Agriculture and Arsenic: Can Overextraction of Groundwater Make Us Sick?*

Natasha Jha

Preliminary draft : Do not cite

Abstract

This paper examines how groundwater irrigation in India, while crucial for poverty reduction, created an environmental externality through arsenic contamination. Using historical variation in groundwater stocks and high-yielding variety seed diffusion, I find that districts with richer groundwater endowments have 70% more habitats exceeding safe arsenic levels in their groundwater. The relationship is strongest in districts with medium-thick aquifers, which had higher early adoption of deep tubewells. Analysis reveals that historical irrigation practices, rather than contemporary agricultural inputs, drive arsenic contamination, highlighting the long-term environmental consequences of irrigation technology choices.

JEL Codes: Q53, Q56, Q15, 013

Keywords: Groundwater Contamination, Environmental Externalities, Irrigation Technology, Green Revolution, Indian Agriculture

*Jha:University of Notre Dame (njha@nd.edu).

1 Introduction

Groundwater contamination poses a major threat to public health and economic development in the developing world. Of the 220 million people exposed to arsenic-contaminated groundwater globally (Podgorski and Berg, 2020), 50 million reside in India (Shaji et al., 2021). The health costs are severe, as arsenic exposure leads to skin lesions, cardiovascular and hypertension disorders, adverse consequences on pregnancy, and higher mortality. Beyond health impacts, arsenic exposure also impairs socioeconomic outcomes, with negative impacts on labor supply, cognitive abilities, educational attainment, earnings, and marriage market outcomes.[1]

Though arsenic contamination is largely naturally occurring, I examine how anthropogenic actions, particularly irrigation, intensify this contamination in India. Irrigation has been transformative for poverty reduction, but its environmental costs are mounting with depleting surface water sources. As climate change accelerates this depletion, communities will become increasingly dependent on groundwater. However, intensive groundwater extraction mobilizes arsenic contamination, creating a vicious cycle of deteriorating water quality. I study how this cycle unfolds by examining the mechanisms through which groundwater irrigation triggers arsenic contamination and quantifying the resulting environmental externality.

I begin by documenting the increased reliance on groundwater irrigation starting in the 1960s, using data on agricultural electricity connections, which primarily powered groundwater pumping through tubewells. This period coincided with India's Green Revolution - a time of agricultural transformation through high-yielding variety (HYV) seeds of rice, wheat and maize. I find the growth in electricity connections tracked closely with the expansion of HYV cultivation area, highlighting how groundwater irrigation enabled timely crop irrigation crucial for HYV success. Further analysis of cultivation patterns across groundwater stocks reveals that areas with deeper aquifers saw the largest gains in HYV area, demonstrating how groundwater accessibility shaped the intensity of irrigation.

I next study how the increased reliance on groundwater irrigation generated environmental externalities through arsenic contamination. My empirical strategy exploits historical variation in groundwater stocks combined with the diffusion of HYV seeds that increased irrigation's marginal value (D'Agostino, 2017). Using water testing data from 176 million sites in rural

[1] See Ahmad et al. (2018); Argos et al. (2010); Mukherjee et al. (2006); Sohel et al. (2009) for health costs and Carson et al. (2011); Chowdhury and Singh (2021); Pitt et al. (2021) for socioeconomic consequences.

India over 2009-2017, I find that communities in districts with richer groundwater stocks face higher arsenic exposure, with these districts having 70% more habitats exceeding safe arsenic levels in their groundwater. This relationship persists even after controlling for time-varying district characteristics that could influence contamination, including population, electricity access, forest cover, economic activities, and fertilizer use.

I then examine the mechanisms through which groundwater extraction triggers arsenic contamination by analyzing variations in aquifer depth across water-abundant districts (Erban et al., 2013). Based on hydrological evidence linking extraction intensity to arsenic release, I find the relationship between groundwater extraction and contamination is strongest in districts with medium-thick aquifers. Historical data on irrigation technology diffusion reveals these districts had a 150% higher stock of deep tubewells in the 1980s, suggesting early adoption of high-extraction technologies created pathways for arsenic mobilization.

I further address alternative explanations for arsenic contamination, particularly those related to agricultural practices - rice cultivation patterns, fertilizer application, and surface water runoff. Using spatial analysis and panel data estimation, I find that household access to groundwater is a stronger predictor of arsenic exposure compared to these alternative pathways, particularly in the Gangetic Plain where contamination is most severe.

This paper contributes to three strands of literature. First, while the Green Revolution successfully reduced poverty (Gollin et al., 2021), recent studies highlight its unintended health consequences, from negative effects of agrichemicals on child health (Brainerd and Menon, 2014) to increased cardiovascular diseases from dietary changes (Sekhri and Shastry, 2024). I add to this literature by examining an overlooked externality - the impact of irrigation technology choices on groundwater arsenic contamination. Second, while economics research has examined groundwater irrigation's effects on agricultural productivity (Badiani and Jessoe, 2011), efficiency, and equity (Ryan and Sudarshan, 2022), these studies primarily focus on short-term impacts. I extend this literature by investigating long-term effects of irrigation choices, showing how early adoption of high-extraction technologies created persistent pathways for contamination. Although the Green Revolution began in the 1960s, arsenic contamination emerged as a policy concern in India only in the 2010s. Third, I contribute to research on groundwater quality degradation. Prior work has documented consequences of declining groundwater levels

(Sayre and Taraz, 2019; Zaveri et al., 2016), but changes in water quality remain understudied. This is particularly important as heavy metal contamination, unlike groundwater depletion, can permanently compromise water sources. While hydrology studies suggest similar relationships, they rely on small-scale, correlational evidence. This paper provides causal estimates linking anthropogenic activity to deep groundwater quality degradation, focusing specifically on how extraction mechanisms lead to contamination.

The paper proceeds as follows: Section 2 provides background on the Green Revolution and Aquifers; Section 3 presents a conceptual framework linking hydrology insights to empirical design; Section 4 describes data sources; Section 5 discusses empirical strategy and identification assumptions; Section 6 presents first stage and reduced form results; Section 7 explores mechanisms; Section 8 provides robustness tests; and Section 9 concludes.

2 Background

2.1 Green Revolution

The Green Revolution marked India's shift to modern agriculture through high-yielding varieties (HYV) of rice and wheat in 1966 (Evenson and Gollin, 2003). This transformation, evidenced by sharp increases in HYV cultivation (panel (b) and (c) in Figure 1), fundamentally changed irrigation needs. While traditional crops could rely on rain-fed agriculture, HYV seeds required precise water management that only groundwater irrigation could provide. Agricultural electricity connections, primarily used for groundwater pumping through tubewells, offer a reliable measure of groundwater irrigation adoption during this period (panel (a) in Figure 1). This technological complementarity drove the largest productivity gains in Asia's irrigated regions (Gollin et al., 2021). Sekhri (2014) documents how HYV adoption followed groundwater accessibility patterns across India, providing variation that I exploit to examine the environmental consequences of increased groundwater extraction.

2.2 Aquifers

Groundwater accessibility depends critically on aquifer characteristics. An aquifer system consists of two key zones: the unsaturated zone below land surface containing both water and air, and the saturated zone where groundwater pools above bedrock. The boundary between these zones, called the water table, determines well depths and extraction costs. Areas with deeper unsaturated zones require costlier well installation and operation. However, thicker aquifers help maintain higher water tables, reducing extraction costs and enabling sustainable groundwater use. These geological features shaped both the initial adoption of groundwater irrigation and its long-term environmental consequences.

2.3 Groundwater Extraction and Arsenic Release

Hydrology research suggests that groundwater extraction can trigger arsenic release through land subsidence. As aquifers are depleted, soil compression can mobilize arsenic from solid to liquid phase (Erban et al., 2013). This process varies systematically with aquifer depth - deeper aquifers require more intensive extraction to trigger arsenic release, leading to longer release periods. Medium-depth aquifers are particularly vulnerable, as they balance extraction feasibility with arsenic mobilization potential. While most arsenic contamination occurs naturally over centuries through geological processes (Chakraborty et al., 2015; Fendorf et al., 2010), the recent emergence of contamination across India's Indo-Gangetic plain suggests a role for human activity. I exploit this variation in aquifer depths and historical irrigation adoption to examine how extraction intensity influences contamination patterns.

3 Data

My analysis combines district-level data on aquifer characteristics, agricultural technology adoption, and water quality across India. Districts are the relevant administrative unit for agricultural policy implementation, including irrigation expansion and mechanization. This administrative structure, combined with consistent data availability at the district level, makes it the most suitable unit for examining irrigation's environmental impacts. The dataset spans 1966-2017 and integrates aquifer depth measures, agricultural data on HYV adoption and electricity connections, water quality testing from rural sites (2009-2017), and district controls. The following sections provide more details.

3.1 World Bank India Agricultural and Climate Data

The India Agriculture and Climate (IAC) dataset provides district-level data on HYV adoption and aquifer characteristics from 1957-1987 across 271 districts in 13 major states. This period captures the pre- and post-Green Revolution transition in Indian agriculture. For aquifer depth, districts are classified as "Thickest" ($>$ 150m), "Medium-Thick" (100-150m), or "Thick" ($<$ 100m), with data available for 124 districts. Using Water Resource plates from the National Atlas of India, I classify remaining districts as "Sporadic", though these are largely excluded from analysis due to distinct irrigation patterns.[2] Panel (a) in Figure 3 presents the spatial distribution across the categories of aquifer thickness.

3.2 Arsenic Exposure

My measure of arsenic exposure comes from the National Rural Drinking Water Program's (NRDWP) water quality testing data over 2009-2017.[3] The program collects water samples from rural sites across India for testing in state laboratories. Data is reported as the number of habitats and their population exposed to contaminated water sources, where contamination is defined as exceeding the Indian safety threshold (50 g/L).[4] This dataset, combined with the 1957-1987 IAC data, allows me to examine how historical irrigation choices affect contemporary arsenic exposure. I aggregate habitat-level exposure to district level using population weights, then merge with the IAC dataset. Panel (b) in Figure 3 presents the spatial distribution of arsenic exposure for the country.

3.3 Groundwater Extraction

To examine extraction mechanisms, I digitized the 1986-87 Minor Irrigation (MI) census data, which records district-wise information on irrigation schemes with command areas under 2000 hectares. The census provides annual counts of dugwells, shallow tubewells, and deep tubewells from 1983-1987, along with pre-1982 totals. Deep tubewells, with output approximately 15

[2] States included are Andhra Pradesh, Bihar, Gujarat, Haryana, Karnataka, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal.

[3] NRDWP is a centrally sponsored scheme aimed at improving access to safe drinking water in rural India.

[4] The WHO threshold is set at 10 g/L. The data only indicates whether a source exceeds the threshold, without reporting actual contamination levels.

times that of average tubewells, serve as a proxy for high-extraction irrigation technology.[5] This historical data on irrigation technology adoption, combined with aquifer characteristics and contemporary arsenic exposure, allows me to test whether early adoption of intensive extraction methods created pathways for contamination.

3.4 Covariate Data

I supplement the analysis with district-time varying controls that could influence arsenic contamination. Using the Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG), ICRISAT-TCI database, and IAC, I control for power supply (affecting extraction intensity), rural population (exposure), proximity to water bodies (alternative sources), employment (economic activity), and fertilizer use (potential contaminant). As shown in Appendix Table 9, summary statistics across aquifer depths reveal systematic differences in agricultural intensity. Districts with the thickest aquifers show highest power supply, fertilizer use, and cropping intensity, indicating greater agricultural intensification. Surface irrigation remains consistent across aquifer depths, suggesting aquifer depth primarily affects groundwater irrigation choices. Contemporary groundwater irrigation is highest in thin and thickest aquifer districts, reflecting the combined use of shallow and deep tubewells.

4 Empirical Strategy

The empirical strategy closely follows that of Sekhri and Shastry (2024) and combines historical variation in water resource distribution with a policy shock that increased groundwater irrigation extraction to measure the impact on arsenic exposure. The following sections provide details on identification assumptions and threats to identification.

4.1 Identification

My identification strategy exploits historical variation in aquifer depth to instrument for Green Revolution-induced groundwater extraction.[6] I establish aquifer depth as a valid instrument by

[5]Further details on the MI census are provided in Appendix section C

[6]See (Glaeser et al., 2015; Hornbeck and Keskin, 2014; Michaels, 2011) for similar designs using natural resource endowments as instruments.

showing its relationship with HYV adoption - districts with deeper aquifers faced lower extraction costs and consequently saw higher Green Revolution "take-up." The first-stage estimation examines how aquifer depth influenced HYV adoption:

$$y_{ist} = \beta_o + \beta_1 Post_t * TKST_{is} + \beta_2 Post_t * MTKST_{is} + \beta_3 Post_t * TNST_{is} + \lambda X_{ist} + \gamma_t + \mu_{is} + \epsilon_{ist} \quad (1)$$

where y_{ist} is HYV cultivation area in district i, state s, year t; $TKST$, $MTKST$ and $TNST$ indicate "thickest," "medium-thick," and "thin" aquifers respectively (with "sporadic" as reference); $Post_t$ indicates post-1966; X_{ist} controls for agro-climatic variables; and γ_t, μ_i are year and district fixed effects. The coefficients β_1, β_2, and β_3 capture differential HYV adoption across aquifer depths. With near-zero HYV adoption pre-1966, pre-trends are not a concern. Results are reported in Table 1.

For causal identification, historical aquifer depths must affect arsenic exposure only through groundwater extraction. Several factors support this assumption. First, aquifer formation is historically determined, eliminating concerns of endogenous placement. Second, while water abundance could influence firm location and industrial pollution (Hagerty and Tiwari, 2022; Liu and Sekhri, 2021), existing research suggests these effects operate through surface water and water table variation rather than aquifer depth differences.

Migration poses potential challenges to identification. However, aquifer depths are largely unobservable to individuals, suggesting migration decisions are likely orthogonal to this characteristic. While prior work links groundwater access to poverty (Sekhri, 2014), making income-driven migration a concern, I address this through robustness tests. Additionally, arsenic-induced migration is unlikely given the contaminant's undetectable nature and limited public awareness campaigns across India.

4.2 Empirical Equation

Regressing arsenic exposure on groundwater extraction raises endogeneity concerns due to correlated policy measures. States better equipped for electricity provision often have stronger infrastructure for contamination monitoring. As Figure 5 shows, eastern states with poor infrastructure also have lower electric pump access. While state fixed effects could address time-invariant characteristics, state infrastructure capacity likely varied during the study period. Historical aquifer thickness provides an exogenous source of variation unlikely to correlate with policy measures beyond groundwater extraction. Panel (b) and (c) in Figure 1 show that post-1966 HYV adoption was substantially higher in ”thickest” and ”medium-thick” aquifer districts, while ”thin” and ”sporadic” districts show similar, low adoption. This pattern motivates combining ”thickest” and ”medium-thick” districts as water-rich areas distinct from water-scarce districts. I estimate:

$$y_{ist} = \beta_o + \beta_1(TKST_{is}, MTKST_{is}) + \lambda X_{it} + \gamma_t + \delta_s + \epsilon_{ist} \tag{2}$$

where Y_{ist} measures arsenic exposure, X_{ist} controls for alternate contamination pathways (fertilizer use, economic activity, industrial presence, river proximity), and γ_t, δ_s capture year and state effects. The coefficient β_1 measures differential arsenic exposure between water-rich and water-scarce districts. Population-weighted estimates are documented in Table 2. To understand heterogeneity in contamination patterns, I further decompose the analysis within water-rich districts:

$$y_{ist} = \beta_o + \beta_1 TKST_{is} + \beta_2 MTKST_{is} + X_{ist} + \gamma_t + \delta_s + \epsilon_{ist} \tag{3}$$

separating ”thickest” and ”medium-thick” effects (β_1, β_2) relative to ”thin” aquifers. Estimation results are detailed in Table 3.

5 Results

5.1 First Stage - Adoption of High-Yield Varieties

Table 1 presents estimation results using equation 1 which are consistent with the graphics presented in Figure 2b & 2c. Area under cultivation post Green Revolution saw a significant increase for “thickest” aquifer depths (112%) followed by “medium-thick” aquifer depth (61%) as compared to “sporadic aquifers”. For “thin” category there is no change in area under cultivation as compared to sporadic aquifers. These results are indicative that aquifer depth played a role in take up of the Green Revolution. Greater water availability led to expansion in area under cultivation for high-yield variety seeds. Given the similarity of magnitude for thickest and medium-thick districts, going ahead the analysis clubs these as "water-rich" districts and uses thin and sporadic aquifers as "water-scarce" districts for comparison.

5.2 Arsenic Exposure

Table 2 reports second stage results on habitats exposed to arsenic contamination using specification 2. I find a significantly higher exposure to arsenic contamination in water-rich districts as compared to water-scarce districts. Water-rich districts have 900 - 700 higher number of habitats exposed to arsenic contamination in drinking water above safety thresholds. These coefficient magnitudes translate to an effect size in the range of 50%-70% higher number of habitats exposed to arsenic contamination in water-rich districts as compared to a sample average of 1278 habitats. This result is qualitatively consistent and persists across differences in specifications controlling for year fixed effects, district time-varying controls and state fixed effects.

To test if there is a specific aquifer depth within water-abundant districts driving the result in Table 2, I present estimation results from equation 3 in Table 3. A decomposition of the within water-abundant districts finds that the increase in arsenic exposure accrues to "medium-thick" aquifer districts, which have larger magnitudes on arsenic exposure. "medium-thick" aquifer districts have approximately 900-1200 higher habitats exposed to arsenic contamination as compared to "thin" aquifer districts. This effect size is larger and represents a size in the range of approximately 70% - 90% greater arsenic exposure in "medium-thick" aquifer districts. For "thickest" aquifer districts, the result is less consistent. In the absence of controls for district-time varying characteristics, "thickest" aquifer districts have a negative and significantly lower arsenic exposure as compared to "thin" aquifer districts. However, the inclusion of district time varying covariates suggests that there are no significant differences in the magnitude of arsenic exposure between "thickest" and "thin" aquifers.

I interpret the difference in arsenic exposure within water-abundant districts as speculative evidence to support arsenic release due to compaction in "medium-thick" aquifers. While "thickest" aquifer districts saw an increase in in the diffusion of HYV seeds and the subsequent increase in groundwater pumping, it is plausible that groundwater extraction rates were unable to (or have not yet) triggered land subsidence to an extent to create mobile arsenic in the water source. I explore this dynamic in the next section.

6 Mechanisms

To examine differences in arsenic exposure for within water-rich districts I explore differences in stocks of groundwater irrigation technology across different aquifer depths. Using information on number of dugwells, shallow tubewells and deep tubewell irrigation schemes, I examine the temporal distribution of each groundwater irrigation technology in Figures 6, 7 and 8 respectively. Dugwells are by far the most popular in "thin" aquifer districts, and the distribution of shallow tubewells across aquifer depths is fairly consistent until the dramatic growth for "thickest" aquifer districts post 1986. However, the most interesting result is the variation in distribution of deeptubewells with a consistently higher stock of deep tubewells in "medium-thick" aquifer districts.

Deep tubewells operate round the clock during irrigation season[7] and have an annual output roughly 15 times that of an average shallow tubewell. These features make deep tubewells a "high-extraction" groundwater irrigation technology. To provide evidence on groundwater extraction as the mechanism behind arsenic contamination, I create a district year panel for all districts in the IAC over the period pre 1982, 1983 - 1987.

To provide an empirical estimate of the stock of deep tubewell on aquifer depth I estimate the following regression specification:

$$y_{ist} = \beta_o + \beta_1 TKST_{is} + \beta_2 MTKST_{is} + \beta_3 TNST_{is} + \gamma_t + \delta_s + \epsilon_{ist} \tag{4}$$

y_{ist} is the number of deep tubewells in district i in state s in year t. $TKST$, $MTKST$ and $TNST$ are aquifer thickness that equal one if district i contains "thickest", "medium-thick" or "thin" aquifer respectively. The reference category is "sporadic" aquifer depth. γ_t and δ_s represent year and state level fixed effects to control for year-specific shocks and time-invariant state characteristics. Robust standard errors are clustered at the district-level given the implementation of agricultural policies at the district level. The coefficients of interest are β_1, β_2, and β_3 which represent the difference in stocks of deep tubewells across "thickest","medium-thick",

[7] depending on the availability of power

and “thin” aquifer depths as compared to “sporadic” aquifer depths. Estimation results from equation 4 are reported in Table 4.

The results indicate a significantly greater number of deep tubewells in “medium-thick” aquifer districts and is indicative of high levels of groundwater extraction experienced by these districts since the 1980s.

6.1 Competing Mechanisms

A major arsenic exposure pathway is through the consumption of rice grains that are irrigated using arsenic contaminated groundwater (Rahman et al., 2019). Irrigation of rice fields using arsenic contaminated groundwater can lead to accumulation of arsenic in soils, thereby permeating a cycle. The Green Revolution contributed towards a rapid spread of rice-wheat systems in the Indo-Gangetic plane (Pingali, 1999), which is highlighted in figure 9.[8]. Examining cropping patterns provides insights into the link between cropping patterns and exposure to arsenic. Visually comparing Figure 9 with panel (b) of Figure 3 suggests poor spatial correlation between incidence of arsenic exposure and dominant rice patterns. Rather, arsenic exposure in panel (b) of Figure 3 correlates closer with Figure 10 which documents groundwater stressed blocks of India.[9]

To provide empirical evidence on alternate competing mechanisms such as fertilizer application and surface water runoff, I collect data on agricultural input usage across districts. I combine these data with information on how households access groundwater for own consumption to create a district year panel. Using a two-way fixed effects model with district and time fixed effects I present motivating regression estimates to highlight that household accesses to groundwater is a consistent predictor of arsenic exposure over alternate input usage such as surface water irrigation or fertilizer use.[10]

Tables 7 and 8 present estimation results on the the proportion of total population affected for the entire sample and a sub-sample of states in the Gangetic Plane. Across specifications access to groundwater as measured by % hh with groundwater access is a significant predictor of arsenic exposure (Columns (3) and (4) in Table 7). This result has a higher magnitude

[8] Replicated Figure from the TCI annual report

[9] The eastern Indian state of West-Bengal shares geo-logical properties with Bangladesh and has always been a high arsenic contaminated region.

[10] More details on the data and methodology can be found in the appendix section D.

among households that reside in the Gangetic Plane (Columns (3) and (4) in Table 8). Taken together these results suggest that increased arsenic exposure is less likely to accrue from fertilizer application or surface water runoff, rather it is the access of households to arsenic contaminated sources of groundwater.

7 Robustness

Given the cross-sectional variation in the instrument used in the paper, there might be concerns on estimating 2 on a district year panel. To rule out that the reduced form results are not driven by idiosyncratic time variation, I estimate the cross-sectional version of 2 after aggregating the data across years. Table 5 and Table 6 represent this result for water-abundant and within water-abundant analysis. The results are qualitatively unchanged, with greater arsenic exposure in water-abundant districts as compared to water-scarce districts. With results on higher arsenic exposure in water-abundant districts being driven by "medium-thick" aquifers.

8 Discussion

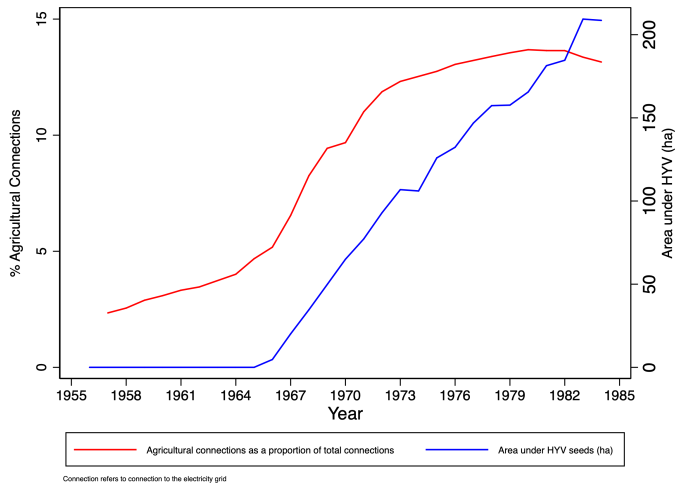
My paper documents a negative externality associated with irrigation practices. Specifically I ask if irrigation practices have the potential to generate arsenic contamination in groundwater across districts in India. Since timely irrigation was a critical factor in the diffusion of HYV seeds, I empirically test if greater expansion in area under HYV seeds are correlated with higher arsenic exposure. The findings of my paper point to unsustainable groundwater extraction as a plausible cause for arsenic contamination of groundwater in India over alternate mechanisms such as fertilizer or pesticide application.

These findings should be interpreted with certain caveats. First, the data used in my paper is in stark contrast to the fine grained and precisely measured variables as is common in the field of hydrology. Thus it is imperative that the qualitative aspect of the results are prioritized over the magnitude of effects. Second, hydrology presents multiple pathways through which arsenic contamination of groundwater may occur. The arguments presented in such papers are nuanced and use rich in situ and satellite data over specific spatial regions. While my paper is unable to rule out all potential pathways due to data limitations, I present compelling evidence in support of groundwater extraction as a plausible channel. Additionally, I rule out contemporary agricultural practices as a potential channel of contamination. Third, in the absence of water testing dates in the data, my paper makes a strong assumption on the orthogonality between water testing dates and testing site characteristics that may drive arsenic levels.

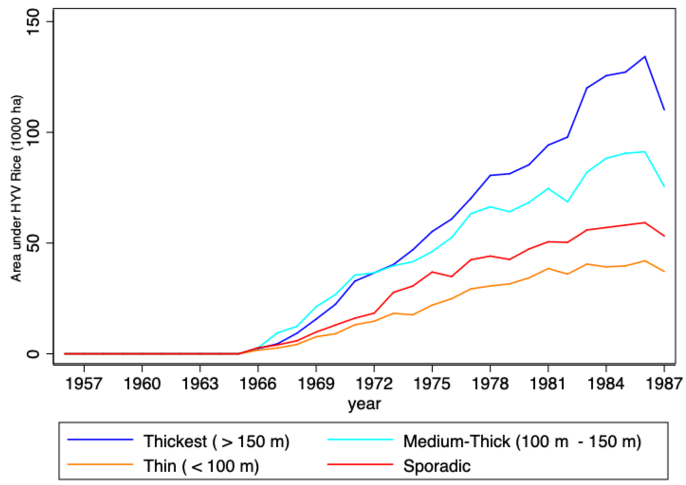
Nonetheless, the findings presented in my paper have policy implications for agricultural policy in India, specifically electricity subsidy for farm connections as well as the fundamentalism around staple food grains. The problem of groundwater over-exploitation in Indian agriculture has been driven by the availability of subsidized, often free power for irrigation (Bhushan et al., 2019). In states with very low tariff, there are limited incentives to use water efficiently which often results in cropping patterns that favor highly water intensive crops (Sarkar and Das, 2014). The problem is exacerbated by agricultural policy that perpetuates the historical and heavy bias towards staple grain production through a combination of input subsidies and price support. effectively crowding out production of traditional non-staple crops (Pingali, 2015). Non-staple crops such as pulses, millets, vegetables, legumes, oilseeds, medicinal plants are not only important sources of micro-nutrients, but are also crops that require lower levels of irrigation (Bhushan et al., 2019). However, a lack of price support and poorly developed market infrastructure come in the way of production system diversification (Pingali, 2015).

As the government of India transitions towards clean energy initiatives with emphasis on solar power (Bhushan et al., 2019), it is important to acknowledge that the fundamental problem of unsustainable groundwater extraction remains unaddressed. The findings of my paper highlight adverse consequences of groundwater extraction beyond water depletion and emphasize the need for complementary changes along both tariff rates and pivoting away from the fundamentalism of staple-food grains.

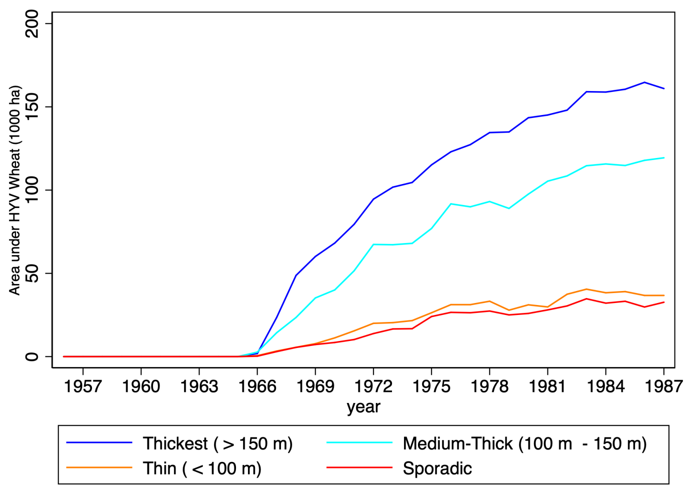
Figures



(a) Electric Connections and HYV Area Growth



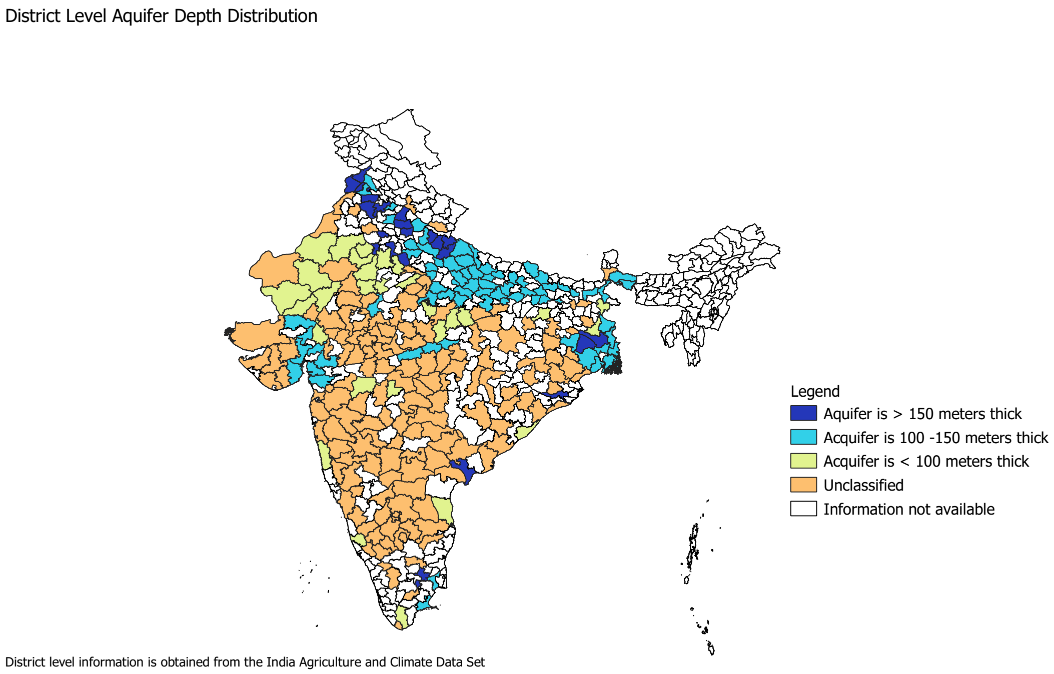
(b) Rice HYV Adoption Across Aquifer Types



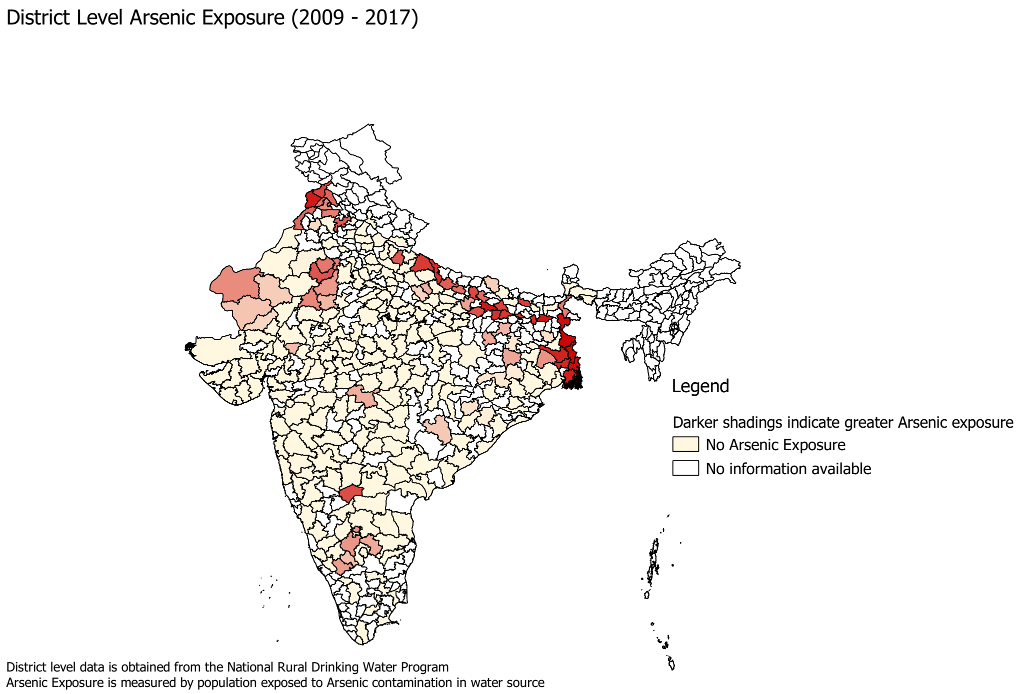
(c) Wheat HYV Adoption Across Aquifer Types

Figure 1: Groundwater Irrigation and the Green Revolution in India

Data for panel (a) are from the Public Electricity Supply - All India Statistics reports (electricity connections). The graph plots the average annual share of agricultural connections as part of the total electricity connections to the grid network. Panels (a), (b) and (c) use data from the IAC dataset for HYV adoption. Districts are categorized based on aquifer depth as "thickest" (>150 m), "medium-thick" (100–150 m), "thin" (< 100 m), and "sporadic". Average area under cultivation are plotted over time.



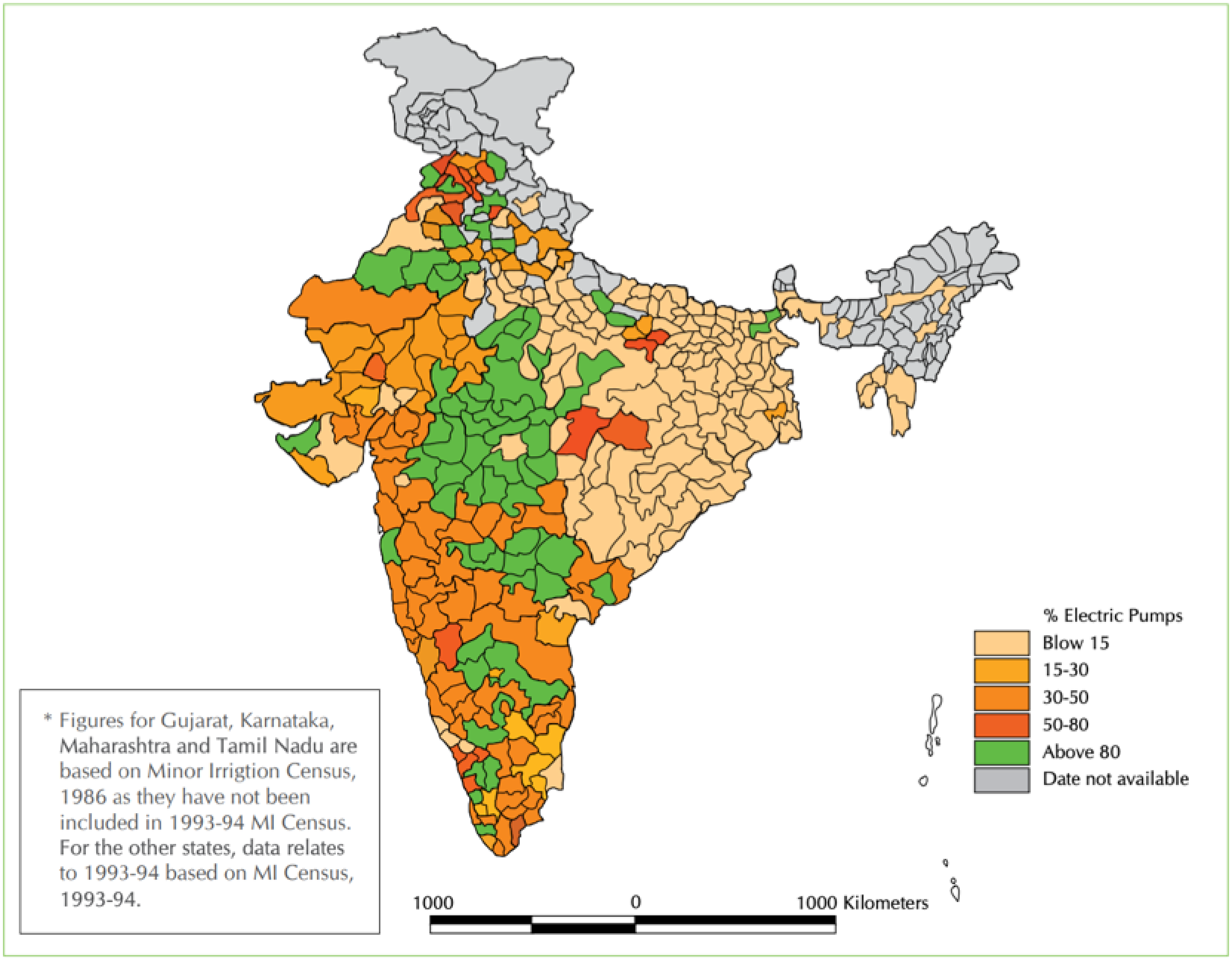
(a) Distribution of Aquifer Depth



(b) Population Exposure to Arsenic Contamination

Figure 3: Groundwater Resources and Arsenic Risk in India

Data for panel (a) are from the Agriculture and Climate in India dataset. Districts are shaded to indicate aquifer depth. Data for panels (b) comes from the National Rural Drinking Water Program. Arsenic exposure is measured by the proportion of population exposed to above safety level arsenic contamination in groundwater source.



Source: IWMI.

Figure 5: Energy Divide in Groundwater Resources

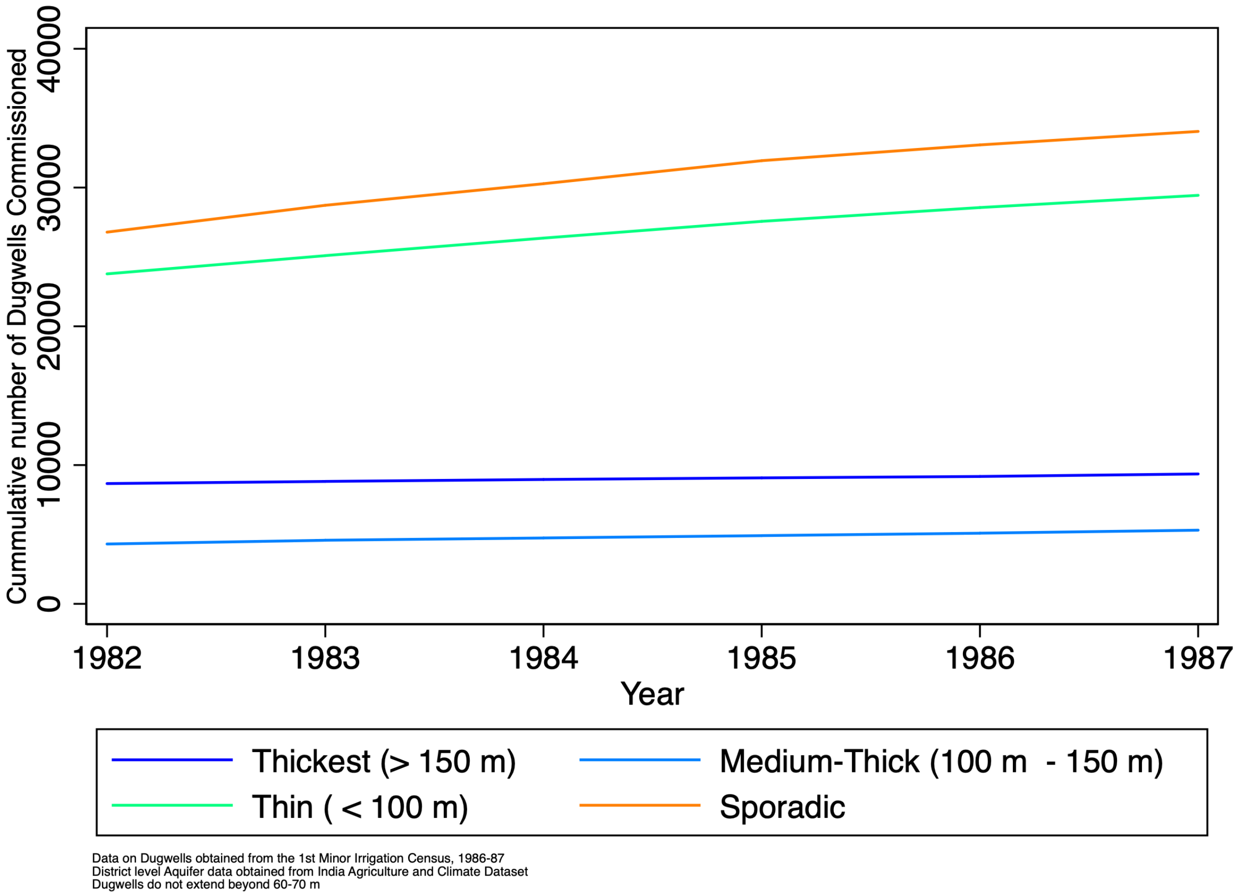


Figure 6: Distribution of Dugwells across Aquifer Depths

Data from the Minor Irrigation Census, 1987. Stock of Dugwells are measured at the district level. Using Agriculture and Climate in India dataset, districts are categorized based on aquifer depth as "thickest" (> 150 m), "medium-thick" (100 - 150 m), "thin" (< 100 m) and "sporadic". Average values across districts are plotted over time.

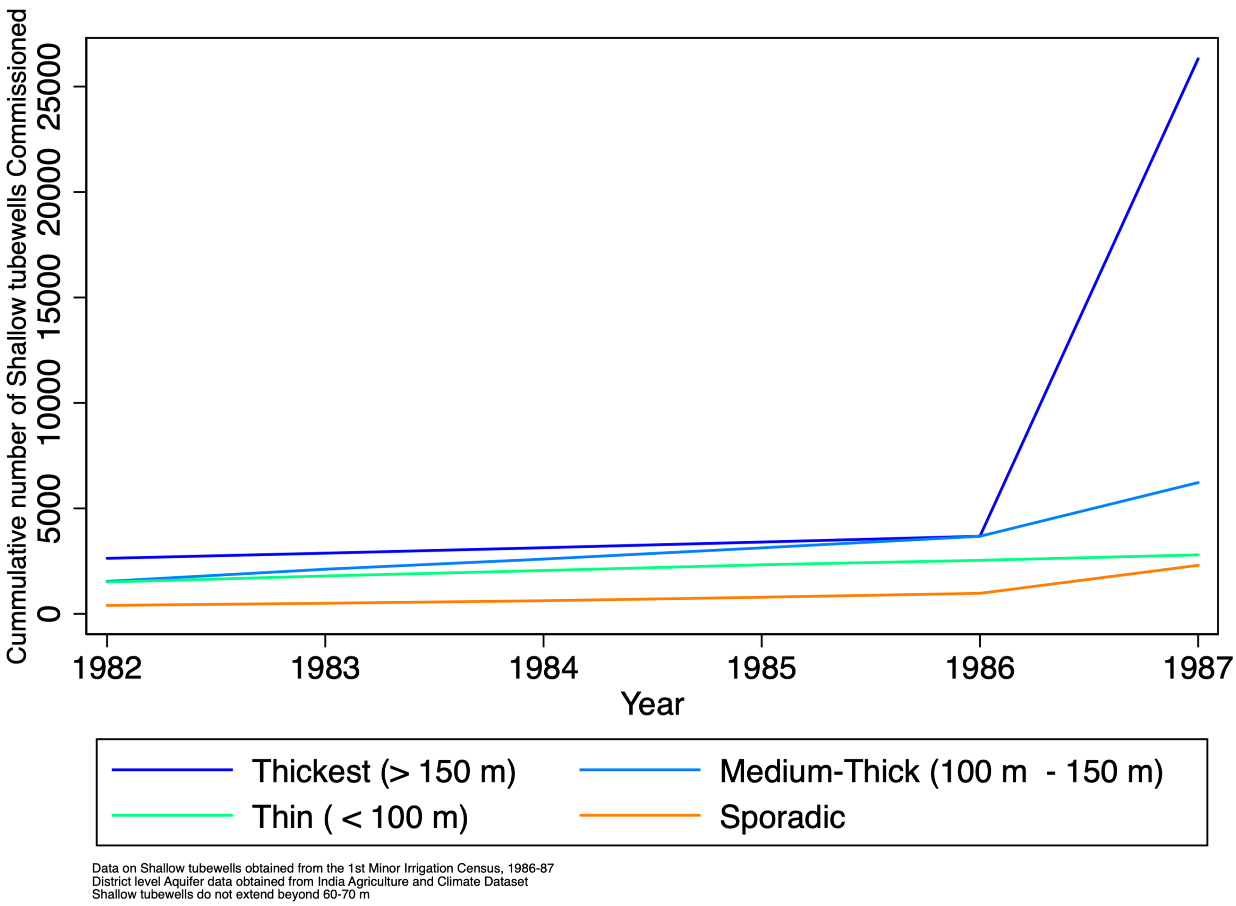


Figure 7: Distribution of Shallow tubewells across Aquifer Depths

Data from the Minor Irrigation Census, 1987. Stock of Shallow tubewells are measured at the district level. Using Agriculture and Climate in India dataset, districts are categorized based on aquifer depth as "thickest" (> 150 m), "medium-thick" (100 - 150 m), "thin" (< 100 m) and "sporadic". Average values across districts are plotted over time.

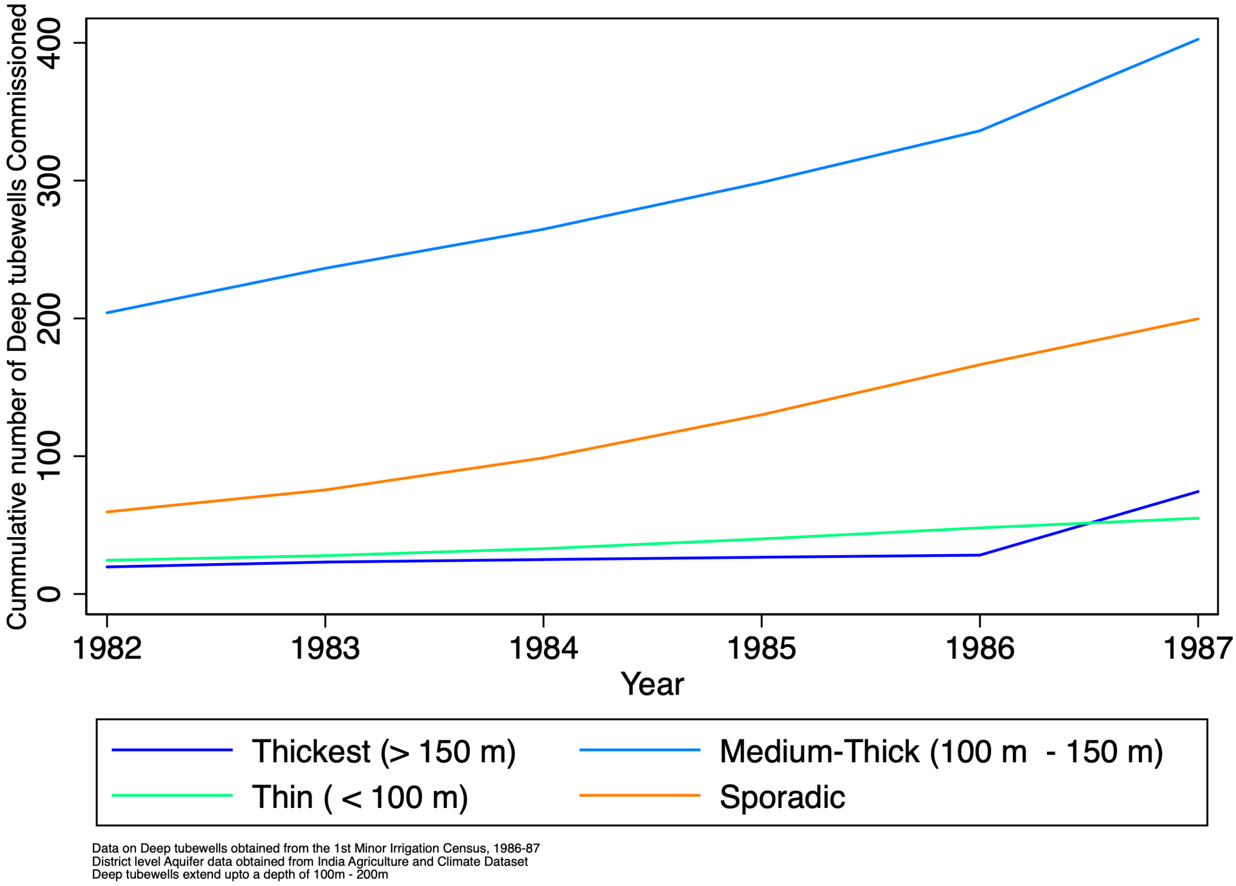


Figure 8: Distribution of Deep tubewells across Aquifer Depths

Data from the Minor Irrigation Census, 1987. Stock of Deep tubewells are measured at the district level. Using Agriculture and Climate in India dataset, districts are categorized based on aquifer depth as “thickest” ($>$ 150 m), “medium-thick” (100 - 150 m), “thin” ($<$ 100 m) and “sporadic”. Average values across districts are plotted over time.

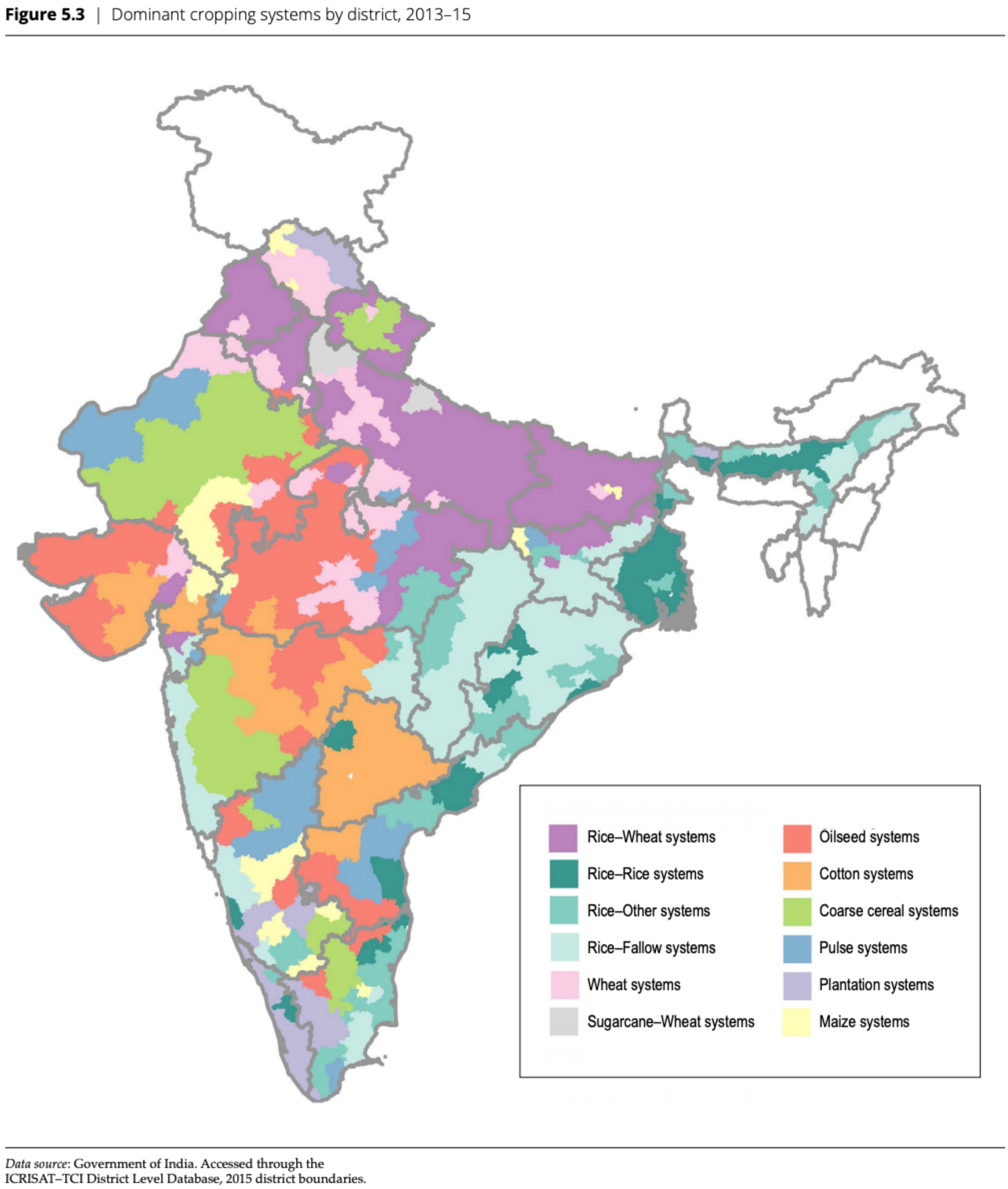
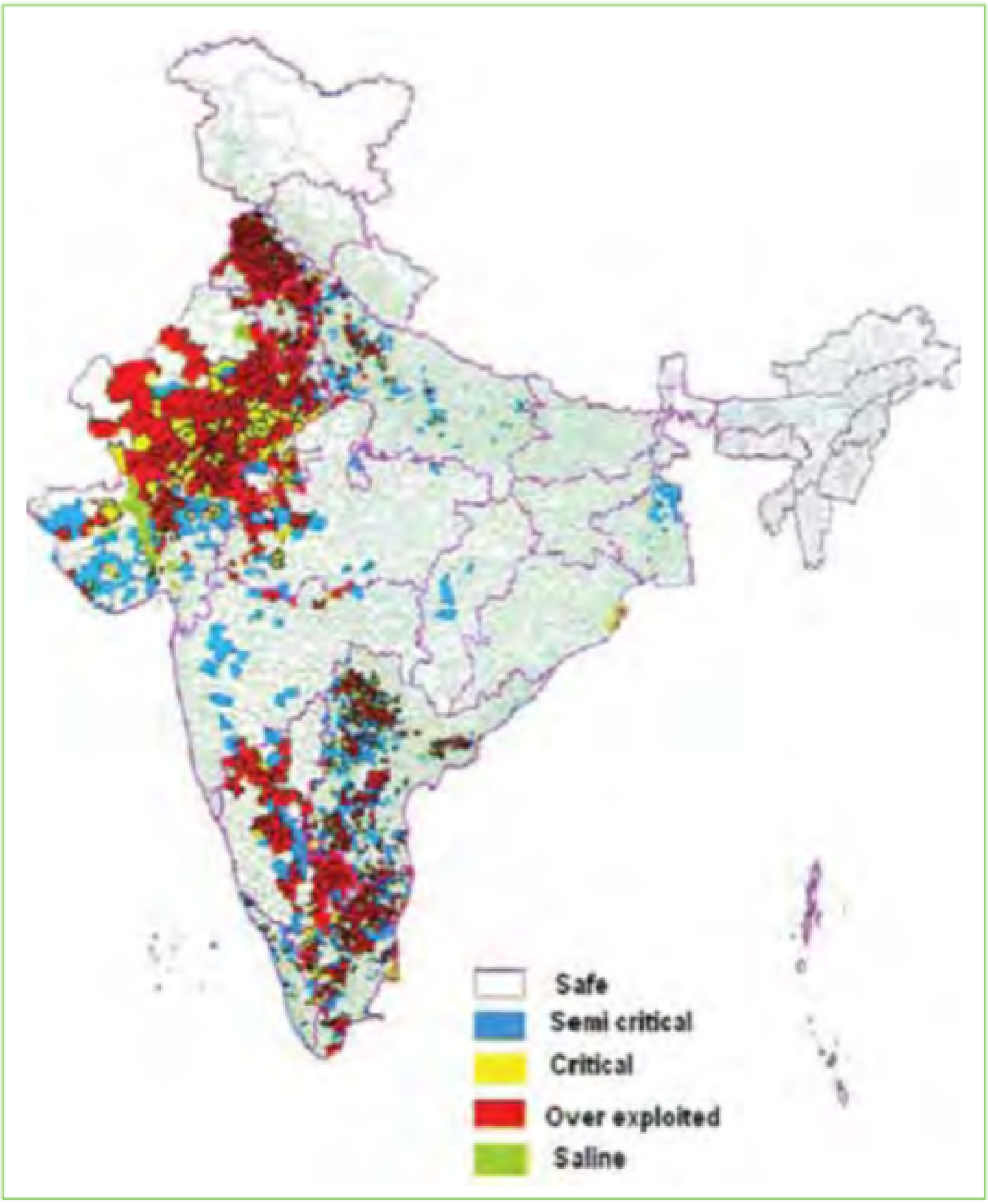


Figure 9: Dominant Cropping System in India



Source: IWMI.

Figure 10: Groundwater Stressed Blocks

Tables

	(1)
	Area under HYV (1000 ha)
Post X Thickest	105.86***
	(21.33)
Post X Mediuim-Thick	57.91***
	(13.81)
Post X Thin	-3.93
	(12.16)
Average area under HYV in Sporadic	93.69
Observations	8640
R^2	0.73

his table reports results from estimating equation (1) on a sample of all districts in the Agriculture and Climate in India dataset over 1957-1987. The outcome variable is a proxy for take up of the Green Revolution and is measured as the area under cultivation for HYV in the district and year . The variable *Post* is an indicator equal to one years after 1966. Thickest, Medium-Thick and Thin are aquifer thickness dummies for thickest (> 150m), medium-thick (100 m - 150m), and thin aquifer (< 100m) depths respectively. Controls include district rainfall and temperature. Regressions control for district fixed effects. Robust standard errors clustered at the district level are reported in parenthesis. * p < 0.10 , ** p < 0.05 , *** p < 0.01 .

Table 1: Adoption of High-Yield Variety

	Number of exposed Habitats			
	(1)	(2)	(3)	(4)
Water Rich (=1)	910***	755**	733**	757**
	(299)	(327)	(358)	(298)
Control Mean	1278	1278	1278	1278
Year Fixed Effect	x	x	x	x
Controls		x		x
State Fixed Effect			x	x
Observations	2407	2273	2407	2273
R^2	0.40	0.67	0.54	0.70

Notes: This table reports results from estimating equation 2 on a sample of 271 districts from the Agriculture and Climate in India dataset over the time period 2009-2017. The outcome variable captures arsenic exposure in a district year as the population weighted number of habitats with above safety threshold arsenic contamination. Water Abundant is a dummy if the district had either thickest or medium-thick aquifer. The comparison category are thin (water-sparse) districts. Thickest, Medium-Thick and Thin are aquifer thickness dummies for thickest (> 150m), medium-thick (100 m - 150m), and thin aquifer (< 100m) depths respectively. Controls include district-time varying characteristics including population, forest cover, access to electricity, economic activities, fertilizer usage as well as distance from river and canal irrigation. Robust standard errors clustered at the district level are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Arsenic affected habitats - Water Rich Districts

	Number of exposed Habitats			
	(1)	(2)	(3)	(4)
Thickest	-767***	219	-657***	-156
	(58)	(612)	(240)	(612)
Medium-Thick	943**	1221***	1015***	1055***
	(367)	(316)	(386)	(281)
Control Mean	1278	1278	1278	1278
Year Fixed Effect	x	x	x	x
Controls		x		x
State Fixed Effect			x	x
Observations	1116	1083	1116	1083
R^2	0.49	0.73	0.58	0.76

Notes: This table reports results from estimating equation 3 on a sample of 271 districts from the Agriculture and Climate in India dataset over the time period 2009-2017. The outcome variable captures arsenic exposure in a district year as the population weighted number of habitats with above safety threshold arsenic contamination. Thickest and Medium-Thick are aquifer thickness dummies for thickest (> 150m) and medium-thick (100 m - 150m)depths respectively. The reference category is "thin" (< 100 m) aquifer depth districts. Controls include district-time varying characteristics including population, forest cover, access to electricity, economic activities, fertilizer usage as well as distance from river and canal irrigation. Robust standard errors clustered at the district level are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Arsenic affected habitats

	(1) Number of Deep tubewells
Thickest	57.15
	(37.89)
Mediuim-Thick	107.81**
	(51.77)
Thin	-4.73
	(11.71)
Number of Deep tubewells in Sporadic	40.93
Observations	1626
R^2	0.10

Notes: This table reports results from estimating equation 4 on a sample of matched districts between Agriculture and Climate in India dataset and Minor Irrigation Census conducted in 1986-87. Data consists of a district year panel of districts across 1982 - 1986. The outcome variable measures the number of deep tubewells in a district. Thickest, Medium-Thick and Thin are aquifer thickness dummies for thickest (> 150m), medium-thick (100 m - 150m), and thin aquifer (< 100m) depths respectively. The reference category is "sporadic" aquifer depth. Robust standard errors clustered at the district level are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: High-extraction Irrigation Technology and Aquifer Depth

	Number of exposed Habitats	
	(1)	(2)
		reg2
Water Rich (=1)	635***	498***
	(130)	(162)
Control Mean	796	796
State Fixed Effect		x
Observations	271	271
R^2	0.22	0.57

Notes: This table reports results from estimating equation 2 on a sample of 271 districts from the Agriculture and Climate in India dataset over the time period 2009-2017. The outcome variable captures arsenic exposure in a district year as the population weighted number of habitats with above safety threshold arsenic contamination over 2009-2017. Water Abundant is a dummy if the district had either thickest or medium-thick aquifer. The comparison category are thin (water-sparse) districts. Thickest, Medium-Thick and Thin are aquifer thickness dummies for thickest ($>$ 150m), medium-thick (100 m - 150m), and thin aquifer ($<$ 100m) depths respectively. Controls include district-time varying characteristics including population, forest cover, access to electricity, economic activities, fertilizer usage as well as distance from river and canal irrigation. Robust standard errors clustered at the district level are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Average Arsenic Exposure - Water Rich Districts

	Number of Exposed Habitats	
	(1)	(2)
		reg2
Thickest	-249***	-170
	(14)	(111)
Medium-Thick	599***	666***
	(117)	(129)
Control Mean	796	796
State Fixed Effect		x
Observations	124	124
R^2	0.29	0.63

Notes: This table reports results from estimating equation 3 on a sample of 271 districts from the Agriculture and Climate in India dataset over the time period 2009-2017. TThe outcome variable captures arsenic exposure in a district year as the population weighted number of habitats with above safety threshold arsenic contamination over 2009-2017. Thickest, Medium-Thick and Thin are aquifer thickness dummies for thickest (> 150m), medium-thick (100 m - 150m), and thin aquifer (< 100m) depths respectively. The reference category is "sporadic" aquifer depth. Controls include district-time varying characteristics including population, forest cover, access to electricity, economic activities, fertilizer usage as well as distance from river and canal irrigation. Robust standard errors clustered at the district level are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Average Arsenic Exposure

	Proportion Total Affected (1)	Proportion Total Affected (2)	Proportion Total Affected (3)	Proportion Total Affected (4)
Nitrogen per ha	-0.000	-0.001	-0.000	-0.002
	(0.000)	(0.001)	(0.000)	(0.001)
Area under surface irrigation (1000ha)	-0.003	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Area under groundwater irrigation (1000ha)	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Gross Crop Area / Net Crop Area	-0.002	-0.004	0.031	0.030
	(0.381)	(0.384)	(0.374)	(0.377)
Phosphorous per ha		0.002		0.002
		(0.002)		(0.002)
Potassium per ha		0.004		0.005
		(0.006)		(0.007)
% hh with groundwater access			9.357**	10.564*
			(4.705)	(5.443)
Control Mean	.46	.46	.46	.46
Control SD	3.82	3.82	3.82	3.82
Observations	2378	2378	2378	2378
R^2	0.018	0.019	0.020	0.022

Notes: This table reports results from estimating a two-way fixed effects model with district and year fixed effects on a sample of 107 matched districts districts across NRDWP, IHDS, DHS and ICRISAT-TCI database. The outcome is the proportion of total population in a district with above safety threshold arsenic contamination over 2009-2017. The sample is constructed using data on 350 districts over 2009-2017. Annual district level data on groundwater access is created using data from the Indian Household Demographic Survey (IHDS) rounds 2005 and 2011 along with the 2016 Democratic and Health Survey. The outcome variable captures the arsenic exposure in a district as the average number of habitats with above safety threshold arsenic contamination over 2009-2017, obtained from NRDWP. Data on fertiliser, irrigation and cropped area comes from the ICRISAT-TCI dataset. Per ha refers to the per hectare application of fertilizer measured in kg. The variable % groundwater access is the proportion of households in a district with primary source of water consumption from tubewell, well or handpump. Robust standard errors clustered at the district level are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Proportion of Total Population Affected

	Proportion Total Affected (1)	Proportion Total Affected (2)	Proportion Total Affected (3)	Proportion Total Affected (4)
Nitrogen per ha	-0.000	-0.001	-0.000	-0.002
	(0.000)	(0.001)	(0.000)	(0.001)
Area under surface irrigation (1000ha)	-0.003	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Area under groundwater irrigation (1000ha)	0.001*	0.001	0.001	0.001
	(0.001)	(0.001)	(0.000)	(0.001)
Gangetic Plane X Area under groundwater irrigation (1000ha)	-0.001	-0.001	-0.000	-0.000
	(0.003)	(0.003)	(0.002)	(0.002)
Gross Crop Area / Net Crop Area	-0.010	-0.013	0.154	0.156
	(0.386)	(0.390)	(0.340)	(0.342)
Phosphorous per ha		0.002		0.003
		(0.002)		(0.002)
Potassium per ha		0.004		0.005
		(0.006)		(0.007)
% hh with groundwater access			1.937	3.048
			(1.372)	(1.962)
Gangetic Plane X % hh with groundwater access			13.570*	13.777*
			(8.148)	(8.190)
Control Mean	.46	.46	.46	.46
Control SD	3.82	3.82	3.82	3.82
Observations	2378	2378	2378	2378
R^2	0.018	0.019	0.025	0.026

Notes: Data consists of arsenic affected population from the NRDWP website. Data on fertiliser, irrigation and cropped area comes from the ICRISAT-TCI website. Standard errors are clustered at the district level. Model includes district and year fixed effects. The outcome variable measures the district level arsenic affected individuals as a proportion of total district population. The variable % groundwater access is calculated under the set III definition. I combine data from 2005 and 2011 rounds of the IHDS as well as the 2015 DHS survey to create measures of groundwater access by households. Groundwater access is defined as a dummy which equals one if the household's primary source of water consumption comprises of tubewell, well or handpump. Data comes from a sub sample of 107 matched districts across IHDS, DHS, Census 2011 and NRDWP arsenic contamination habitats, that belong to the Gangetic Plane.

Table 8: Proportion of Total Population Affected : Gangetic Plane

References

[1] Ahmad, S. A., Khan, M. H., and Haque, M. (2018). Arsenic contamination in groundwater in bangladesh: implications and challenges for healthcare policy. *Risk management and healthcare policy*, 11:251.

[2] Argos, M., Kalra, T., Rathouz, P. J., Chen, Y., Pierce, B., Parvez, F., Islam, T., Ahmed, A., Rakibuz-Zaman, M., Hasan, R., et al. (2010). Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in bangladesh (heals): a prospective cohort study. *The Lancet*, 376(9737):252–258.

[3] Badiani, R. and Jessoe, K. K. (2011). Electricity subsidies for agriculture: Evaluating the impact and persistence of these subsidies in india.

[4] Bhushan, C., Bhati, P., Singh, M., and Jhawar, P. (2019). Silver bullet: Redesigning solar pump programme for water and energy security.

[5] Brainerd, E. and Menon, N. (2014). Seasonal effects of water quality: The hidden costs of the green revolution to infant and child health in india. *Journal of Development Economics*, 107:49–64.

[6] Carson, R. T., Koundouri, P., and Nauges, C. (2011). Arsenic mitigation in bangladesh: a household labor market approach. *American Journal of Agricultural Economics*, 93(2):407–414.

[7] Chakraborty, M., Mukherjee, A., and Ahmed, K. M. (2015). A review of groundwater arsenic in the bengal basin, bangladesh and india: from source to sink. *Current Pollution Reports*, 1:220–247.

[8] Chowdhury, S. and Singh, P. (2021). Information campaign on arsenic poisoning: Unintended consequences in marriage market.

[9] D'Agostino, A. (2017). Technical change and gender wage inequality: long-run effects of india's green revolution. *Available at SSRN 3400889*.

[10] Erban, L. E., Gorelick, S. M., Zebker, H. A., and Fendorf, S. (2013). Release of arsenic to deep groundwater in the mekong delta, vietnam, linked to pumping-induced land subsidence. *Proceedings of the National Academy of Sciences*, 110(34):13751–13756.

[11] Evenson, R. E. and Gollin, D. (2003). Assessing the impact of the green revolution, 1960 to 2000. *science*, 300(5620):758–762.

[12] Fendorf, S., Michael, H. A., and van Geen, A. (2010). Spatial and temporal variations of groundwater arsenic in south and southeast asia. *Science*, 328(5982):1123–1127.

[13] Glaeser, E. L., Kerr, S. P., and Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics*, 97(2):498–520.

[14] Gollin, D., Hansen, C. W., and Wingender, A. M. (2021). Two blades of grass: The impact of the green revolution. *Journal of Political Economy*, 129(8):2344–2384.

[15] Hagerty, N. and Tiwari, A. (2022). Industrial water pollution and agricultural production in india.

[16] Hornbeck, R. and 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.

[17] Liu, J. and Sekhri, S. (2021). Groundwater accessibility and firm behavior. Technical report, Working Paper.

[18] Michaels, G. (2011). The long term consequences of resource-based specialisation. *The Economic Journal*, 121(551):31–57.

[19] Mukherjee, A., Sengupta, M. K., Hossain, M. A., Ahamed, S., Das, B., Nayak, B., Lodh, D., Rahman, M. M., and Chakraborti, D. (2006). Arsenic contamination in groundwater: a global perspective with emphasis on the asian scenario. *Journal of Health, Population and Nutrition*, pages 142–163.

[20] Pingali, P. (1999). Sustaining rice-wheat production systems: Socio-economic and policy issues. Technical report.

[21] Pingali, P. (2015). Agricultural policy and nutrition outcomes–getting beyond the preoccupation with staple grains. *Food security*, 7:583–591.

[22] Pitt, M. M., Rosenzweig, M. R., and Hassan, M. N. (2021). Identifying the costs of a public health success: Arsenic well water contamination and productivity in bangladesh. *The Review of Economic Studies*, 88(5):2479–2526.

[23] Podgorski, J. and Berg, M. (2020). Global threat of arsenic in groundwater. *Science*, 368(6493):845–850.

[24] Rahman, M., Sohel, N., Yunus, F. M., Alam, N., Nahar, Q., Streatfield, P. K., and Yunus, M. (2019). Arsenic exposure and young adult's mortality risk: A 13-year follow-up study in matlab, bangladesh. *Environment international*, 123:358–367.

[25] Ryan, N. and Sudarshan, A. (2022). Rationing the commons. *Journal of Political Economy*, 130(1):210–257.

[26] Sarkar, A. and Das, A. (2014). Groundwater irrigation-electricity-crop diversification nexus in punjab: Trends, turning points, and policy initiatives. *Economic and Political Weekly*, pages 64–73.

[27] Sayre, S. S. and Taraz, V. (2019). Groundwater depletion in india: Social losses from costly well deepening. *Journal of Environmental Economics and Management*, 93:85–100.

[28] Sekhri, S. (2014). Wells, water, and welfare: the impact of access to groundwater on rural poverty and conflict. *American Economic Journal: Applied Economics*, 6(3):76–102.

[29] Sekhri, S. and Shastry, G. K. (2024). The curse of plenty: The green revolution and the rise in chronic disease. *American Economic Journal: Applied Economics*, 16(4):1–33.

[30] Shaji, E., Santosh, M., Sarath, K., Prakash, P., Deepchand, V., and Divya, B. (2021). Arsenic contamination of groundwater: A global synopsis with focus on the indian peninsula. *Geoscience frontiers*, 12(3):101079.

[31] Sohel, N., Persson, L. Å., Rahman, M., Streatfield, P. K., Yunus, M., Ekström, E.-C., and Vahter, M. (2009). Arsenic in drinking water and adult mortality: a population-based cohort study in rural bangladesh. *Epidemiology*, pages 824–830.

[32] Zaveri, E., Grogan, D. S., Fisher-Vanden, K., Frolking, S., Lammers, R. B., Wrenn, D. H., Prusevich, A., and Nicholas, R. E. (2016). Invisible water, visible impact: groundwater use and indian agriculture under climate change. *Environmental Research Letters*, 11(8):084005.

A Balance Table

	(1) Thickest mean	(2) Medium-Thick mean	(3) Thin mean
Area under surface irrigation (1000 ha)	49.71	53.09	53.23
Area under groundwater irrigation (1000 ha)	157.55	140.98	167.52
Nitrogen per gca (Kg/ha)	141.18	107.66	59.89
Cropping Intensity	1.71	1.59	1.40
Rural Population	1933.48	47604.25	1653.66
Number of Households	618.07	39981.69	447.97
Power supply for Agriculture (0/1)	0.87	0.68	0.81
Sum of Forest cover (0-100) of all pixels in district	1442.94	41922.14	1677.08
Maximum Forest Cover (0-100) in unit	17.43	14.61	12.16
Average light luminosity (total light/ num cells)	10.95	5.63	5.47
Observations	198	657	261

Notes: Data consists of variables from SHRUG data-set as well as ICRISAT-TCI database.

Table 9: Covariate Summary Across Aquifer Depths

B HYV adoption for Maize and Bajra

The following graphs plot area under cultivation for HYV seeds for Maize (11) and (12).

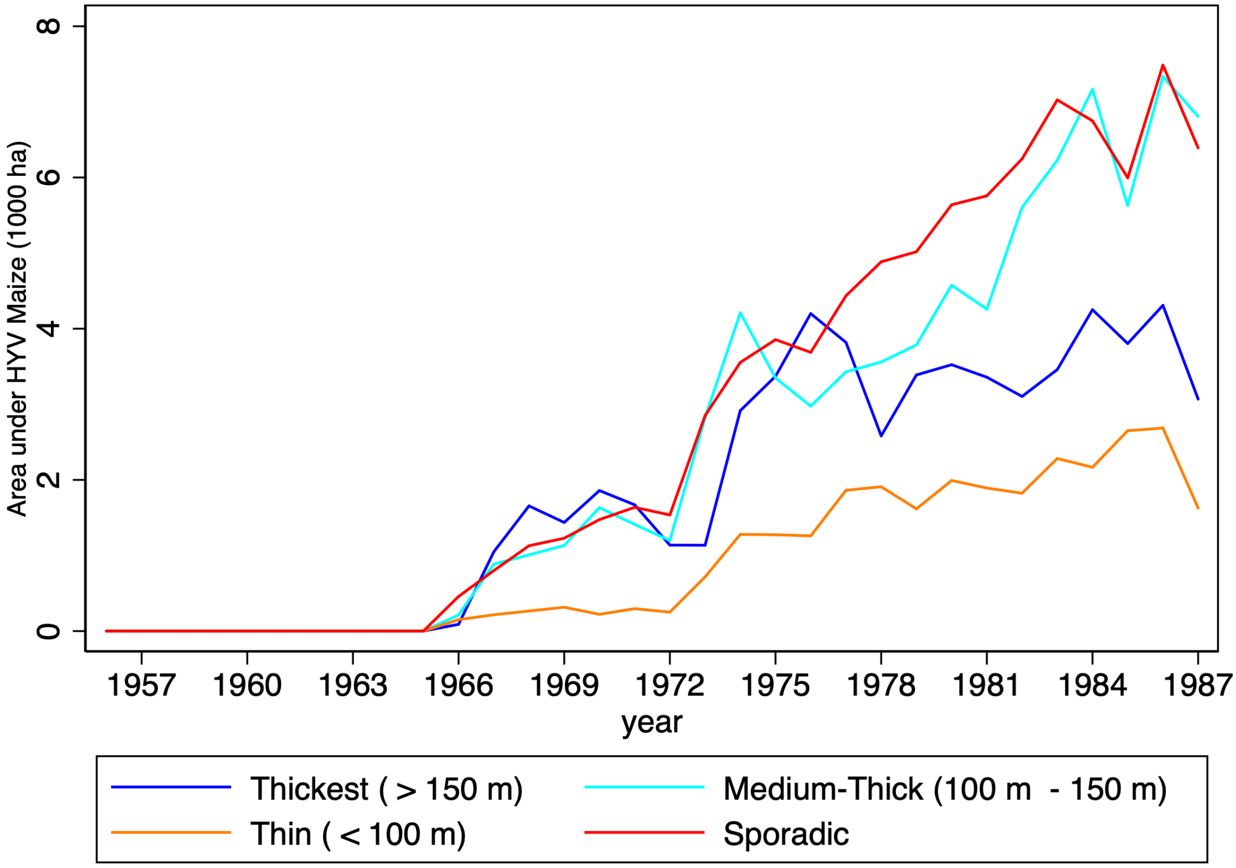


Figure 11: Trends in Maize HYV Adoption

Data from Agriculture and Climate in India dataset. Area under cultivation is measured at the district level. Districts are categorized based on aquifer depth as “thickest” (150 m), ”medium-thick” (100 - 150 m), “thick” (100 m) and “sporadic”. Average values across districts are plotted over time

C MI Data Details

This section provides greater detail on different irrigation technologies captured under the Minor Irrigation Census.

Dugwells are ordinary open wells of varying dimension which are dug or sunk from the ground surface ito water bearing stratum to extract water for irrigation purposes. These are broadly known as Masonary wells, kuccha wells and dug-cum-bore wells. All such schemes are of private nature belonging to individual cultivator.

Private shallow tubewells consist of a bore hole built into the ground with the purpose of tapping fround water from previous zones. In sedimentary formations depth of a shallow tubewell does not exceed 60-70 mts. These tubewells are either cavity tubewells or strainer tubewells. These are usually drilled by percussion method using hand boring sets and sometimes percussion rigs. Success and popularity of the scheme depends on how cheap they are. Coir structures form by binding coir strings over a iron frame is being used as strainer. In shallow water table areas, bamboo frames are also used. SOmetimes steel pipe, casing are replaced by popes constructed by rapping bituminised gunny bags over the bamboo frame. These are called bore wells, in which bore hole is stable without a lining in the bottom portion and a tube is inserted only in the upper zone.

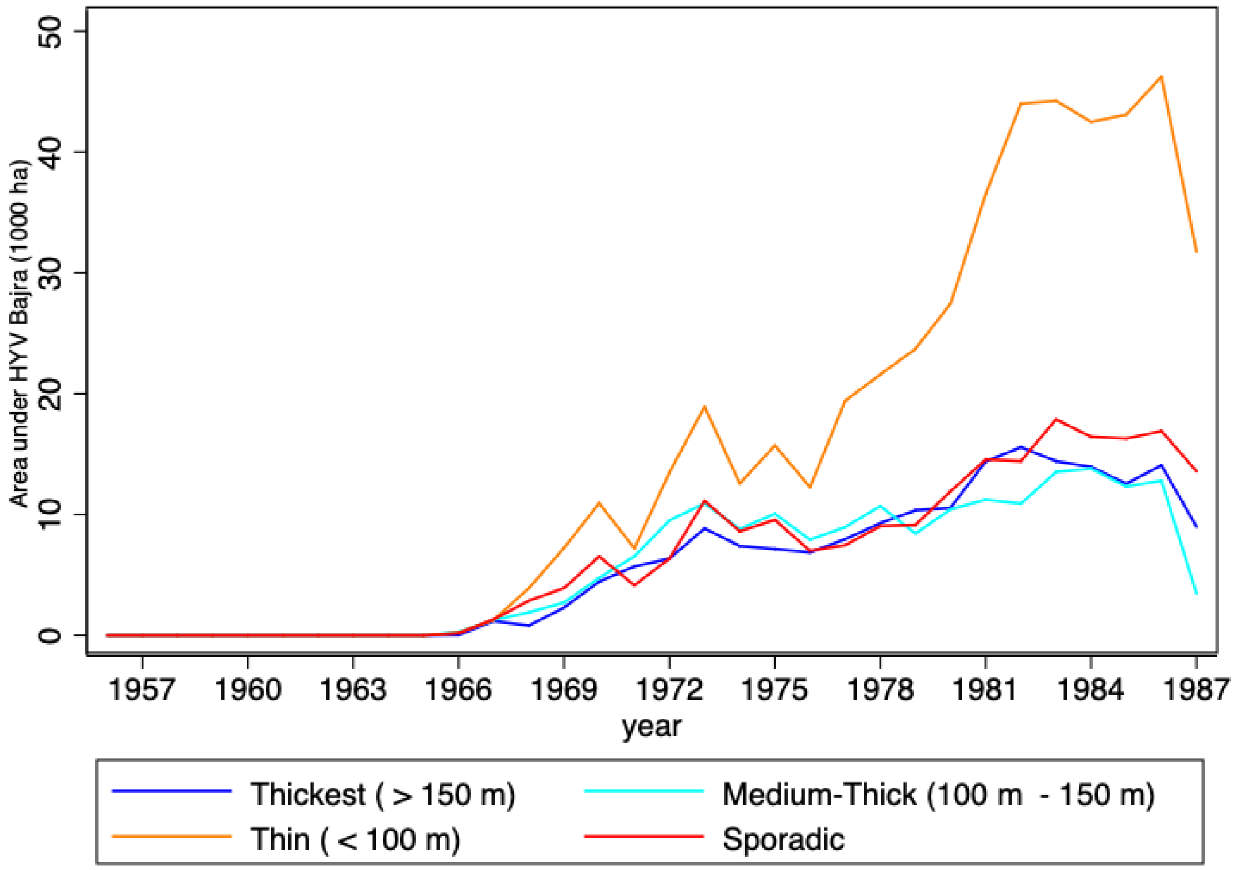


Figure 12: Trends in Bajra HYV Adoption

Data from Agriculture and Climate in India dataset. Area under cultivation is measured at the district level. Districts are categorized based on aquifer depth as “thickest” (150 m), ”medium-thick” (100 - 150 m), “thick” (100 m) and “sporadic”. Average values across districts are plotted over time

Deep Tubewells usually extend to the depth of 100 meters and more and are designed to give a discharge of 100 to 200 cubic meter per hour. These are drilled by rotary percussion or rotary cum percussion rigs. These tubewells operate round the clock during the irrigation season, depending upon the availability of power.

D Agricultural Intensification and Arsenic Contamination

This analysis combines data from the ICRISAT-TCI database and the NRDWP district level arsenic affected data. The main outcome variable is number of individuals affected by arsenic contamination scaled by 2011 census population figures. The initial data matches for 80% (556 districts) of the data. The ICRISAT data does not cover few north-eastern states and so I am unable to match about 18% (126 districts) primarily across the states of Nagaland, Meghalaya, Mizoram, Sikkim and Tripura. Data for Jammu and Kashmir is also not covered in the ICRISAT database. The arsenic data extends from 2009-2017, whereas the ICRISAT data is more historical and up to 2017 - thus providing 8 years of overlap. Unfortunately the pump density data is only till the time period 1992-2009.

I control for source of groundwater consumption by households by combining multiple survey instruments (IHDS 2005, 2011 & DHS 2015) to create district level measure of proportion of households with groundwater as primary source of consumption. I perform the analysis for the subset of districts that match across all sources of data (351 districts).

Groundwater access by households for self consumption is available for 3 years 2005, 2011,2015-16. To impute values for the entire sample, I use the following method which I refer to as Set III. Impute values for missing years by calculating the trend coefficient from a district level regression as follows

$$y_{it} = \beta_0 + i.dist\_code + \beta year + \epsilon_{it}$$

I use β from the above equation to calculate the predicted groundwater for each of the years in the sample using 2005 as the base year : $(1+\beta)^y$, where y is the difference between current year and 2005.

Tables 10 & 11 present results from the estimation strategy discussed above on arsenic exposure as a proportion of rural population.

	Proportion Rural Affected (1)	Proportion Rural Affected (2)	Proportion Rural Affected (3)	Proportion Rural Affected (4)
Nitrogen per ha	0.000	-0.002	0.000	-0.002
	(0.000)	(0.001)	(0.000)	(0.002)
Area under surface irrigation (1000ha)	-0.005	-0.005	-0.005	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)
Area under groundwater irrigation (1000ha)	0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Gross Crop Area / Net Crop Area	0.047	0.042	0.083	0.082
	(0.486)	(0.491)	(0.478)	(0.482)
Phosphorous per ha		0.004		0.004
		(0.003)		(0.003)
Potassium per ha		0.006		0.008
		(0.008)		(0.009)
% hh with groundwater access			10.451*	12.334*
			(5.788)	(6.732)
Control Mean	.46	.46	.46	.46
Control SD	3.82	3.82	3.82	3.82
Observations	2378	2378	2378	2378
R^2	0.019	0.020	0.020	0.022

Notes: Data consists of arsenic affected population from the NRDWP website. Data on fertiliser, irrigation and cropped area comes from the ICRISAT-TCI website. Standard errors are clustered at the district level. Model includes district and year fixed effects. The outcome variable measures the district level arsenic affected individuals as a proportion of total rural district population. The variable % groundwater access is calculated under the set III definition. I combine data from 2005 and 2011 rounds of the IHDS as well as the 2015 DHS survey to create measures of groundwater access by households. Groundwater access is defined as a dummy which equals one if the household's primary source of water consumption comprises of tubewell, well or handpump. Data comes from 350 matched districts across IHDS, DHS, Census 2011 and NRDWP arsenic contamination habitats.

Table 10: Proportion of Rural Population Affected

	Proportion Rural Affected (1)	Proportion Rural Affected (2)	Proportion Rural Affected (3)	Proportion Rural Affected (4)
Nitrogen per ha	0.000	-0.002	-0.000	-0.003
	(0.000)	(0.001)	(0.001)	(0.002)
Area under surface irrigation (1000ha)	-0.005	-0.005	-0.006*	-0.006*
	(0.003)	(0.003)	(0.003)	(0.003)
Area under groundwater irrigation (1000ha)	0.002*	0.002	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Gangetic Plane X Area under groundwater irrigation (1000ha)	-0.001	-0.001	0.001	0.001
	(0.004)	(0.004)	(0.003)	(0.003)
Gross Crop Area / Net Crop Area	0.043	0.037	0.221	0.224
	(0.488)	(0.493)	(0.443)	(0.445)
Phosphorous per ha		0.004		0.005
		(0.003)		(0.003)
Potassium per ha		0.006		0.008
		(0.008)		(0.009)
% hh with groundwater access			2.466	4.194
			(2.099)	(2.731)
Gangetic Plane X % hh with groundwater access			14.645	14.953
			(10.751)	(10.734)
Control Mean	.46	.46	.46	.46
Control SD	3.82	3.82	3.82	3.82
Observations	2378	2378	2378	2378
R^2	0.019	0.020	0.023	0.025

Notes: Data consists of arsenic affected population from the NRDWP website. Data on fertiliser, irrigation and cropped area comes from the ICRISAT-TCI website. Standard errors are clustered at the district level. Model includes district and year fixed effects. The outcome variable measures the district level arsenic affected individuals as a proportion of total rural district population. The variable % groundwater access is calculated under the set III definition. I combine data from 2005 and 2011 rounds of the IHDS as well as the 2015 DHS survey to create measures of groundwater access by households. Groundwater access is defined as a dummy which equals one if the household's primary source of water consumption comprises of tubewell, well or handpump. Data comes from a sub sample of 107 matched districts across IHDS, DHS, Census 2011 and NRDWP arsenic contamination habitats, that belong to the Gangetic Plane.

Table 11: Proportion of Rural Population Affected : Gangetic Plane

The following graphs demonstrate contemporary trends in irrigation across surface irrigation, groundwater irrigation, electric and diesel pump-set usage in Figures 13, 14, 15 and 16 respectively.

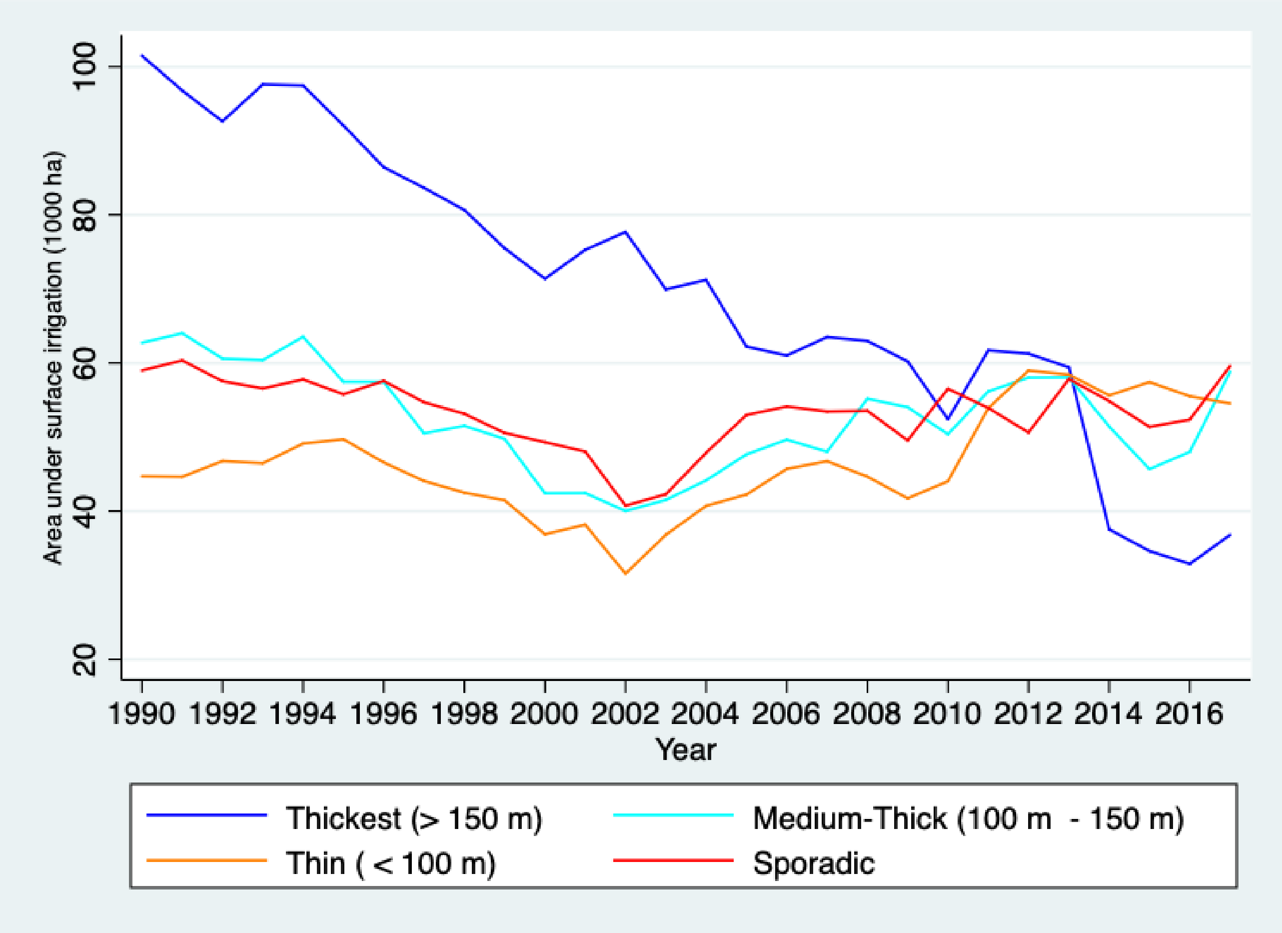


Figure 13: Trends in Surface Irrigated Area

Data from ICRISAT database. Area under Surface Irrigation is aggregated at the District level. Average values are plotted over time.Districts are categorized as "thickest" (150 m), "medium-thick" (100 - 150 m), "thick" (100 m) and "sporadic".

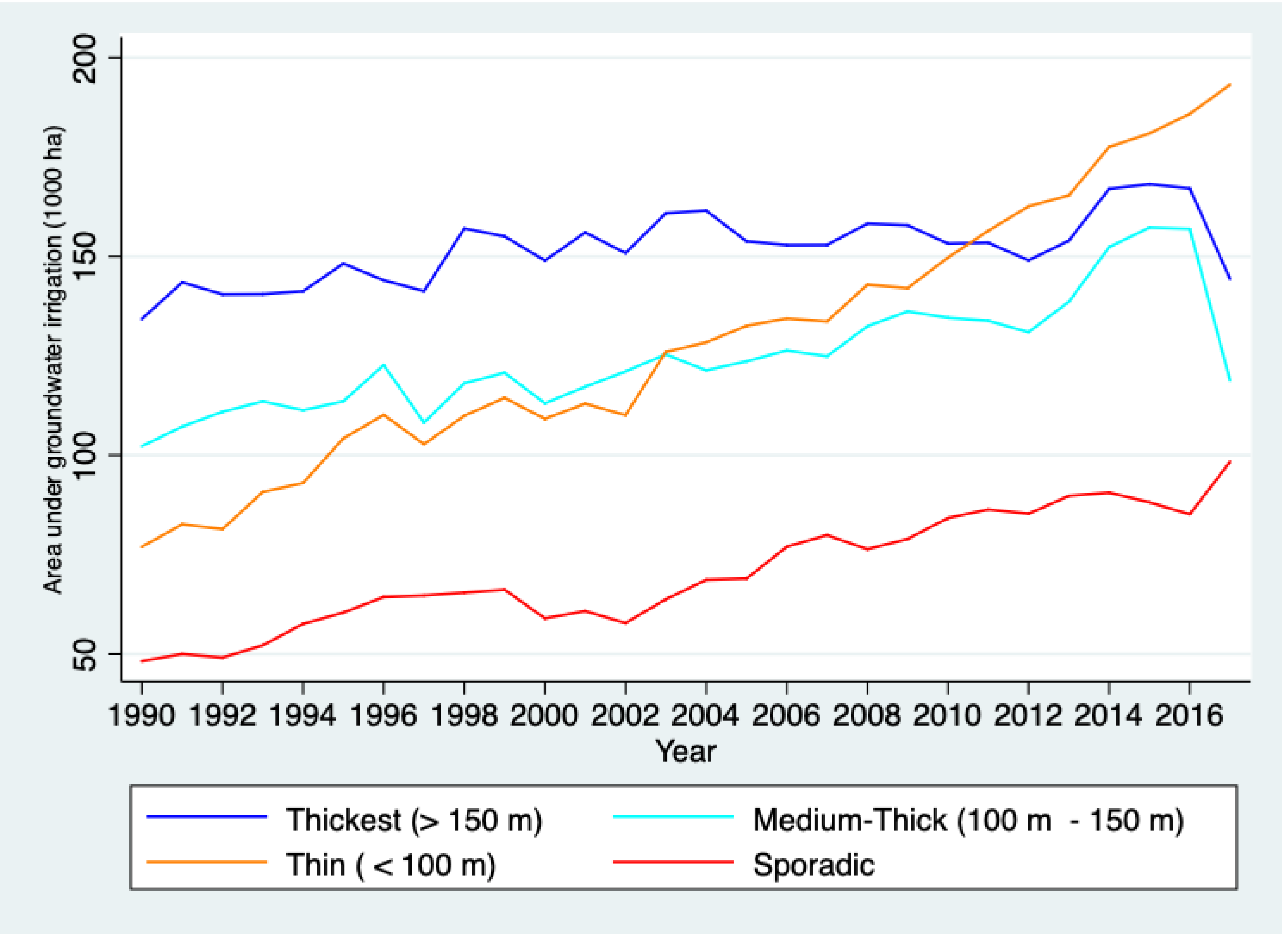


Figure 14: Trends in Groundwater Irrigated Area

Data from ICRISAT database. Area under Groundwater Irrigation is aggregated at the District level. Average values are plotted over time.Districts are categorized as "thickest" (150 m), "medium-thick" (100 - 150 m), "thick" (100 m) and "sporadic".

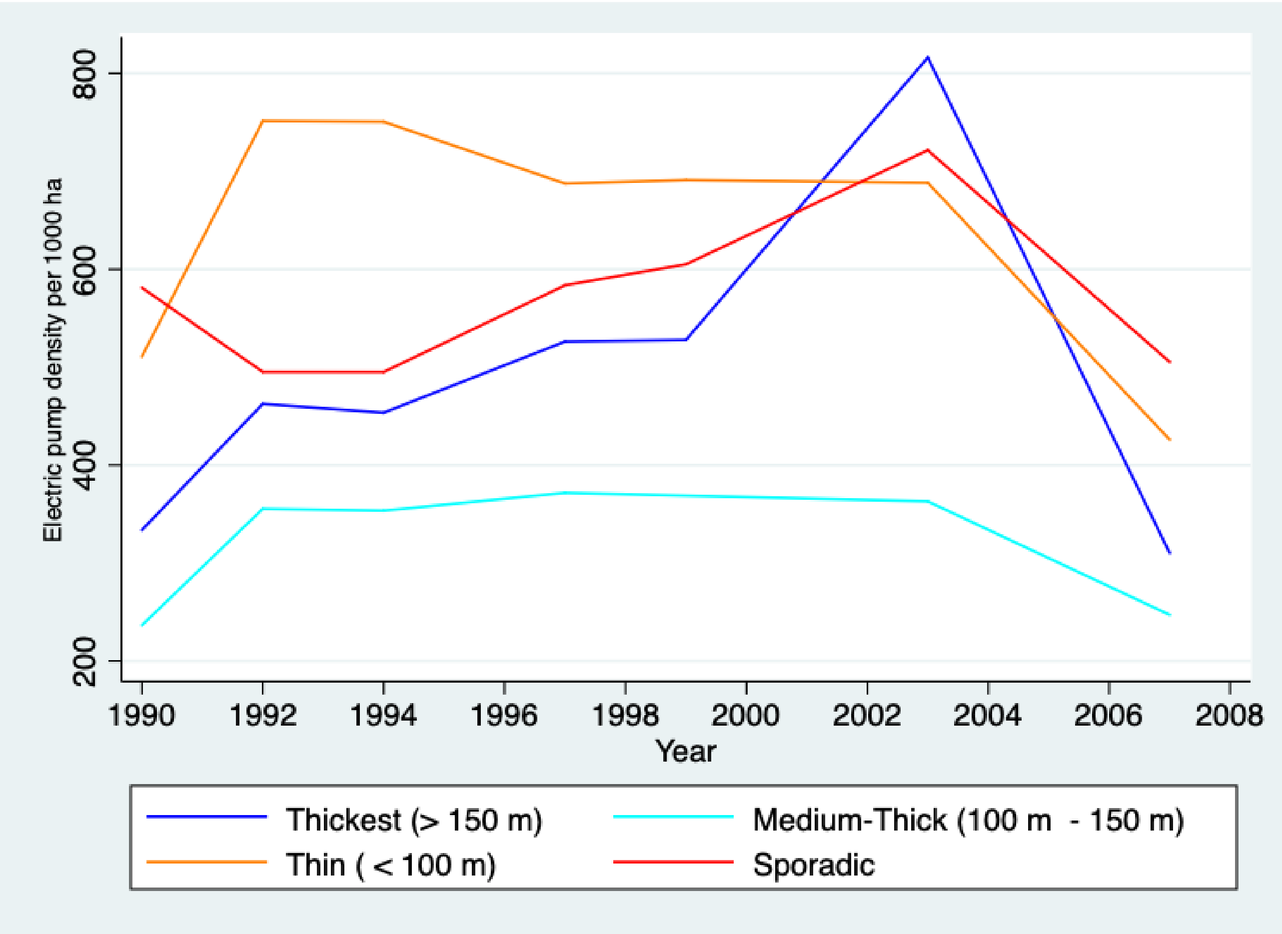


Figure 15: Trends in Electric Pump-set Usage

Data from ICRISAT database. Density of Electric pump-set is aggregated at the District level. Average values are plotted over time.Districts are categorized as “thickest” (150 m), ”medium-thick” (100 - 150 m), “thick” (100 m) and “sporadic”.

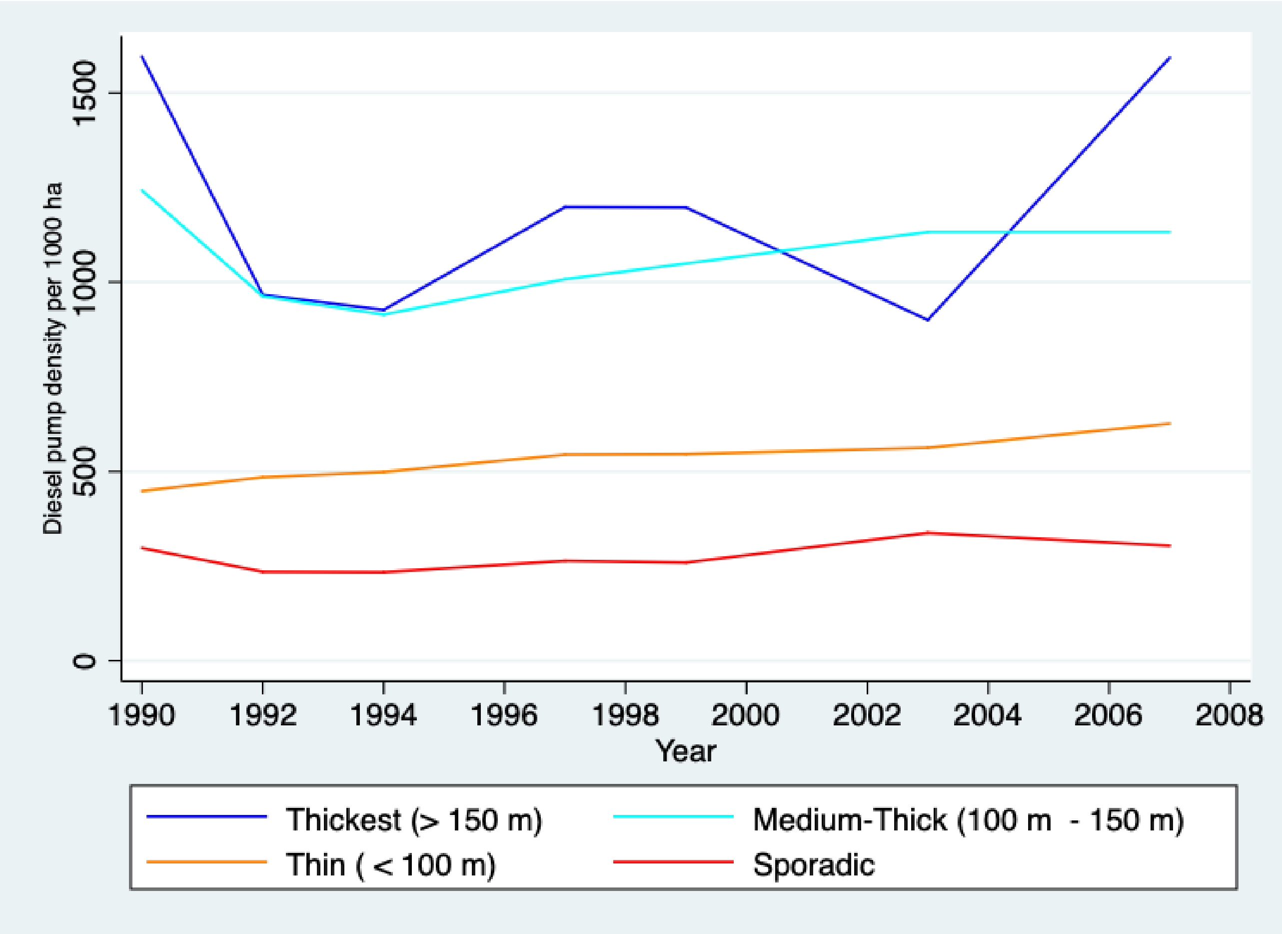


Figure 16: Trends in Diesel Pump-set Usage

Data from ICRISAT database. Density of Diesel pump-set is aggregated at the District level. Average values are plotted over time.Districts are categorized as “thickest” (150 m), ”medium-thick” (100 - 150 m), “thick” (100 m) and “sporadic”.