the sign of Z and as the presence of a dam makes farmers richer they are more likely to exhibit standard risk averse behavior. But in cases where incomes or output levels are below desired thresholds, there is an incentive to increase risk exposure in order to meet the minimum desired output or income level.

6.2 Caveats and Future Work

Although we have endeavored to use the best possible methods and data, we acknowledge several shortcomings.

Firstly, pertaining to our data, we acknowledge that while no dataset is perfect, global gridded data will always have some level of inaccuracy. There are many weather data products, but few offer as wide coverage, both over time and over space, as the Matsuura and Willmott (2001) dataset which we employ here. Remote sensing technology today allows for highly accurate observations of temperature and rainfall. However, this data is not available for much of the necessary time period, nor is it available yet at a global scale. Although it is imperfect, Matsuura and Willmott (2001) is widely used in the economics literature and deemed to be of sufficient accuracy to make statistical inferences from (e.g. Bohra-Mishra, Oppenheimer, and Hsiang 2013; Cole, Healy, and Werker 2012; Dell, Jones, and Olken 2012; Nunn and Qian 2014). Likewise, the cropland dataset which we employ is an approximation of the real-world, and is one of many different approaches used to construct gridded data sets of cropland. Nevertheless, it represents the state-of-the-art for this type of product, and has been widely used in many integrated assessment and land-use change models. Additionally, the global land cover dataset we use to identify cropland for the NPP calculations is of coarse resolution. Therefore, it is likely that our NPP measures only capture changes in agricultural productivity on larger cropland areas and do not account for small-holder farms, unless these are agglomerated. We plan to test the robustness of our results by using other available land cover datasets in a future draft.

When estimating the impact of rainfall variability, we focus on one specific type of variability: variation from year to year. As Fishman (2016) shows, intra-seasonal rainfall variability can have an even larger impact on crop yields. Estimating the impact of intra-seasonal variability is relatively straightforward when dealings with a small geographic area, where growing seasons are homogenous or well-known. At the global level, however, measuring this impact is much more difficult as it would require very detailed knowledge on crop choices (which will be endogenously determined with respect to rainfall variation, adding an additional layer of complexity). Because we examine annual variation, rather than growing season-variation, it is likely that we are underestimating the true impact of increased rainfall variability during the growing season.

In a similar vein, we treat all irrigation dams homogenously, which overlooks some important distinguishing characteristics. All dams included in the GRanD dataset are considered to be ‘large’ dams with a minimum reservoir capacity, making them in some respects, similar. Nevertheless, we cannot account for idiosyncrasies of each dam, including dam managing institutions or physical dam characteristics. Our dam results therefore should not be interpreted as being representative of any particular dam, but should instead be considered indicative of the global large dam population as a whole. In future drafts, an analysis of irrigation only dams versus multi-purpose dams will be conducted.

Finally, several of the datasets which we use, including NPP, cropland, and weather data, likely exhibit some level of spatial autocorrelation. When not accounted for, spatial autocorrelation has the potential to bias regression estimates. We perform one test for spatial correlation by clustering our standard errors at several different geographic levels, and find that the results are generally robust. We therefore do not believe such a bias is significant in generating our results, but do plan on further addressing this issue in future versions.

References

Abdi, A. M., Seaquist, J., Tenenbaum, D. E., Eklundh, L., & Ardö, J. (2014). The supply and demand of net primary production in the Sahel. *Environmental Research Letters*, *9*(9), 094003.

Alderman, H. and Paxson C.H. (1994). Do the poor insure? A synthesis of the literature on risk and consumption in developing countries. *Economics in a Changing World: Proceedings of the Tenth World Congress of the International Economic Association*. St. Martin's Press, New York, pp. 48–78

Auffhammer M., S.M. Hsiang, W. Schlenker and A. Sobel (2013) Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy* 7 (2): 181-198.

Auffhammer, M., & Schlenker, W. (2014). Empirical studies on agricultural impacts and adaptation. *Energy Economics*, *46*, 555-561.

Baccini, A. G. S. J., et al. "Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps." *Nature Climate Change* 2.3 (2012): 182-185.

Blanc, E., & Strobl, E. (2013). The impact of climate change on cropland productivity: evidence from satellite based products at the river basin scale in Africa. *Climatic change*, *117*(4), 873-890.

Blanc, E., & Strobl, E. (2014). Is small better? A comparison of the effect of large and small dams on cropland productivity in South Africa. *The World Bank Economic Review*, *28*(3), 545-576.

Bohra-Mishra, Pratikshya, Michael Oppenheimer, and Solomon M. Hsiang. "Nonlinear permanent migration response to climatic variations but minimal response to disasters." *Proceedings of the National Academy of Sciences*111.27 (2014): 9780-9785.

Brower, C., & Heibloem, M. (1986). Irrigation Water Management Training Manual No. 3: Irrigation Water Management: Irrigation Water Needs. FAO, International Support Program for Irrigation Water Management. *Land and Water Development Division, Rome*.

Burke, M., Hsiang, S. M., & Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*.

Cech, T. V. (2010). Principles of water resources: history, development, management, and policy. John Wiley & Sons.

Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4SF2T42.

Cole, Shawn, Andrew Healy, and Eric Werker. "Do voters demand responsive governments? Evidence from Indian disaster relief." *Journal of Development Economics* 97.2 (2012): 167-181.

Conley, T. (1999) GMM Estimation with Cross Sectional Dependence, *Journal of Econometrics*, 92, 1-45.

Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, *4*(3), 66-95.

Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate–economy literature. *Journal of Economic Literature*, *52*(3), 740-798.

Diamond, P. A., & Stiglitz, J. E. (1974). Increases in risk and in risk aversion. *Journal of Economic Theory*, *8*(3), 337-360.

Duflo, E., & Pande, R. (2005). *Dams* (No. w11711). National Bureau of Economic Research.

Duflo, E. and R. Pande, 2007," Dams," *Quarterly Journal of Economics*.

Fisher, A., Hanemann, M., Roberts, M. and Schlenker, W. (2012) The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment. *American Economic Review*, 102, 3749-60.

Fishman, R. (2016). More uneven distributions overturn benefits of higher precipitation for crop yields. *Environmental Research Letters*, *11*(2), 024004.

Food and Agriculture Organization of the United Nations. FAO GEONETWORK. World map of the major hydrological basins (Derived from HydroSHEDS) (GeoLayer). (Latest update: 04 Jun 2015) Accessed (28 Dec 2016). URI: http://data.fao.org/ref/7707086d-af3c-41cc-8aa5-323d8609b2d1.html?version=1.0

GEM (2011) Global Land Cover 2000 Data Set (Glc 2000). Global Environment Monitoring Unit (GEM)

Heinsch, F. A., Jolly, W. M., Mu, Q., & Kimball, J. S. (2005, May). Regional scaling of cropland net primary production for Nebraska using satellite remote sensing from MODIS and the ecosystem process model Biome-BGC. In *AGU Spring Meeting Abstracts*..

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., & Sohlberg, R. A. (2003). Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, *7*(10), 1-15.

Hicke, J. A., Lobell, D. B., & Asner, G. P. (2004). Cropland area and net primary production computed from 30 years of USDA agricultural harvest data. *Earth Interactions*, *8*(10), 1-20.

Hornbeck, R., & Keskin, P. (2014). The historically evolving impact of the Ogallala aquifer: Agricultural adaptation to groundwater and drought. *American Economic Journal: Applied Economics*, *6*(1), 190-219.

Houghton R A. (2008). Carbon flux to the atmosphere from land use changes: 1850–2005. In: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, TN, USA.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., ... & Wang, Y. P. (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Climatic Change, 109(1-2), 117-161.

Kautilya, and L. N. Rangarajan. "IV.v Aspects of the Economy: Agriculture." The Arthashastra. New Delhi: Penguin India, 1992.

Klein Goldewijk, K., Beusen, A., Van Drecht, G., & De Vos, M. (2011). The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. Global Ecology and Biogeography, 20(1), 73-86.

Koide, N A.W. Robertson, A.V.M. Ines, J.-H. Qian, D.G. DeWitt, A. Lucero 2013 Prediction of rice production in the Philippines using seasonal climate forecasts; J. Appl. Meteorol. Climatol., 52, 552–569

Lehner, B., C. Reidy Liermann, C. Revenga, C. Vorosmarty, B. Fekete, P. Crouzet, P. Doll, M. Endejan, K. Frenken, J. Magome, C. Nilsson, J.C. Robertson, R. Rodel, N. Sindorf, and D. Wisser. 2011a. Global Reservoir and Dam Database, Version 1 (GRanDv1): Dams, Revision 01.

Lehner, B., C. Reidy Liermann, C. Revenga, C. Vörösmarty, B. Fekete, P. Crouzet, P. Döll, M. Endejan, K. Frenken, J. Magome, C. Nilsson, J.C. Robertson, R. Rodel, N. Sindorf, and D. Wisser. 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.

Lobell, D. B., Hicke, J. A., Asner, G. P., Field, C. B., Tucker, C. J., & Los, S. O. (2002). Satellite estimates of productivity and light use efficiency in United States agriculture, 1982–98. *Global Change Biology*, *8*(8), 722-735.

Lu, X., & Zhuang, Q. (2010). Evaluating climate impacts on carbon balance of the terrestrial ecosystems in the Midwest of the United States with a process-based ecosystem model. *Mitigation and adaptation strategies for global change*, *15*(5), 467-487

MacDonald, A. M., Bonsor, H. C., Ahmed, K. M., Burgess, W. G., Basharat, M., Calow, R. C., et al. (2016). Groundwater quality and depletion in the Indo-Gangetic Basin mapped from in situ observations. *Nature Geoscience*, *9*(10), 762-766.

Moscardi, E. and A. de Janvry. (1977). Attitudes toward Risk among Peasants: An Econometric Approach. *American Journal of Agricultural Economics,* 59(4):710-716

Monfreda, C., Ramankutty, N., & Foley, J. A. (2008). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global biogeochemical cycles*, *22*(1).

Mendelsohn, R., Nordhaus, W. D., & Shaw, D. (1994). The impact of global warming on agriculture: a Ricardian analysis. *The American economic review*, 753-771.

Narain, U., S. Margulis, and T. Essam (2011) Estimating costs of adapting to climate change, *Clim. Policy,* 11: 1001–1019

Naylor, RL W.P. Falcon, D. Rochberg, N. Wada (2001) Using El Niño/Southern Oscillation climate data to predict rice production in Indonesia; Clim. Change, 50, 255–265

Nunn, Nathan, and Nancy Qian. "US food aid and civil conflict." *The American Economic Review* 104.6 (2014): 1630-1666.

Olmstead, Sheila M., and Hilary Sigman 2016 Droughts, dams, and economic activity, Mimeo unpublished paper presented at AEA conference

Olmstead, Sheila M., and Hilary Sigman. 2015 "Damming the commons: An empirical analysis of international cooperation and conflict in dam location." *Journal of the Association of Environmental and Resource Economists* 2.4 (2015): 497-526.

Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4N877QK.

Prince, S. D., Haskett, J., Steininger, M., Strand, H., & Wright, R. (2001). Net primary production of US Midwest croplands from agricultural harvest yield data. *Ecological Applications*, *11*(4), 1194-1205.

Ramankutty, N., & Foley, J. A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global biogeochemical cycles*, *13*(4), 997-1027.

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, *22*(1).

Roy, A. D. (1952). Safety-first and the Holding of Assets. *Econometrica*, 20(3):431–449.

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *Bioscience*, *54*(6), 547-560.

Sattinger, M. (2013). Safety first consumption. *Journal of Economic Theory*, *148*(1), 306-321.

Sawano, S T., Hasegawa, S. Goto, P. Konghakote, A. Polthanee, Y. Ishigooka, T. Kuwagata, H. Toritani (2008) Modeling the dependence of the crop calendar for rain-fed rice on precipitation in Northeast Thailand; Paddy Water Environ, 6, 83–90

Schlenker, W., W. M. Hanemann, and A. C. Fisher (2006) The impact of global warming on US agriculture: an econometric analysis of optimal growing conditions, *Review of Economics and Statistics*, 88, 113-125.

Schlenker, W. and M. J. Roberts (2009) Nonlinear temperature e_ects indicate severe damages to US crop yields under climate change, Proceedings of the National Academy of Sciences, 106, 15594-15598.

Shahabuddin, Q., Mestelman, S. and Feeny, D. (1986). Peasant Behavior Toward Risk and SocioEconomic and Structural Characteristics of Farm Household in Bangladesh. Oxford Economic Papers 38:122-130

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., & Scanlon, B. R. (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, R. O. (2011). The distributional impact of large dams: Evidence from cropland productivity in Africa. *Journal of Development Economics*, *96*(2), 432-450.

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, *108*(50), 20260-20264.

Tum, M., & Günther, K. P. (2011). Validating modelled NPP using statistical yield data. *biomass and bioenergy*, *35*(11), 4665-4674.

Turner, D. P., Ritts, W. D., Cohen, W. B., Gower, S. T., Running, S. W., Zhao, M., et al (2006). Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sensing of Environment*, *102*(3), 282-292.

Wale, E. and Yalew, A. (2007). Farmers' Variety Attribute Preferences: Implications for Breeding Priority Setting and Agricultural Extension Policy in Ethiopia. *African Development Review*, 19(2):379-396.

Willmott, C. J. and K. Matsuura (2001) Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1900 - 2014), http://climate.geog.udel.edu/~climate/html_pages/download.html

You, Liangzhi, et al. "What is the irrigation potential for Africa? A combined biophysical and socioeconomic approach." Food Policy 36.6 (2011): 770-782.

Zeng, R., Cai, X., Ringler, C & Zhu, T. (2016). Hydropower versus Irrigation-An Analysis of Global Patterns.

Zhang, Y., Yu, Q., Jiang, J. I. E., & Tang, Y. (2008). Calibration of Terra/MODIS gross primary production over an irrigated cropland on the North China Plain and an alpine meadow on the Tibetan Plateau. *Global Change Biology*, *14*(4), 757-767.

Zhao, M., Heinsch, F. A., Nemani, R. R., & Running, S. W. (2005). Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote sensing of Environment*, *95*(2), 164-176.

Tables

Table 1: Crop Water Requirements

Crop	Water Need (mm/growing season)	Sensitivity to Drought
Barley	450-650	low-medium
Cotton	700-1300	low
Maize	500-800	medium-high
Millet	450-650	low
Peanut	500-700	low-medium
Potato	500-700	high
Pulses	350-500	medium-high
Rice	450-700	high
Sorghum	450-650	low
Soybean	450-700	low-medium
Sugarcane	1500-2500	high
Sunflower	600-1000	low-medium
Wheat	450-650	low-medium

Source: Brouwer and Heibloem (1986).

Table 2: Summary Statistics: Intensive Margin

Variable	Observations	Mean	Std. Dev.	Min	Max
Full Sample					
Δ Log NPP	121,823	0.0020315	0.1555586	-2.605602	2.395623
Positive 1 SD precipitation shock	131,194	0.1539705	0.3609218	0	1
Negative1 SD precipitation shock	131,194	0.1456317	0.3527379	0	1
Upstream irrigation dams	131,194	0.05921	0.3269501	0	6
Aquifer	131,194	0.6723936	0.4693422	0	1
Arid Regions					
Δ Log NPP	56,043	0.004265	0.1821419	-1.875797	1.943092
Positive 1 SD precipitation shock	60,354	0.1538423	0.3608005	0	1
Negative1 SD precipitation shock	60,354	0.1497332	0.3568127	0	1
Upstream irrigation dams	60,354	0.0623654	0.320856	0	6
Aquifer	60,354	0.6782649	0.4671458	0	1
Humid Regions					
Δ Log NPP	65,247	0.0001323	0.129081	-2.605602	2.395623
Positive 1 SD precipitation shock	70,266	0.1538013	0.3607607	0	1
Negative1 SD precipitation shock	70,266	0.1422452	0.3493039	0	1
Upstream irrigation dams	70,266	0.0555888	0.3296766	0	6
Aquifer	70,266	0.6672644	0.4711961	0	1

Table 3: Distribution of the Frequency of Shocks

Years	# of positive shocks	# of negative shocks
0	92,489	112,736
1	99,879	105,718
2	78,990	69,491
3	47,087	35,116
4	19,993	14,862
5	6,966	6,134
6	1,841	2,710
7	505	906
8	162	292
9	170	86
10	8	39

Table 4: Summary Statistics: Extensive Margin

Variable	Observations	Mean	Std. Dev.	Min	Max
Full Sample					
Δ Log Cropland	228,111	0.001084	0.0439956	-0.5380473	3.125128
Number of Positive 1sd+ shocks in 10 years	228,111	1.462832	1.322833	0	10
Number of Negative 1sd+ shocks in 10 years	228,111	1.406219	1.410588	0	10
log(Population)	228,111	8.034819	3.731299	-14.13268	16.69055
Contemporaneous Annual Rainfall (mm)	228,111	697.2635	685.6145	0	10884.5
Contemporaneous Mean temperature (C)	228,111	10.52946	13.52876	-26.35	37.55833
Upstream irrigation dams	228,111	0.0218096	0.2009196	0	13
Arid Regions					
Δ Log Cropland	161,317	-0.0005896	0.0192585	-0.5380473	1.151934
Number of Positive 1sd+ shocks in 10 years	161,317	1.491665	1.35288	0	10
Number of Negative 1sd+ shocks in 10 years	161,317	1.256799	1.342835	0	10
log(Population)	161,317	6.977274	3.721091	-14.13268	16.57498
Contemporaneous Annual Rainfall (mm)	161,317	391.6998	305.1831	0	6179.9
Contemporaneous Mean temperature (C)	161,317	6.982183	13.81474	-26.35	33.54167
Upstream irrigation dams	161,317	0.0180762	0.1842319	0	13
Humid Regions					
Δ Log Cropland	65,322	0.0052118	0.0762545	-0.5332266	3.125128
Number of Positive 1sd+ shocks in 10 years	65,322	1.393696	1.241004	0	8
Number of Negative 1sd+ shocks in 10 years	65,322	1.77366	1.500328	0	10
log(Population)	65,322	10.67981	2.012343	-2.246583	16.69055
Contemporaneous Annual Rainfall (mm)	65,322	1445.438	770.4847	39.8	10884.5
Contemporaneous Mean temperature (C)	65,322	19.36522	7.341205	-8.9	37.55833
Upstream irrigation dams	65,322	0.030939	0.2367886	0	6

Table 5: Geography and Upstream irrigation dams

	(1)	(2)	(3)	(4)
		Full sample		
	2000-2013	2000-2013	1990-2005	1990-2005
River length * $\overline{D}_{ct}$	0.000110***	0.000104***	0.000007***	0.000014***
	(0.000)	(0.000)	(0.000)	(0.000)
River slope share *$\overline{D}_{ct}$	0.000046*	0.000066***	0.000023**	0.000054***
	(0.000)	(0.000)	(0.000)	(0.000)
Positive 1SD rainfall shock		0.001043		
		(0.003)		
Negative 1SD rainfall shock		0.005230+		
		(0.003)		
Average Temp(C)		0.004774***		
		(0.000)		
Average Temp(C) Sq.		0.000089***		
		(0.000)		
4+ Positive 1SD rainfall shocks				-0.001861
				(0.001)
4+ Negative 1SD rainfall shocks				-0.001002
				(0.002)
Contemporaneous Annual Rainfall (mm/year)				-0.000004**
				(0.000)
Contemporaneous Mean temperature (C)				0.000850***
				(0.000)
Log(Population)		0.017172***		0.005399***
		(0.001)		(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	49.411	44.263	13.26	23.10
N	125258	125258	347034	227801
R-sq	0.092	0.103	0.054	0.063

Notes: Dependent variable is the number of upstream irrigation 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

Table 6: Impact of Precipitation Shocks on Crop Productivity

	Dependent variable: Δ Log NPP								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	With Cell Fixed Effects						With Country Fixed Effects		
	Full sample	Arid	Humid	High-income	Middle-income	Low-income	Full sample	Arid	Humid
Positive 1SD precip shock	0.050768***	0.076959***	0.027882***	0.031694***	0.055773***	0.014575*	0.050461***	0.074372***	0.030541***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.001)	(0.002)	(0.001)
Negative 1SD precip shock	-0.067176***	-0.094896***	-0.036016***	-0.069800***	-0.070494***	-0.020810***	-0.067026***	-0.093976***	-0.037963***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)	(0.001)	(0.002)	(0.002)
Avg Temp(C)	-0.034625***	-0.027589***	-0.052127***	-0.005782	-0.039311***	0.012226	-0.001827***	-0.002372***	-0.000916
	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)	(0.009)	(0.000)	(0.000)	(0.001)
Avg Temp(C) Sq.	-0.000071	-0.000682***	0.000577***	-0.001236***	0.000439***	-0.001541***	0.000022*	-0.000011	0.000007
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log(Population)	0.000557	0.025868**	-0.004066	-0.024360*	0.009467	-0.039122+	-0.000194	0.000592	-0.001237**
	(0.006)	(0.010)	(0.007)	(0.012)	(0.006)	(0.022)	(0.000)	(0.001)	(0.000)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cell fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Country Fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Country specific trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	121823	56043	65247	27742	80353	11128	121823	56043	65247
R-sq	0.110	0.161	0.100	0.170	0.128	0.129	0.086	0.132	0.079

Notes: Dependent variable is the change in log NPP in a gridcell. Each column presents the regression coefficients from a separate regression based on gridcells in arid and humid climatic zones as per GAEZ FAO classification, and in different income categories as per the World Bank classification Standard errors are reported in parentheses. Columns 1-6 show standard errors clustered at the gridcell level, while columns 7-9 show cluster-robust standard errors. Positive (Negative) 1 SD rainfall shock 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. Statistical significance is given by + $p<0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table 7: Impact of Precipitation Shocks on Cropland Expansion

	Dependent Variable: Δ Log Cropland								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	With Gridcell Fixed Effects						With Country Fixed Effects		
	Full Sample	Arid Regions	Humid Regions	High Income	Middle income	Low Income	Full Sample	Arid Regions	Humid Regions
Number of Positive 1sd+	0.000670***	-0.000174*	0.001988***	0.000026	0.001275***	-0.000525***	-0.000095	-0.000222***	-0.000298
shocks in 10 years	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of Negative 1sd+	0.000644***	0.000576***	0.001153***	-0.000004	0.000556***	0.001528***	0.000461***	0.000466***	0.000531***
shocks in 10 years	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Contemporaneous Annual	-0.000000	0.000002**	-0.000000	0.000005***	-0.000002	-0.000007***	0.000002***	0.000002***	-0.000000
Rainfall (mm)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Contemporaneous Mean	0.000467	-0.000191	0.008293	0.000621***	-0.001081***	-0.001711***	-0.000013	-0.000098***	0.000428***
temperature (C)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
log(Population)	0.000534	-0.000376	-0.009218	-0.001558***	-0.001383	0.002983+	-0.000727***	-0.000311***	-0.002216***
	(0.000)	(0.000)	(0.006)	(0.000)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gridcell Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Country Fixed Effects	No	No	No	No	No	No	Yes	Yes	Yes
Country Specific Time Trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	228111	161317	65322	66731	138052	18388	228111	161317	65322
R-sq	0.299	0.389	0.290	0.371	0.287	0.562	0.082	0.195	0.070

Notes: Dependent variable is the change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression based on gridcells in arid and humid climatic zones as per GAEZ FAO classification and income groups according to World Bank classification. Standard errors are reported in parentheses. Columns 1-6 show standard errors clustered at the gridcell level, while columns 7-9 show cluster-robust standard errors. Number of Positive (Negative) shocks 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 8: Joint Impact of Precipitation Shocks and Dams on Crop Productivity

	Δ Log NPP								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Full sample			Arid Regions			Humid Regions		
	OLS	OLS	2SLS	OLS	OLS	2SLS	OLS	OLS	2SLS
Upstream irrigation dams	-0.0285***	0.0019	0.0357	-0.0346***	0.0037	0.2044*	0.0000	0.0000	0.0005
	(0.005)	(0.001)	(0.030)	(0.008)	(0.003)	(0.096)	(.)	(0.001)	(0.021)
Pos 1SD rainfall shock	0.0514***	0.0510***	0.0991***	0.0781***	0.0752***	0.1549***	0.0282***	0.0308***	0.0501***
	(0.001)	(0.001)	(0.011)	(0.003)	(0.002)	(0.028)	(0.002)	(0.001)	(0.009)
x Upstream irrigation dams	-0.0107***	-0.0091**	-0.7812***	-0.0204**	-0.0151*	-1.3145**	-0.0044	-0.0041	-0.3139*
	(0.003)	(0.003)	(0.178)	(0.007)	(0.006)	(0.453)	(0.003)	(0.003)	(0.145)
Neg 1SD rainfall shock	-0.0676***	-0.0674***	-0.0706***	-0.0958***	-0.0947***	-0.0799***	-0.0364***	-0.0384***	-0.0444***
	(0.002)	(0.001)	(0.004)	(0.003)	(0.002)	(0.006)	(0.002)	(0.002)	(0.005)
x Upstream irrigation dams	0.0066	0.0065*	0.0319	0.0137+	0.0099+	-0.2298**	0.0061	0.0071*	0.0867
	(0.004)	(0.003)	(0.056)	(0.007)	(0.006)	(0.087)	(0.004)	(0.003)	(0.077)
Avg Temp(C)	-0.0346***	-0.0019***	-0.0009*	-0.0275***	-0.0024***	-0.0016**	-0.0521***	-0.0009	-0.0003
	(0.002)	(0.000)	(0.000)	(0.002)	(0.000)	(0.001)	(0.003)	(0.001)	(0.001)
Avg Temp(C) Sq.	-0.0001	0.0000*	0.0000	-0.0007***	-0.0000	-0.0000	0.0006***	0.0000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log(Population)	0.0006	-0.0003	0.0003	0.0260**	0.0004	0.0008	-0.0041	-0.0014**	-0.0015**
	(0.006)	(0.000)	(0.001)	(0.010)	(0.001)	(0.003)	(0.007)	(0.000)	(0.001)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cell fixed effects	Yes	No	No	Yes	No	No	Yes	No	No
Country fixed effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Country specific trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	121823	121823	116311	56043	56043	52494	65247	65247	63388
R-sq	0.110	0.086		0.161	0.132		0.100	0.079	
First stage:									
Dependent Variable: Upstream irrigation dams									
River Slope Share 1.5-3% * $\overline{D}_{ct}$			0.0000622*			0.0002210***			-0.0000641*
			(0.000026)			(0.000045)			(0.000030)
River Slope Share 1.5-3% * Neg 1SD shock * $\overline{D}_{ct}$			0.0001367*			0.0003435**			-0.0000840
			(0.000070)			(0.000122)			(0.000080)
River Slope Share 1.5-3% * Pos 1SD shock * $\overline{D}_{ct}$			-0.0000244			-0.0000910			0.0000866+
			(0.000042)			(0.000076)			(0.000048)
Total River Length * $\overline{D}_{ct}$			0.0000001***			0.0000000***			0.0000001***
			(0.000000)			(0.000000)			(0.000000)
Hansen p-value			0.432			0.804			0.587
Kleibergen-Papp LM p-value			0.000			0.003			0.000
Kleibergen-Papp F-stat			12.821			3.131			13.885

Notes: Dependent variable is the change in log NPP in a gridcell. Each column presents the regression coefficients from a separate regression based on gridcells in arid and humid climatic zones as per the GAEZ FAO classification. Standard errors are reported in parentheses. Columns which include gridcell fixed effects show standard errors clustered at the gridcell level, while columns which include country fixed effects show cluster-robust standard errors. Positive (Negative) 1 SD rainfall shock 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. Upstream irrigation dams is a count variable which indicates the number of upstream irrigation dams from the gridcell according to the methodology described in Section 2. Statistical significance is given by + $p<0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table 9: Water Intensive Crop Production and Dam Access

Variable	Obs	Mean	Std. Dev.	Min	Max
Cropland Cells *with* upstream irrigation dams in Arid regions					
Rice harvest area ha.	2,842	12149.25	29763.33	0	157717.1
Cotton harvest area ha.	2,842	1954.299	5100.52	0	40318.32
Sugar harvest area ha.	2,842	1971.822	7672.468	0	82723.04
Cropland Cells *without* upstream irrigation dams in Arid regions					
Rice harvest area ha.	57,512	4123.428	16994.03	0	208244.2
Cotton harvest area ha.	57,512	2142.328	9338.82	0	210628.4
Sugar harvest area ha.	57,512	742.1464	5066.367	0	125368.6

Table 10: Crop Productivity, Cropping Patterns and Income Groups

	Dependent Variable: Δ Log NPP				
	(1)	(2)	(3)	(4)	(5)
	Arid	Arid	High-income	Middle-income	Low-income
		without water-intensive cells			
	2SLS	2SLS	2SLS	2SLS	2SLS
Upstream irrigation dams	0.2044*	0.0452	0.0289	0.0950***	0.1157*
	(0.096)	(0.033)	(0.067)	(0.026)	(0.048)
Pos 1SD rainfall shock	0.1549***	0.0808***	0.0613***	0.1023***	0.0094
	(0.028)	(0.004)	(0.013)	(0.011)	(0.006)
x Upstream irrigation dams	-1.3145**	-0.0853	-0.5249*	-0.7215***	0.0204
	(0.453)	(0.106)	(0.239)	(0.160)	(0.079)
Neg 1SD rainfall shock	-0.0799***	-0.0965***	-0.1003***	-0.0511***	-0.0219***
	(0.006)	(0.004)	(0.012)	(0.005)	(0.004)
x Upstream irrigation dams	-0.2298**	0.0765	0.3418*	-0.2086**	-0.3269***
	(0.087)	(0.054)	(0.146)	(0.065)	(0.057)
Avg Temp(C)	-0.0016**	-0.0034***	0.0021	-0.0015***	0.0031*
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Avg Temp(C) Sq.	-0.0000	-0.0001+	-0.0001	0.0000**	-0.0001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log(Population)	0.0008	0.0033**	-0.0010	-0.0005	-0.0040*
	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Country specific trends	Yes	Yes	Yes	Yes	Yes
N	52494	18551	26663	77571	9477
Hansen p-value	0.804	0.013	0.000	0.832	0.474
Kleibergen-Papp LM p-value	0.003	0.018	0.002	0.000	0.000
Kleibergen-Papp F-stat	3.131	2.761	2.997	16.633	83.401

Notes: Dependent variable is the change in log NPP in a gridcell. Each column presents the regression coefficients from a separate regression. Robust standard errors are reported in parentheses. Positive (Negative) 1 SD rainfall shock 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. Upstream irrigation dams is a count variable which indicates the number of upstream irrigation dams from the gridcell according to the methodology described in Section 2. Water-intensive gridcells are those whose rice, cotton and sugarcane harvest areas are greater than the global mean. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001

Table 11: Joint Impact of Precipitation Shocks and Dams on Cropland Expansion

	Dependent Variable: Δ Log Cropland								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Full Sample			Arid Regions			Humid Regions		
	OLS	OLS	2SLS	OLS	OLS	2SLS	OLS	OLS	2SLS
Upstream irrigation dams	0.000909	0.000668***	-0.016236*	-0.000160	0.000054	-0.041221***	0.001018	0.000971	0.025903**
	(0.001)	(0.000)	(0.007)	(0.002)	(0.000)	(0.012)	(0.002)	(0.001)	(0.008)
4+ Positive 1SD rainfall shocks	0.002280***	0.000731*	0.000813*	-0.000498*	-0.000496**	-0.000430*	0.007584***	0.002659**	-0.004365
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.003)
* Upstream irrigation dams	-0.003384***	-0.00189***	-0.007642	-0.002765***	-0.00228***	-0.005664	-0.004849+	-0.001568	0.350394*
	(0.001)	(0.001)	(0.008)	(0.001)	(0.001)	(0.008)	(0.003)	(0.002)	(0.157)
4+ Negative 1SD rainfall shocks	0.002917***	0.002634***	0.004742***	0.002508***	0.002596***	0.005450***	0.004974***	0.003149***	0.007115***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)
* Upstream irrigation dams	-0.001934*	-0.001510*	-0.094519**	-0.000454	-0.000668	-0.182662**	-0.003169+	-0.002056	-0.127713**
	(0.001)	(0.001)	(0.031)	(0.001)	(0.001)	(0.070)	(0.002)	(0.010)	(0.048)
Contemporaneous Annual Rainfall	-0.000000	0.000002***	0.000002***	0.000003**	0.000002***	0.000002***	-0.000000	-0.000000	-0.000000
(mm/year)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Contemporaneous Annual temperature (C)	0.000449***	-0.000013	0.000005	-0.000208***	-0.00010***	-0.000079***	0.008141***	0.000427***	0.000236*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Log(Population)	0.000647*	-0.000719***	-0.000597***	-0.000314***	-0.00030***	0.000017	-0.009510	-0.00225***	-0.002615***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gridcell fixed effects	Yes	No	No	Yes	No	No	Yes	No	No
Country fixed effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Country specific time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	228111	224103	227527	161317	160841	161269	65322	61814	64862
R-sq	0.299	0.080		0.388	0.195		0.290	0.069	
First stage:									
Dependent Variable: Upstream irrigation dams									
River Slope Share 1.5-3% *			0.0000594***			0.0000276			-0.0000311*
National Dam Propensity			(0.0000105)			(0.0000169)			(0.0000122)
River Slope Share 1.5-3% * 4+ Negative			-0.0000495*			-0.000106****			0.0000617
rainfall shocks * National Dam Propensity			(0.0000226)			(0.0000243)			(0.0000549)
River Slope Share 1.5-3% * 4+ Positive			-0.0000217			-0.0000351			-0.0000287
rainfall shocks * National Dam Propensity			(0.0000195)			(0.0000243)			(0.0000198)
Total River Length * National Dam			0.00000002***			0.00000002***			-0.0000001***
Propensity			(0.000000003)			(0.000000003)			(0.00000001)
Hansen J p-value			0.5132			0.2805			0.5797
Kleibergen-Papp LM p-value			0.0000			0.0000			0.0012
Kleibergen-Papp F-stat			10.214			7.135			3.321

Notes: Dependent variable is the change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression based on gridcells in arid and humid climatic zones as epr GAEZ FAO classification. Standard errors are reported in parentheses. Columns which include gridcell fixed effects show standard errors clustered at the gridcell level, while columns which include country fixed effects show cluster-robust standard errors. 4+ Positive (Negative) rainfall shocks 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 four or more years out of the prior ten years. Upstream irrigation dams is a count variable which indicates the number of upstream irrigation dams from the gridcell, according to the methodology described in Section 2. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001.

Table 12: Joint Impact of Precipitation Shocks and Groundwater on Crop Productivity

	Dependent Variable: Δ Log NPP		
	(1)	(2)	(3)
		By climatic zones	
	Full sample	Arid	Humid
Positive 1SD rainfall shock	0.053217***	0.078533***	0.032376***
	(0.003)	(0.005)	(0.003)
x Aquifer	-0.003533	-0.002271	-0.006633+
	(0.003)	(0.005)	(0.003)
Negative 1SD rainfall shock	-0.072482***	-0.110142***	-0.033728***
	(0.003)	(0.005)	(0.003)
x Aquifer	0.008060*	0.023223***	-0.003465
	(0.004)	(0.006)	(0.004)
Avg Temp(C)	-0.034495***	-0.027048***	-0.052292***
	(0.002)	(0.002)	(0.003)
Avg Temp(C) Sq.	-0.000076	-0.000700***	0.000581***
	(0.000)	(0.000)	(0.000)
Log(Population)	0.000474	0.024788*	-0.004084
	(0.006)	(0.010)	(0.007)
Year fixed effects	Yes	Yes	Yes
Cell fixed effects	Yes	Yes	Yes
Country specific trends	Yes	Yes	Yes
N	121823	56043	65247
R-sq	0.110	0.161	0.100

Notes: Dependent variable is the change in log NPP. Each Column presents the regression coefficients from a separate regression. Standard errors are reported in the parentheses and are clustered by gridcell. Positive (Negative) 1 SD shock 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. Aquifer represents a dummy for gridcells that have access to large groundwater reserves.
Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001.

Table 13: Impact of Precipitation Shocks on Crop Productivity by Rainfall Quantiles

	Quantiles by long- term average rainfall					
	Q1(driest)	Q2	Q3	Q4	Q5	Q6 (wettest)
Positive 1SD precip shock	0.090912***	0.066339***	0.041608***	0.047834***	0.033626***	0.013390***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Negative 1SD precip shock	-0.101878***	-0.113187***	-0.083649***	-0.057028***	-0.036690***	-0.003967
	(0.004)	(0.005)	(0.005)	(0.004)	(0.003)	(0.003)
Avg Temp(C)	-0.025873***	-0.034324***	-0.031939***	-0.027143***	0.018203*	-0.031515**
	(0.003)	(0.004)	(0.005)	(0.007)	(0.007)	(0.010)
Avg Temp(C) Sq.	-0.000343	-0.000448+	-0.000148	-0.000819***	-0.001026***	0.000182
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log(Population)	0.064926***	0.039898**	-0.069449***	-0.096740***	-0.019786	0.018052*
	(0.016)	(0.015)	(0.015)	(0.024)	(0.016)	(0.008)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cell fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country specific trends	Yes	Yes	Yes	Yes	Yes	Yes
N	20306	20306	20306	20306	20306	20293
R-sq	0.204	0.198	0.137	0.166	0.084	0.139

Notes: Dependent variable is the change in log NPP. Each Column presents the regression coefficients from a separate regression. Standard errors are reported in the parentheses and are clustered by gridcell. Positive (Negative) 1 SD shock 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. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001.

Table 14: The Impact of Dams on Agricultural Productivity at Varying Distances

	Δ Log NPP								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Full sample 2SLS			Arid sample 2SLS			Humid sample 2SLS		
	0-25km	25-50km	50-75km	0-25km	25-50km	50-75km	0-25km	25-50kn	50-75km
Upstream dam	0.3051*	0.0357	-0.0109	0.8345*	0.2044*	0.0920***	0.0848	0.0005	-0.0108
	(0.130)	(0.030)	(0.026)	(0.409)	(0.096)	(0.023)	(0.109)	(0.021)	(0.018)
Pos 1SD rainfall shock	0.0890***	0.0991***	0.0983***	0.1300***	0.1549***	0.1245***	0.0457***	0.0501***	0.0511***
	(0.012)	(0.011)	(0.009)	(0.029)	(0.028)	(0.012)	(0.008)	(0.009)	(0.009)
x Upstream irrigation dam	-3.2608**	-0.7812***	-0.4392***	-5.1323*	-1.3145**	-0.4461***	-1.2674+	-0.3139*	-0.1913*
	(1.022)	(0.178)	(0.088)	(2.592)	(0.453)	(0.105)	(0.729)	(0.145)	(0.084)
Neg 1SD rainfall shock	-0.0678***	-0.0706***	-0.0714***	-0.0799***	-0.0799***	-0.0839***	-0.0413***	-0.0444***	-0.0435***
	(0.003)	(0.004)	(0.004)	(0.007)	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)
x Upstream irrigation dam	0.0331	0.0319	0.0308	-1.3294*	-0.2298**	-0.0896**	0.2438	0.0867	0.0401
	(0.280)	(0.056)	(0.030)	(0.552)	(0.087)	(0.030)	(0.321)	(0.077)	(0.047)
Avg Temp(C)	-0.0015***	-0.0009*	-0.0005	-0.0020**	-0.0016**	-0.0018***	-0.0011+	-0.0003	-0.0000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
Avg Temp(C) Sq.	0.0000	0.0000	0.0000	-0.0000	-0.0000	-0.0000+	0.0000	-0.0000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log(Population)	-0.0002	0.0003	0.0011	0.0009	0.0008	-0.0006	-0.0014*	-0.0015**	-0.0009+
	(0.001)	-0.001	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Country specific trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	116324	116311	116324	52494	52494	52494	63401	63388	63401
Hansen p-value	0.988	0.432	0.540	0.554	0.804	0.050	0.895	0.587	0.649
Kleibergen-Papp LM p-value	0.000	0.000	0.000	0.096	0.003	0.000	0.000	0.000	0.000
Kleibergen-Papp F-stat	4.484	12.821	11.919	1.172	3.131	5.410	5.371	13.885	18.125

Notes: Dependent variable is the change in log NPP in a gridcell. Each column presents the regression coefficients from a separate regression based on gridcells in arid and humid climatic zones as per the GAEZ FAO classification. Cluster-robust standard errors are reported in parentheses.Positive (Negative) 1 SD rainfall shock 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. Upstream irrigation dams is a count variable which indicates the number of upstream irrigation dams from the gridcell according to the methodology described in Section 2. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001

Table 15: Joint Impact of Precipitation Shocks and Dams on Cropland Expansion, Multiple Year Thresholds

	Dependent Variable: Δ Log Cropland											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Full Sample				Arid Regions				Humid Regions			
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Upstream irrigation dams	-0.006111	-0.009531	-0.016236*	-0.01878**	-0.046121**	-0.040333**	-0.0412***	-0.0449***	-0.021199	0.039606*	0.025903**	0.027035***
	(0.009)	(0.008)	(0.007)	(0.006)	(0.017)	(0.014)	(0.012)	(0.012)	(0.032)	(0.017)	(0.008)	(0.006)
Positive rainfall shocks	-0.000308	0.000612*	0.000813*	0.00091*	-0.00064***	-0.000191	-0.00043*	-0.00095**	-0.00617***	-0.000149	-0.004365	0.006835*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.003)	(0.003)
* Upstream irrigation dams	0.000576	-0.003523	-0.007642	0.004797	0.013548	0.007351	-0.00566	0.017917+	0.17522***	0.037574	0.350394*	-0.149288
	(0.006)	(0.005)	(0.008)	(0.011)	(0.008)	(0.006)	(0.008)	(0.009)	(0.040)	(0.037)	(0.157)	(0.202)
Negative rainfall shocks	0.0015***	0.0021***	0.0047***	0.00942**	0.00127***	0.00210***	0.0055***	0.012384*	0.001866	0.004690**	0.00712***	0.010049***
	(0.000)	(0.000)	(0.001)	(0.003)	(0.000)	(0.000)	(0.001)	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)
* Upstream irrigation dams	-0.038***	-0.0388***	-0.09452**	-0.24249	-0.02480***	-0.04404***	-0.18266**	-0.447489	-0.043910	-0.106658**	-0.127713**	-0.214477*
	(0.006)	(0.009)	(0.031)	(0.153)	(0.007)	(0.012)	(0.070)	(0.357)	(0.053)	(0.035)	(0.048)	(0.092)
Annual Precipitation	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.00000***	0.000000	-0.000000	-0.000000	-0.000000
(mm/year)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Annual Temperature (C)	-0.000001	-0.000003	0.000005	0.000007	-0.00008***	-0.00008***	-0.0001***	-0.0001***	0.000192+	0.000293***	0.000236*	0.000358***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log(Population)	-0.0006***	-0.0006***	-0.0006***	-0.0006***	0.000004	-0.000016	0.00002	0.000019	-0.00262***	-0.00254***	-0.00262***	-0.00242***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of years with shocks	2 years	3 years	4 years	5 years	2 years	3 years	4 years	5 years	2 years	3 years	4 years	5 years
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Time Trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	227527	227527	227527	227527	161269	161269	161269	161269	64862	64862	64862	64862
Hansen J p-value	0.3888	0.2549	0.5132	0.6450	0.9599	0.6909	0.2805	0.3160	0.1910	0.0035	0.5797	0.0000
Kleibergen-Papp LM p-value	0.0000	0.0000	0.0000	0.0816	0.0000	0.0000	0.0000	0.0528	0.0000	0.0000	0.0012	0.0507
Kleibergen-Papp F-stat	11.668	11.237	10.214	1.245	8.140	8.511	7.135	1.477	16.022	7.697	3.321	1.264

Notes: Dependent variable is the change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression based on gridcells in arid and humid climatic zones as per GAEZ FAO classification. Cluster-robust standard errors are reported in parentheses. Columns 1, 5, and 9 show results when the dummy variables for precipitation shocks equal 1 when there were 2 or more years with 1+ SD shocks in the past 10 years; columns 2, 6, and 10 when there are 3 or more years with 1+ SD shocks; columns 3, 7, and 11 when there are 4 or more years with 1+ SD shocks, and columns 4, 8, and 12 when there are 5 or more years with 1+ SD shocks. Upstream irrigation dams is a count variable which indicates the number of upstream irrigation dams from the gridcell, according to the methodology described in Section II. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001.

Table 16: Impact of Precipitation Shocks on Cropland Expansion, Alternative Time Periods

	Dependent Variable: Δ Log Cropland								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Full Sample			Arid Regions			Humid Regions		
	1990-2005	1990-2005	1980-2005	1990-2005	1990-2005	1980-2005	1990-2005	1990-2005	1980-2005
Number of Positive 1sd+	0.000670***	0.000660***	0.000583***	-0.000174**	-0.000179**	-0.000190***	0.001988***	0.001990***	0.001810***
shocks in 10 years	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of Negative 1sd+	0.000644***	0.000643***	0.000555***	0.000576***	0.000570***	0.000480***	0.001153***	0.001162***	0.001041***
shocks in 10 years	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Contemporaneous Annual	-0.000000	-0.000000	0.000003*	0.000002*	0.000002*	-0.000000	-0.000000	-0.000000	0.000005*
Precipitation (mm/year)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Contemporaneous Annual	0.000467***	0.000447***	0.000312***	-0.000191**	-0.000164**	-0.000355***	0.008293***	0.008242***	0.006802***
Temperature (C)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
log(Population)	0.000534+			-0.000376***			-0.009218		
	(0.000)			(0.000)			(0.006)		
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gridcell Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Specific Time Trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	228111	231432	289290	161317	164612	205765	65322	65324	81655
R-sq	0.299	0.299	0.257	0.389	0.388	0.351	0.290	0.290	0.247

Notes: Dependent variable is the change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression based on gridcells in arid and humid climatic zones as per GAEZ FAO classification. Cluster-robust standard errors are reported in parentheses. Columns 1, 4, and 7 show results from the main specification, with log(population) included and for the time period 1990-2005, for the full sample, arid regions, and humid regions, respectively. Columns 2, 5, and 8 then omit log(population) but cover the same 1990-2005 time period. Finally, columns 3, 6, and 9 still omit log(population), but cover the time period 1980-2005, extending it by 10 years. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001.

\

Table 17: Joint Impact of Precipitation Shocks and Dams on Cropland Expansion, Alternative Time Periods

	Dependent Variable: Δ Log Cropland								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Full Sample			Arid Regions			Humid Regions		
	1990-2005	1990-2005	1980-2005	1990-2005	1990-2005	1980-2005	1990-2005	1990-2005	1980-2005
Upstream irrigation dams	-0.016236*	-0.008915	-0.000444	-0.041221***	-0.041958**	-0.034269**	0.025903**	0.021202*	0.019929**
	(0.007)	(0.007)	(0.006)	(0.012)	(0.014)	(0.011)	(0.008)	(0.008)	(0.007)
4+ Positive 1SD rainfall shocks	0.000813*	0.000796*	0.001168***	-0.000430*	-0.000451*	-0.000287+	-0.004365	-0.007479*	-0.006051*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.003)	(0.003)
* Upstream irrigation dams	-0.007642	-0.009659	-0.027246***	-0.005664	-0.005443	-0.005909	0.350394*	0.476063*	0.451777*
	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.006)	(0.157)	(0.190)	(0.189)
4+ Negative 1SD rainfall shocks	0.004742***	0.004487***	0.003309***	0.005450***	0.005425***	0.004285***	0.007115***	0.007310***	0.005709***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
* Upstream irrigation dams	-0.094519**	-0.085404**	-0.041634**	-0.182662**	-0.184635*	-0.136741**	-0.127713**	-0.128127*	-0.088228**
	(0.031)	(0.027)	(0.014)	(0.070)	(0.073)	(0.051)	(0.048)	(0.050)	(0.032)
Contemporaneous Annual	0.000002***	0.000001***	0.000001***	0.000002***	0.000002***	0.000001***	-0.000000	0.000001*	0.000001*
Precipitation (mm/year)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Contemporaneous Annual	0.000005	-0.000149***	-0.000151***	-0.000079***	-0.000078**	-0.000074**	0.000236*	0.000174+	0.000153+
Temperature (C)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(Population)	-0.000597***			0.000017			-0.002615***		
	(0.000)			(0.000)			(0.000)		
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Country	Country	Country	Country	Country	Country	Country	Country	Country
Country Time Trends	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Sample	Full	Full	Full	Arid	Arid	Arid	Humid	Humid	Humid
Ses	Robust	Robust	Robust	Robust	Robust	Robust	Robust	Robust	Robust
N	227527	230848	288560	161269	164564	205705	64862	64864	81080
Hansen J-stat	0.429	2.150	0.631	1.165	0.691	0.004	0.307	0.300	0.014
Hansen p-value	0.5132	0.1426	0.4269	0.2805	0.4058	0.9476	0.5797	0.5841	0.9064
Kleibergen-Papp LM-stat	40.691	35.984	42.799	28.435	20.898	25.089	13.437	13.247	12.177
Kleibergen-Papp p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0012	0.0013	0.0023
Kleibergen-Papp F-stat	10.214	9.027	10.731	7.135	5.226	6.267	3.321	3.273	3.018

Notes: Dependent variable is the change in log Cropland in a gridcell. Each column presents the regression coefficients from a separate regression based on gridcells in arid and humid climatic zones as per GAEZ FAO classification. Cluster-robust standard errors are reported in parentheses. Columns 1, 4, and 7 show results from the main specification, with log(population) included and for the time period 1990-2005, for the full sample, arid regions, and humid regions, respectively. Columns 2, 5, and 8 then omit log(population) but cover the same 1990-2005 time period. Finally, columns 3, 6, and 9 still omit log(population), but cover the time period 1980-2005, extending it by 10 years. Statistical significance is given by + $p<0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$.

Table 18: Impact of Precipitation Shocks on Crop Productivity, Alternative Standard Errors

	Dependent Variable: Δ log NPP				
	(1)	(2)	(3)	(4)	(5)
Positive 1SD precip shock	0.050461***	0.050461***	0.050461***	0.050461***	0.050461***
	(0.001)	(0.009)	(0.007)	(0.004)	(0.004)
Negative 1SD precip shock	-0.067026***	-0.067026***	-0.067026***	-0.067026***	-0.067026***
	(0.001)	(0.010)	(0.007)	(0.006)	(0.005)
Avg Temp(C)	-0.001827***	-0.001827*	-0.001827	-0.001827**	-0.001827
	(0.000)	(0.001)	(0.002)	(0.001)	(0.001)
Avg Temp(C) Sq.	0.000022*	0.000022	0.000022	0.000022	0.000022
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log(Population)	-0.000194	-0.000194	-0.000194	-0.000194	-0.000194
	(0.000)	(0.001)	(0.002)	(0.001)	(0.001)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Country Fixed effects	Yes	Yes	Yes	Yes	Yes
Country specific trends	Yes	Yes	Yes	Yes	Yes
Cluster Standard errors	Robust	Country	Country-year	Province	Province-year
N	121823	121823	121823	121823	121823
R-sq	0.086	0.086	0.086	0.086	0.086

Notes: Dependent variable is the change in log NPP in a gridcell. Each column presents the regression coefficients from a separate regression. Standard errors are reported in parentheses. Positive (Negative) 1 SD rainfall shock 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. Statistical significance is given by + p<0.10 * p<0.05 ** p <0.01 ***p < 0.001

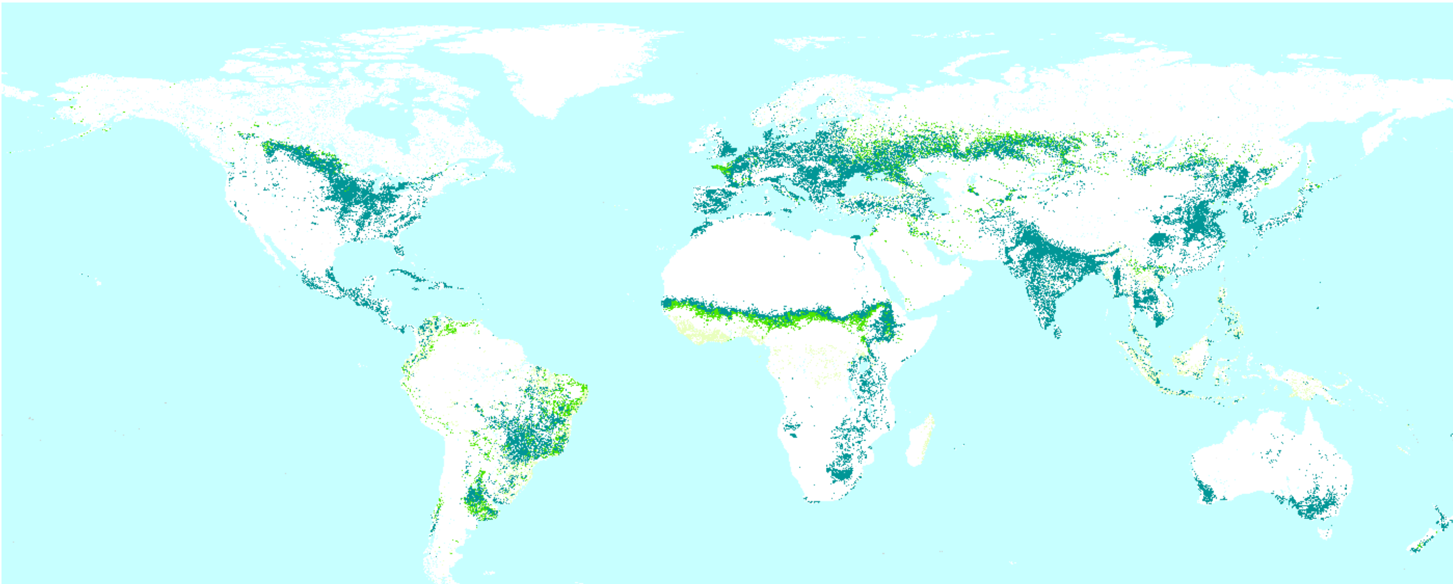
Table 19: Impact of Precipitation Shocks on Cropland Expansion, Alternative Standard Errors

	Dependent Variable: Δ log Cropland				
	(1)	(2)	(4)	(3)	(5)
Number of Positive 1sd+ shocks in 10 years	-0.000095	-0.000095	-0.000092	-0.000095	-0.000095
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of Negative 1sd+ shocks in 10 years	0.000461***	0.000461	0.000466+	0.000461*	0.000461**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Contemporaneous Annual Rainfall (mm)	0.000002***	0.000002	0.000002	0.000002	0.000002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Contemporaneous Mean temperature (C)	-0.000013	-0.000013	-0.000012	-0.000013	-0.000013
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
log(Population)	-0.000727***	-0.000727+	-0.000727	-0.000727*	-0.000727
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes
Country Specific Time Trends	Yes	Yes	Yes	Yes	Yes
Cluster Standard Errors	Robust	Country	Country-year	Province	Province-year
N	228111	228111	224103	228111	228111
R-sq	0.082	0.082	0.080	0.082	0.082

Notes: Dependent variable is the change in log cropland in a gridcell. Each column presents the regression coefficients from a separate regression. Standard errors are reported in parentheses. Positive (Negative) 1 SD rainfall shock 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. Statistical significance is given by + $p<0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Figures

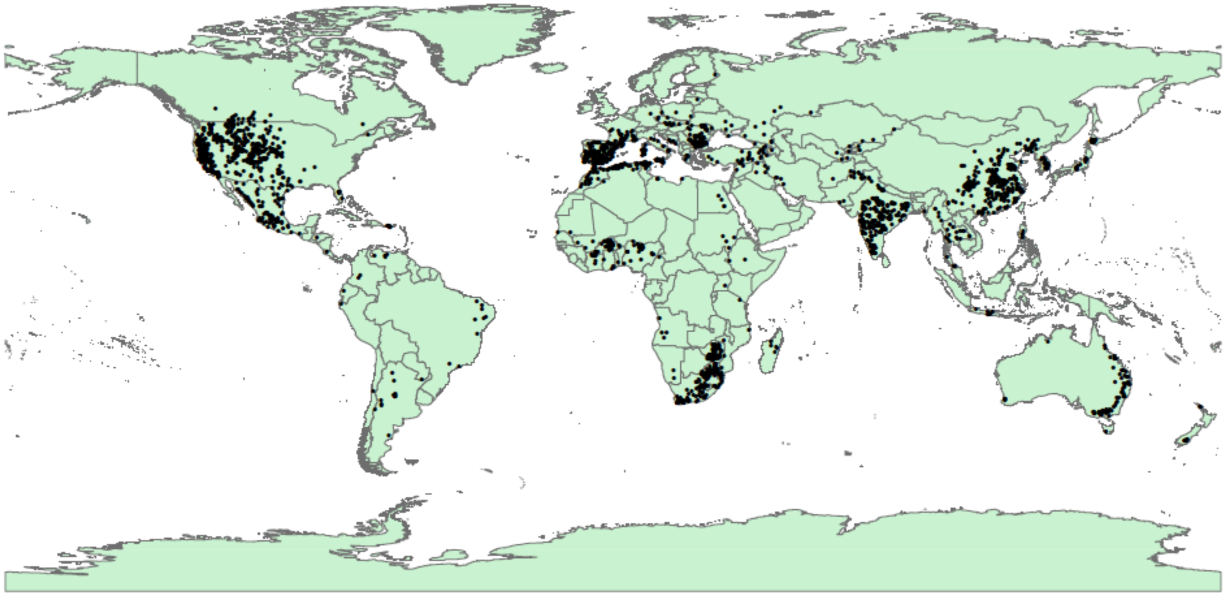
Figure 1: Global Land Cover 2000



Water Bodies
Mosaic: Cropland/Tree Cover/Other natural
Mosaic: Cropland/Shrub and/or grass cover
Cultivated and managed areas

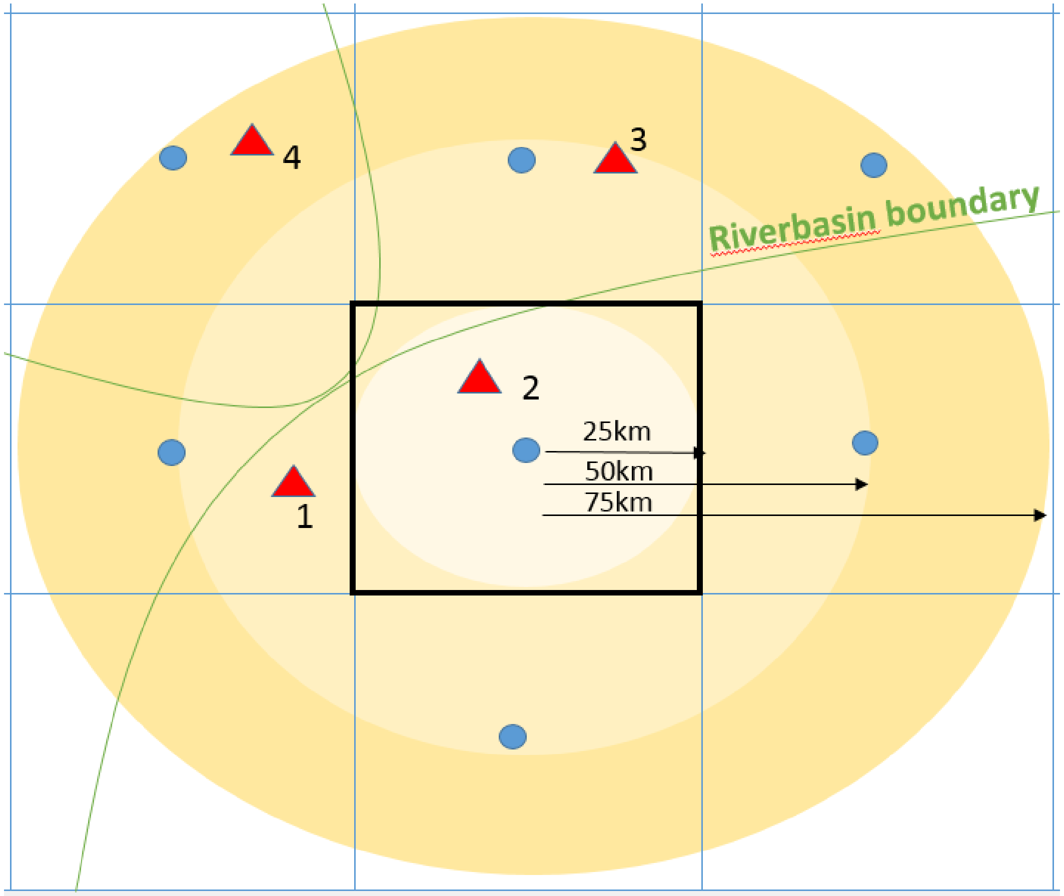
Notes: Gridcells falling into either of these categories are identified as cropland areas. Source: GEM (2011) Global Land Cover 2000 Data Set

Figure 2: GRanD Irrigation Dam Locations



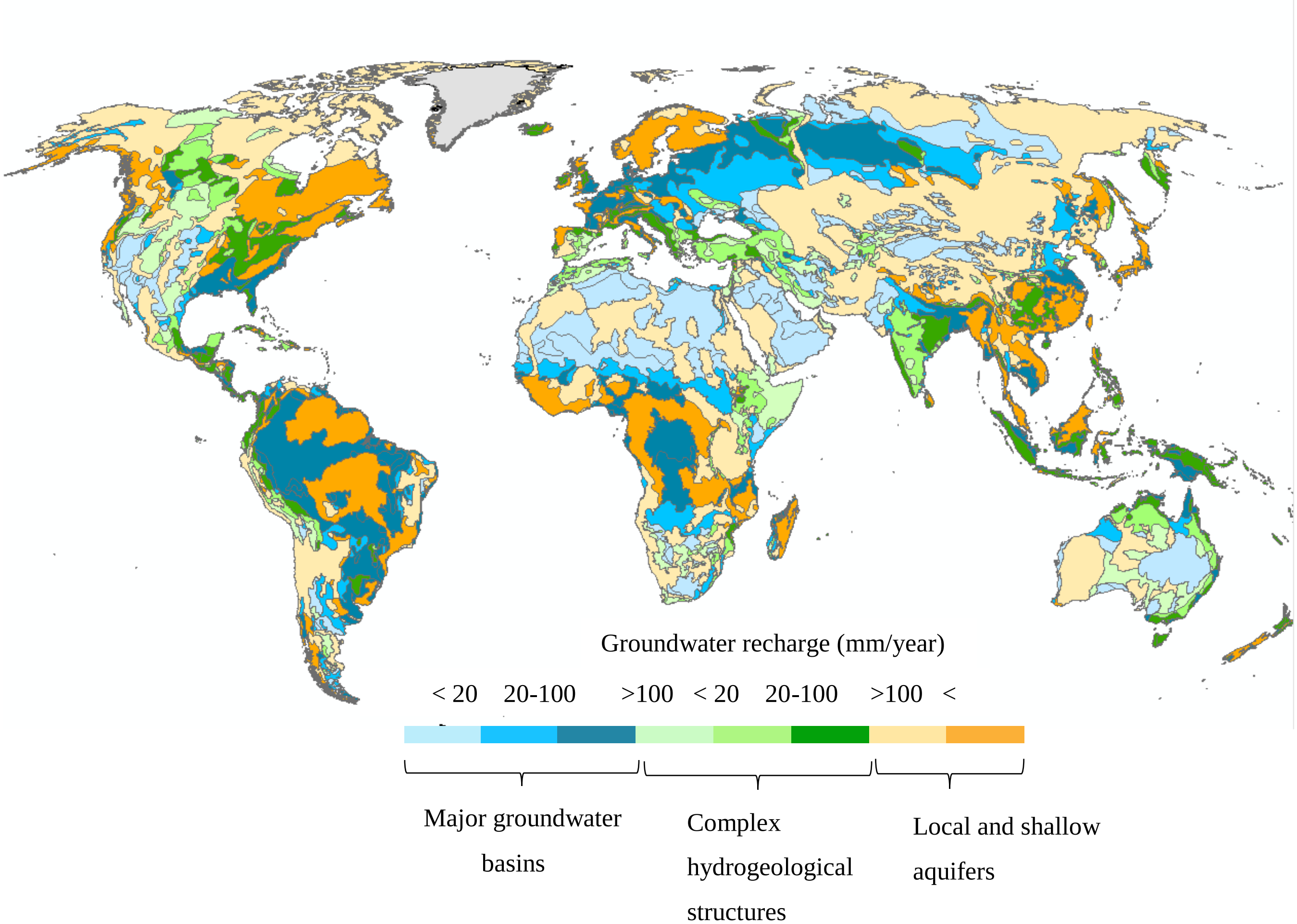
Notes: Image shows locations of dams in the GRanD dataset which have an irrigation purpose. Dams with no data on year of construction were excluded.

Figure 3: Defining the Command Area of Dams



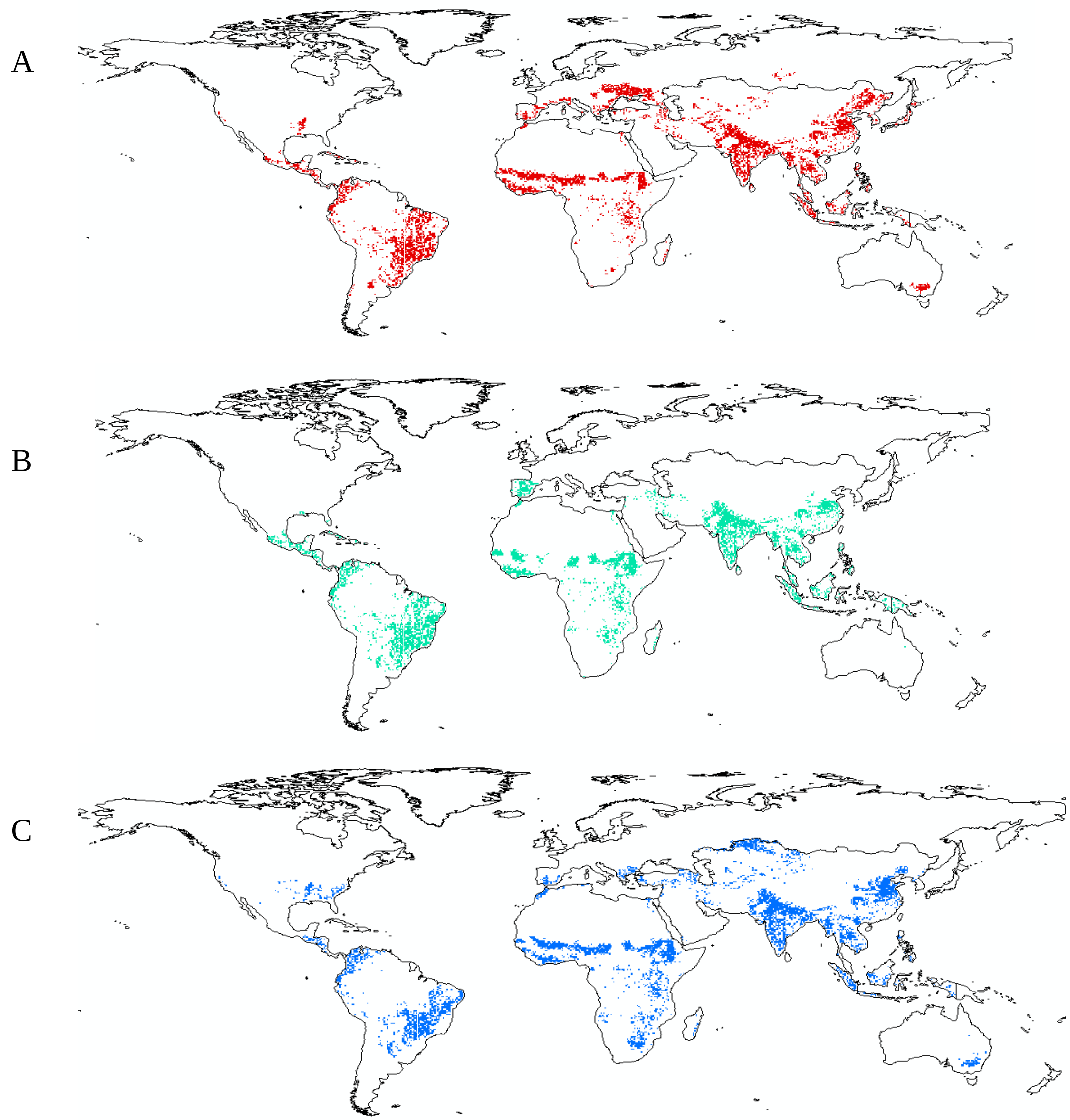
Notes: In this figure, the gridcells represent 0.5 x 0.5 degree grids (approximately 50kms x 50 kms at the equator) used in our sample. Gridcell centroids are marked by circles, and dams are labeled by triangles. Concentric circles around the centroid of the selected gridcell show distance thresholds 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.

Figure 4: Groundwater resources of the world



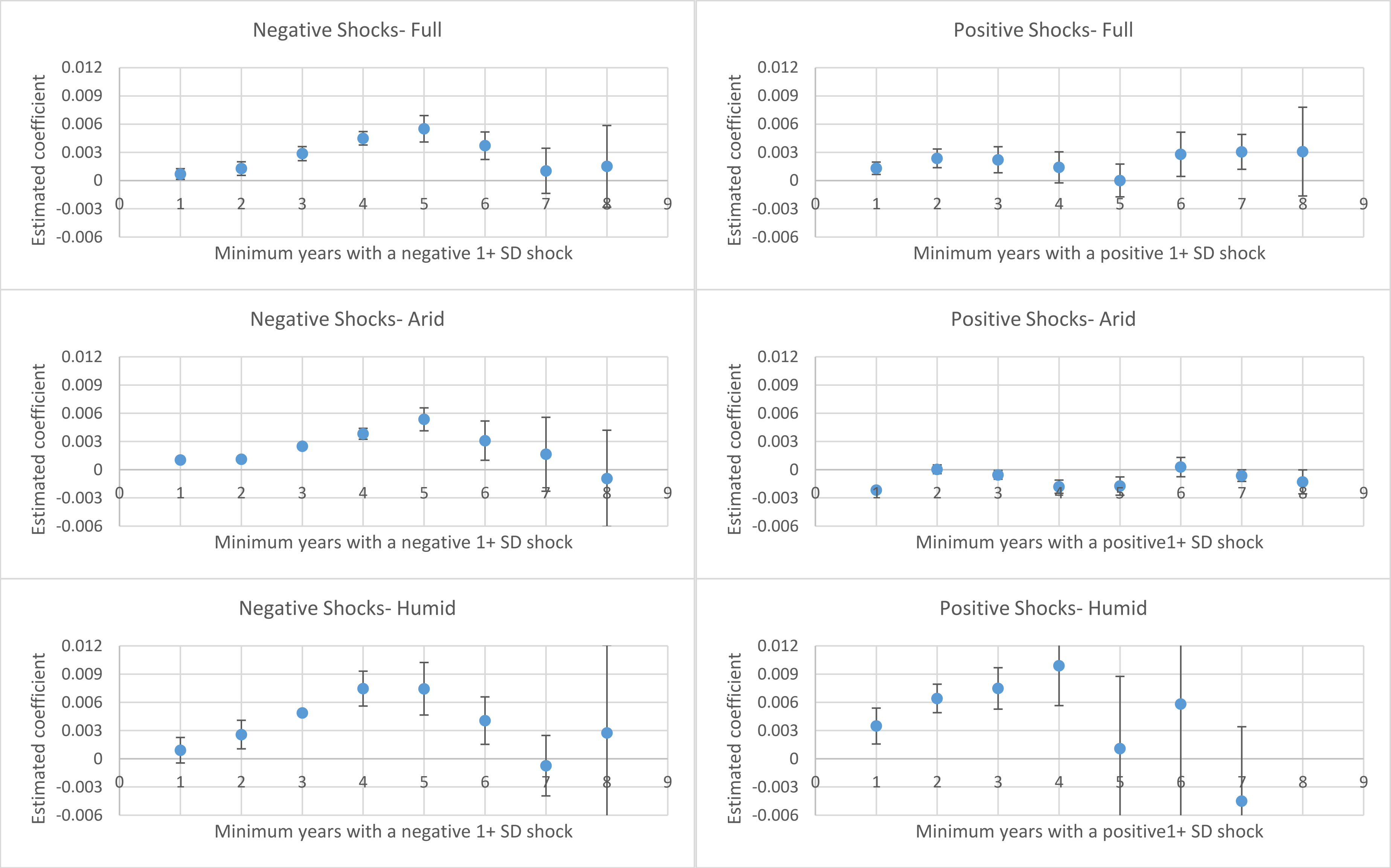
Notes: Large groundwater regions are depicted in blue and green with color gradations indicating recharge rates. Darker(lighter) colors indicate higher (lower) recharge rates.
Source: WHYMAP- World-wide Hydrogeological Mapping and Assessment Programme, BGR/UNESCO

Figure 5: Global distribution of rice, sugarcane and cotton growing cells



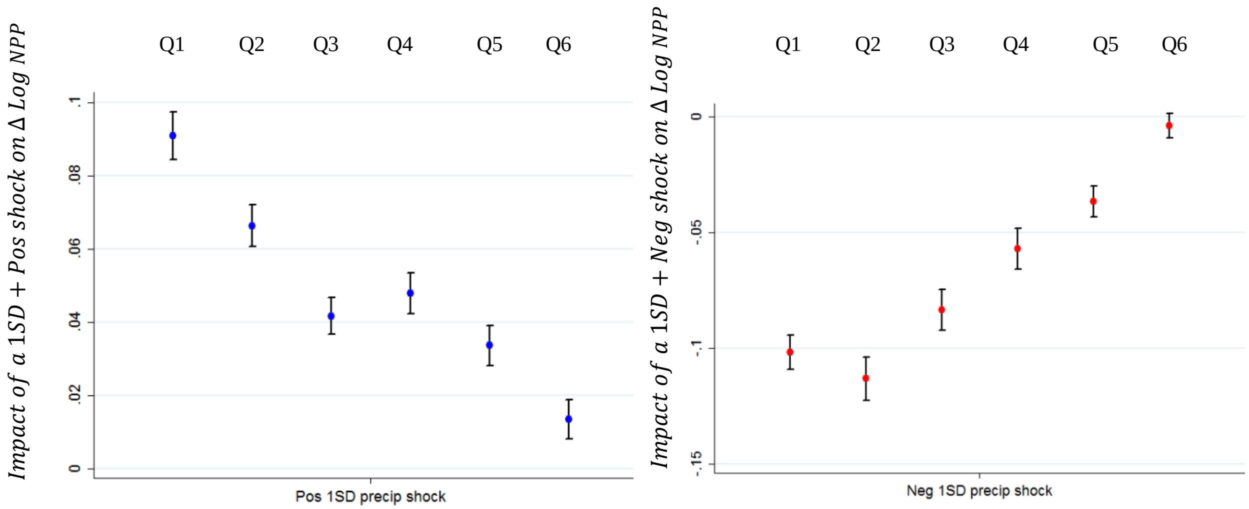
Notes: Panel A, B and C together depict gridcells whose harvested area for rice, sugarcane and cotton is greater than the global median respectively. Source: Monfreda et al. (2008)

Figure 6: Non-linear Impact of Precipitation Shocks on Cropland Production



Notes: Figure shows point estimates and 95% confidence intervals from coefficients obtained from estimating equation 2 with $Prec10_{it}^{-}$ and $Prec10_{it}^{+}$ replaced with binary variables equal to 1 if the gridcell experienced at least X years of precipitation shocks within the past 10 years, where X is the value given on the x-axis. Row 1 shows regressions using the full sample, row 2 uses only arid (as defined by FAO GAEZ) gridcells, and row 3 uses only humid gridcells. In each graph, plotted coefficients are estimated using separate regressions, but negative and positive coefficients are estimated jointly for corresponding samples values on the x-axis. For instance, for the full sample, the coefficient corresponding to 3 years of negative shocks is estimated in the same regression as the coefficient corresponding to 3 years of positive shocks.

Figure 7: Impact of Precipitation Shocks on Crop Productivity by Rainfall Quantiles



Notes: Figure shows point estimates and 95% confidence intervals from coefficients obtained from estimating equation (1) for each of the 6 quantiles, Q1 to Q6, representing gridcells with lowest amounts of long term average rainfall, to those with the highest amounts of long term average rainfall. In each graph, plotted coefficients are estimated using separate regressions, but negative and positive coefficients are estimated jointly. Full regression results are in Table 13