Left out in crowds: how spatial segregation, aquifer characteristics, and mechanized pumping energy shape drinking water access in rural India

Sudatta Ray[1,2] and Isabella Gupta[1]

[1]Department of Geography, National University of Singapore, Singapore

[2]Environmental Studies, Division of Social Science, Yale-NUS College, Singapore

Inequality in accessing public resources is a persisting problem among marginalized communities globally. In developing countries, the lack of access extends to even the most basic services such as reliable drinking water supply and domestic electricity. Literature investigating these abiding challenges largely focus on social rather than biophysical factors. We find both contribute towards the poor outcomes of the most marginalized community in India – the Scheduled Tribes (ST). Using high-resolution, interdisciplinary data across seven Indian states, we find ST populations to be the most spatially concentrated within villages and, in at least one state, across low transmissivity aquifers. Nearly 50% of ST population lived in villages where they made up more than 80% of the village population. In contrast, less than 3% of other marginalized communities were in similarly concentrated villages. We find villages with high ST concentrations to suffer an average loss of 20% households with machine-pumped water, that instead, source drinking water from uncovered wells. Uncovered wells suffer from low productivity, and high seasonal fluctuations. ST-concentrated villages in Odisha are also more numerous in low transmissivity aquifers, where high energy pumping is required to offset deep drawdowns. Whereas we find ST villages to be further constrained by limited access to electricity suffering an average loss of 18% in electrified households. Future policies need to consider biophysical elements that affect public resource access and identify areas of overlap with the spatial distribution of marginalized populations. These overlaps, while creating unforeseen poverty traps for marginalized communities, could also offer opportunities of targeting public resources and compounding their positive impacts.

Introduction

Safe and reliable drinking water supply is one of the greatest abiding challenges of the rural developing world. In 2022, between 8%-42% of rural population on average lacked access to even basic drinking water services across Central and South Asia, North, West and Sub-Saharan Africa [1].[1] Over 90% of all rural drinking water was based on groundwater sources in the same regions [1]. Sufficiency of water volumes however is not guaranteed by the mere presence of water infrastructure, latter being the most often recorded data. Groundwater levels can fluctuate by many meters at any given location across seasons, hydrogeologies and with changes in precipitation. Climate change induced uncertainty in rainfall is likely to further increase groundwater fluctuations and future stocks [2–4]. Therefore, continuous monitoring of existing water infrastructure is needed to ascertain true extent of coverage today, and to plan for sufficient water supply in the future. Groundwater based drinking water supply requires two key inputs - groundwater resource and pumping energy to treat and transport it. Yet, with a few exceptions, drinking water literature rarely includes analyses of either, largely focusing on the distribution of, and access to water infrastructure [5, 6]. More broadly, lack of access to basic public resources is a persistent problem among marginalised communities globally. Competition among diverse ethnic groups (ethnic diversity) and more recently, localized concentrations of certain marginalized groups (spatial segregation) have been identified as two key factors contributing to inequities in access to resources across different geographical contexts [7–11]. We add to this literature by studying high concentrations of marginalized communities within villages and their

[1]Basic services include an improved source and a roundtrip collection time of less than 30 minutes including queuing [1]

association with the distribution of drinking water infrastructure and electricity in rural India [12].

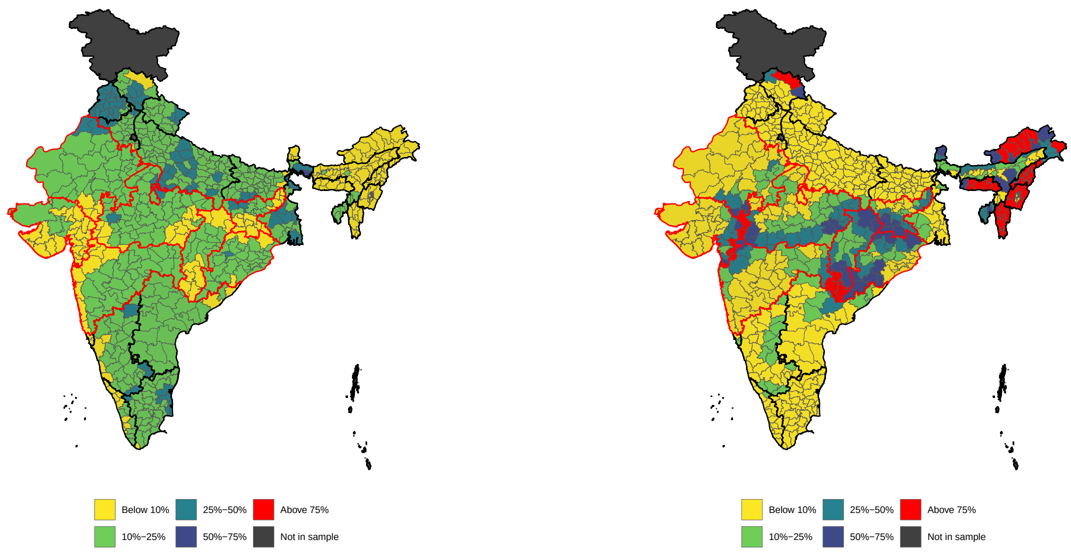
We create a novel dataset by stitching together socioeconomic and biophysical data at the village-level. Such granular and interdisciplinary data allows us to make three main contributions - first, we identify the association between high concentrations of marginalized populations and the distribution of drinking water infrastructure. Unlike existing analyses conducted at district and state levels, our village-level analysis reveals the extreme concentration of Scheduled Tribe (ST) communities within villages, which we use to identify and measure the association of population concentration with access to drinking water structures. ST populations are among the most marginalized in India, and yet, are rarely the foci of even literature studying the role of marginalization in public resource distribution. Second, we analyse the role of aquifer characteristics in determining the suitability of existing drinking water infrastructure for serving current domestic demands. Although aquifer characteristics are among the greatest determinants of pumped water volumes, they are understudied in the context of drinking water supply [5, 12]. On the other hand, government policies and targets rarely consider the variation in the types, scales and costs of drinking water infrastructure required to serve populations residing across different types of aquifers. We find ST populations in one sample state to not only be concentrated within villages, but also in hydrogeologies that experience deeper drawdowns making groundwater extraction energy intensive, which brings us to our third major contribution on the important role of mechanized pumping energy in drinking water access. Besides groundwater resource, pumping energy is the other essential input required in the provision of reliable and safe drinking water supply. While rural drinking water literature largely overlooks the role of pumping energy, literature on rural electrification finds recently constructed electricity infrastructure to be incapable of serving high power needs required for water pumping [13–15]. We differentiate the various drinking water structures based on pumping energy needs and study the role of rural electrification in shaping their respective distributions. Our results therefore highlight the need to consider biophysical elements in addition to socioeconomic characteristics in public goods allocation, especially while examining drinking water access.

Scheduled tribe communities are among the most marginalized and least studied in India

The impact of ethnic divisions and concentration of marginalized population on the distribution of public resources has been widely studied across different geographic contexts including parts of Africa, the United States, India, Brazil, and Indonesia [7, 8, 10, 16, 17]. In India, it has largely been studied for two major marginalized communities of Scheduled Castes (SC) and Scheduled Tribes (ST), and to a smaller extent for the Muslim community.[2] However, treatment of SC and ST communities differs widely in this literature. While STs are often acknowledged as the most marginalized, they are largely lumped with SCs, used as a comparison group representative of 'weakest' political representation, or left out entirely. On the other hand, the same studies treat SCs with nuanced and varied lenses of numerical dominance, sub-caste divisions and neighborhood-level segregation [11, 16, 25–30]. This contrasting treatment likely stems from lack of granular data combined with the relative geographic isolation of STs near forested areas making it hard to identify them at broad geographic levels [9, 16, 27, 29]. Outside the north-eastern Indian states, where almost entire populations are tribal, ST households comprise 10% or more of the total population only in seven states spread across the central belt comprising of Chhattisgarh, Gujarat, Jharkhand, Maharashtra, Madhya Pradesh, Odisha and Rajasthan. SC populations on the other hand are more evenly spread throughout the country (figure 1). We restrict our analyses to these seven states with sizable tribal populations and focus on village outcomes to isolate the associations with ST concentration.

Although the geographical concentration of ST households is apparent at the state and district-levels (figure 1(b)), employing granular village-level data, we find that in fact concentration of ST population stems from

[2]Scheduled Castes or dalits occupy the lowest rung of the Hindu Varna (caste) system which stratifies society into hierarchical categories. Whereas, Scheduled Tribes is a term first coined by the British colonists to identify indigenous people living in relative isolation near forested land [18, 19]. There is rich literature challenging distinction between the two groups, and the relative isolation of Scheduled Tribes [20, 21]. Whereas, the strong association of Scheduled Tribes with forested land, and their subsequent dispossession, and the latter's despoliation is widely recognized and deeply studied [22–24]



(a) (b)

Figure 1: District-level distribution of (a) Scheduled Caste and (b) Scheduled Tribe population across India. Thin black lines represent district boundaries, thick black lines represent state boundaries. and thick red lines indicate the seven states in our analysis sample. Data are based on Census 2011, taken from SHRUG [31].

village-levels. Figure 2 illustrates the distribution of ST and SC populations across villages in varying proportions, where x-axis denotes village-level and y-axis state-level population proportions of STs and SCs. Across the seven states, the greatest proportion of STs reside in villages where they constitute 90% or greater of the village populations. Whereas, majority of the SC populations reside in villages where they make up 10%-20% of village populations. The village-level concentration of ST population has important implications for access to public resources, since villages are the smallest unit of nearly all government programmes in India [32]. For instance, drinking water and rural electrification, the two public resources included in our analyses, began with village based targets before expanding to universal household coverage. Even today, despite the household-level targets of electricity access and drinking water supply, the physical infrastructure responsible for their respective deliveries is still largely village-wide i.e. transformers for electricity and handpumps or wells for water. Additionally, the decentralised administrative structure of India gives rise to the possibility of political capture at different administrative levels [33–35]. Therefore understanding the varying degree of distributional inequities across districts, sub-districts and villages is a key component of our analysis.

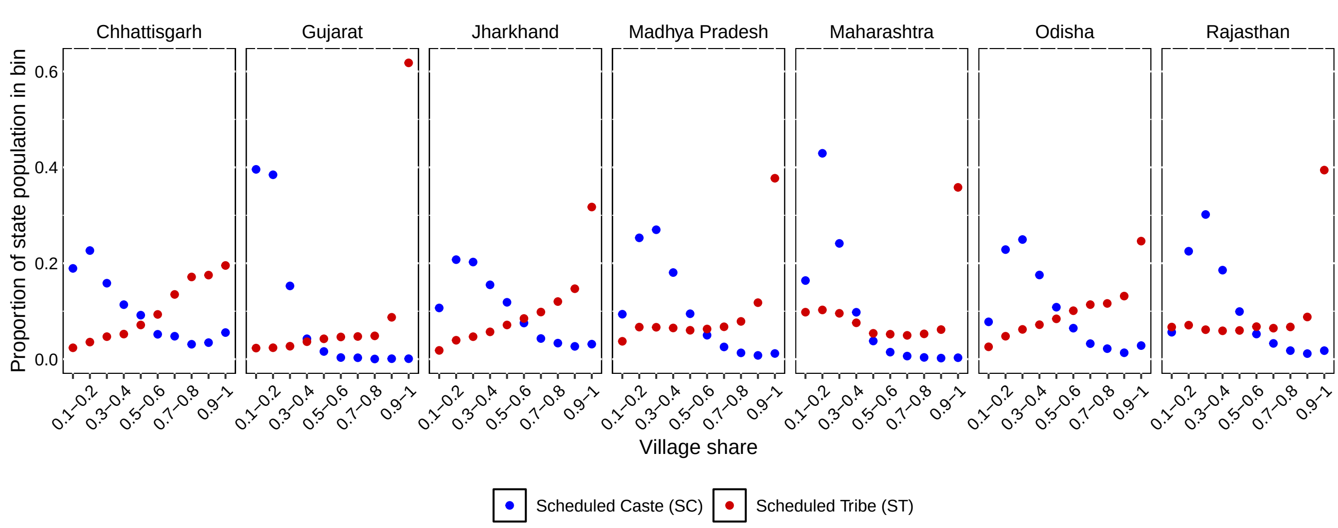


Figure 2: Distribution of Scheduled Caste (SC) and Scheduled Tribe (ST) village-level populations in 2011 for the seven states included in the sample. X-axis represents deciles of villages population shares of SC and ST. Y-axis represents the proportion of the state population in the respective village share bin. Data taken from SHRUG [31] (n=235,444 villages).

We treat villages with concentrations of ST population of at least 80% as ST villages and those with at least 20% SC population as SC villages. Overall, approximately 20% of villages are identified as ST and 26% as SC. We drop 0.1% of villages which are identified as both ST and SC villages from our sample.

Moving targets and persisting inequities in drinking water coverage

The Accelerated Rural Water Supply Programme (ARWSP) introduced in the early 1970s provided a set of targets for the roll-out of drinking water infrastructure across rural India. The programme was the first to introduce numerical targets, which included 40l of drinking water per capita per day and access to a minimum of one handpump or a public tap for every 250 people at a distance of no more than 1.6km [36]. The ARWSP has gone through many iterations since its introduction with two lasting impacts - first the structure of the numerical targets has remained the same with progressive changes made to the volume, source and distance criteria. For instance, a 2013 version mandates 55l per capita per day requirement via piped water supply systems within 100m of the household (or 30 minutes of fetching time) [37]. Second, handpumps and to a lesser extent, public taps have remained the water structures of choice for the expansion [37]. Even in the most recent census conducted in 2011, handpumps were the most common source of drinking water in the seven states in our sample, with an average of 43% households using handpumps as their primary drinking water source (figure 3(a)). The remaining water structures include uncovered and covered wells, taps, tubewells, and just over 5% of households cumulatively relied on surface water based sources such as springs, rivers, lakes etc. Jal Jeevan Mission (JJM) launched in 2019 is the most recent iteration of the rural drinking water policy. It aims to equip all houses with piped water connections supplying a daily minimum of 55l per capita. Although impressive progress has been reported by the programme (see JJM dashboard), an independent evaluation is pending [38]. Whereas, JJM data have been contested for their accuracy, found to be inconsistent with other datasets such as the Mission Antyodaya survey, and as such, are not included in our analyses (see figure A1 for a summarized comparison) [39–41].

The different groundwater-sourced water structures are largely based on two types of wells - dug wells and bore wells. Bore wells are constructed using drilling machines and can be as deep as required for accessing water to the order of thousands of meters (Karnataka's deep drinking water wells are one such example [42]). Handpumps and tubewells require bore wells, but greatly differ in water productivity. Handpumps relying on manual pumping are functional only up to 50m on average. Whereas mechanized pumping is essential in tubewells. Dug wells are constructed using manual or animal labour, and are therefore, limited to depths of 12m on average [43]. Uncovered and covered wells are largely based on dug wells (see table A1 for complete description of the various water structures). Finally, taps represent some degree of ambiguity as they largely require mechanized pumping but can also be gravity-fed.

Water structures and whether they are based on mechanized pumping influence rates of extraction and thereby drinking water coverage. We examine the distribution of, and estimate the extraction volumes from handpumps, uncovered wells, taps and tubewells. We assume handpumps and uncovered wells to be based on manual modes of extraction, and taps and tubewells on mechanized pumping. In our main results, we refer to taps and tubewells cumulatively as machine-pumped water sources.

Besides depths and the presence of pumping energy, productivities of water structures are also influenced by the aquifer characteristics of the inter-laying rocks between the land surface and groundwater sources [44]. There is growing recognition of their role in determining suitability of groundwater based drinking water supply, particularly in hard-rock aquifers which cover nearly two-thirds of the country's landmass [45–47]. Yet aquifer characteristics are rarely considered in research that investigates and informs the distribution of drinking water access in India. Indian drinking water literature largely focuses on the distribution of tap water (despite its relative low presence), and the role of caste and religion in determining the same [48–51]. We find a distinct socioeconomic pattern to the distribution of water structures in our sample of states which motivates our analysis (figure 3(a)). While close to 27% of households on average reported taps and tubewells as their primary source of drinking water, the proportion was less than half of that at approximately 11% among ST villages on average. Instead, over a quarter of households on average relied on uncovered wells for their drinking water supply in ST villages.

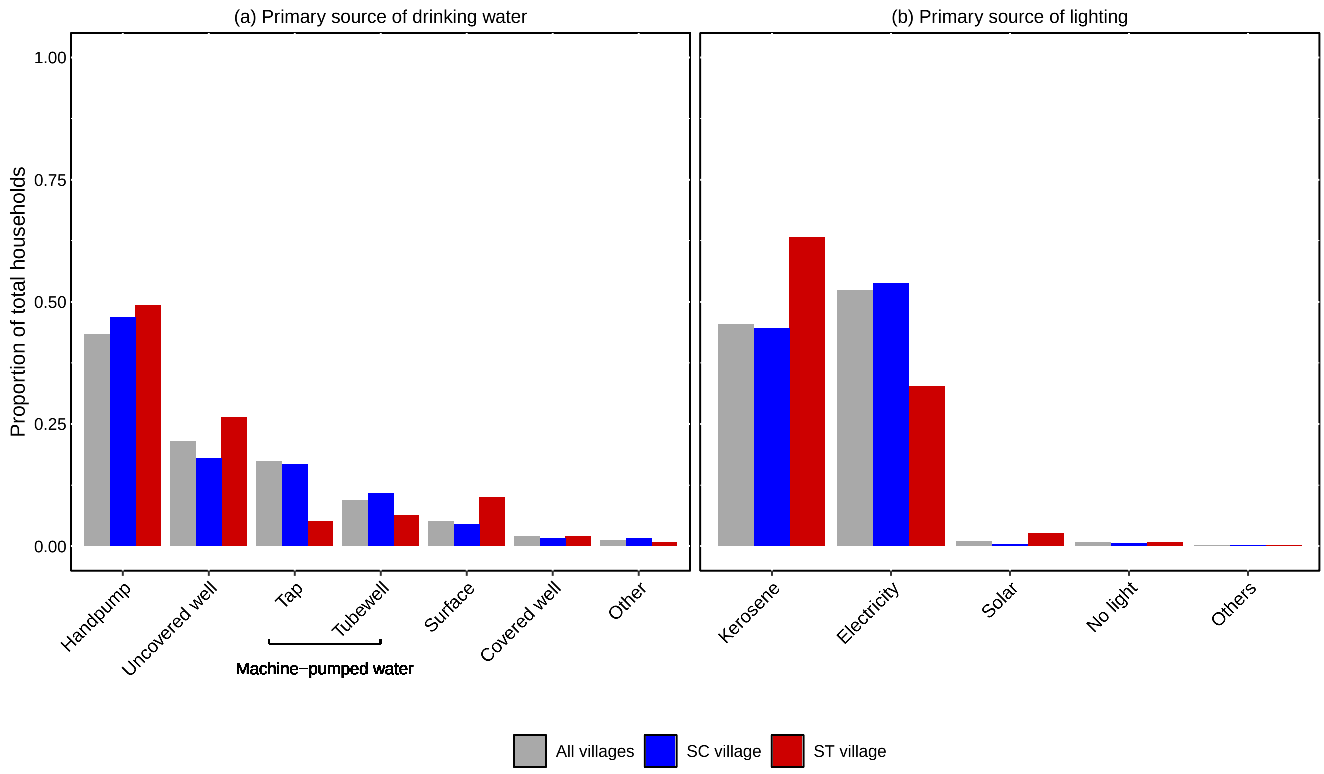


Figure 3: Distribution of drinking water and lighting sources across SC villages (≥ 20% SC) ST villages (≥ 80% ST), and all villages. Y-axis represents the average proportion of households in a village reporting the respective source as their primary source of (a) drinking water or (b) lighting. Data are based on Census 2011 (n=235,444) [52], full details can be found in the Methods section.

Pumping energy: the missing piece in drinking water supply

Pumping energy is a key component of groundwater abstraction, evidenced by the vast energy-groundwater irrigation nexus literature in India [53–59]. Yet, there is scarce mention of energy in the rural drinking water supply literature despite the fact that over 85% of all drinking water sources in rural India are based on groundwater [60]. The relatively low prevalence of machine-pumped water sources among existing drinking water structures could be a possible explanation. Mechanized pumping is required to operate in only two of the six categories of drinking water structures recorded by the population census of India (figure 3(a)). In reality, pumps can be, and are often used across all types of drinking water structures. They reduce drudgery, ensure consistent flow, and are essential for piped water supply. Additionally, pumping energy is essential for village-based drinking water structures only in regions where water depths exceed those accessible by handpumps of approximately 50m (UNICEF [61]). Water at deeper levels requires submersible pumps for extraction which can only be run on electricity.

On the face of it, electricity is an unlikely constraint in drinking water supply due to the tremendous expansion of rural electrification across India in the last decade (see *Saubhagya* dashboard). However, the focus of rural electrification has been largely low power household-based supply since the early 2000s with limited expansion in high capacity uses such as submersible pumps [13, 15]. As a result, adequacy of current water supply in newly electrified regions with deep water levels remains doubtful in the absence of additional electric infrastructure upgrades. Key policy documents on rural drinking water too fail to explicitly mention pumping energy as an essential input [37, 62].

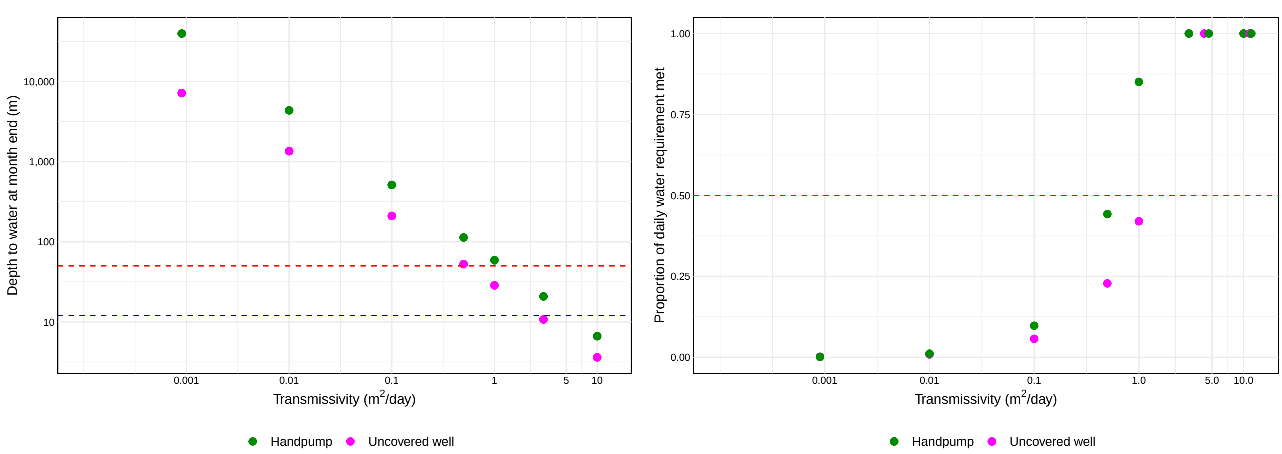
An important manifestation of this oversight is in recorded data - we found no government collected data on energy used in drinking water structures. In its stead, we use household electrification as a proxy for the presence of electricity infrastructure across villages to study the role of electricity in drinking water provision. Household electrification is a weak proxy as it does not ensure the presence of electricity infrastructure required to power submersible pumps, but it is the only reliable data on electricity available at the village-level. On average, a little over half of all households reported using electricity as their primary light source, the remaining largely used kerosene. The average proportion of households using electricity was less than a third among ST villages (figure 3 (b)). We also use satellite-based nightlight data as a supplementary measure of electrification. We test whether drinking water demand has high mechanized pumping needs among geographically concentrated ST populations, and whether, the relative absence of electricity explains the greater presence of shallow uncovered wells in ST dominated villages.

Results

Hydrogeologies play an important role in determining drinking water availability

Productivity of groundwater structures can vary widely depending upon aquifer characteristics, permeability, groundwater stock, presence of faults or fractures, and depth to water in addition to water infrastructure related characteristics [44, 63, 64]. Despite the near ubiquitous presence of groundwater-based drinking water infrastructure across Indian rural households, the actual availability of water is likely to vary dramatically across space and time. Transmissivity, an aquifer characteristic governing drawdowns is among the most important in determining pumped volumes, and therefore, also in predicting long-term sustainability of drinking water infrastructure [12, 44, 64, 65]. Drawdowns are temporary drops in water-levels caused by continuous pumping and can lead to temporary water failures when they exceed depths of respective water structures. For instance, if drawdowns exceed 50m due to a dry spell or over-pumping, a handpump capable of lifting water from a maximum depth of 50m will be rendered non-functional until the water level recovers.

Since transmissivity can vary widely within and across hydrogeologies, we examine the respective suitabilities of different water structures across a range of transmissivity values by simulating drawdowns resulting from meeting daily domestic water needs of 55l per capita for a village with 750 residents [44]. We estimate the drawdowns after a month of pumping to meet this assumed daily water requirement. More details about the simulation model can be found in the Methods section. Simulated drawdowns allow us to evaluate the suitability of two of the three commonly used water structures - handpumps and uncovered wells - in extracting the requisite daily volumes of water. Machine-pumped sources are not similarly constrained by drawdowns, so long as drawdowns do not exceed wells depths, associated pump-set capacities, and adequate electricity supply.



(a) (b)

Figure 4: **Simulated drops in water-levels (drawdowns) across transmissivity and water structures, and the latter's adequacy in providing sufficient volumes of drinking water.** (a): Drawdowns at the end of a month based on 55 litres per capita requirement for an average-sized village of 750 people. Simulated depths are based on Equation 2 in the Methods section. The blue dotted line denotes an average depth of uncovered wells of 12m and red dotted line denotes average depth of deep handpumps of 50m. (b): Proportions of required daily drinking water met by handpumps and uncovered wells at the end of a month. Simulated volumes are based on Equation 2 in the Methods section. The red dotted line denotes 50% of the total daily requirement.

Drawdowns in hydrogeologies with transmissivity greater than $5m^2/day$ rarely go beyond tens of meters compared to the exponentially deeper levels water plummets to in hydrogeologies with transmissivity less than $3m^2/day$ (figure 4(a)). We classify hydrogeologies with median transmissivity values of $3m^2/day$ and above as high transmissivity and those with less than $3m^2/day$ as low transmissivity (appendix figure A2 and table A2 include a full list of transmissivity ranges and our classification for hydrogeologies found in Odisha). Consequently, even non-mechanized uncovered wells are sufficient in meeting drinking water needs, albeit labor intensive, in villages occurring in high transmissivity hydrogeologies. Whereas, deep drawdowns limit the volumes of water that can be pumped from non-mechanized sources in villages occurring in low transmissivity hydrogeologies (figure 4(b)).[3] In low transmissivity hydrogeologies, handpumps built for deep aquifers and operational up to 50m are unlikely to meet more than 85% of the daily water needs, while uncovered wells with average depths of 12m are likely to provide approximately 40% of the assumed daily requirement (figure 4(b)). Water demand fulfilled by both handpumps and uncovered wells decreases rapidly with decreasing transmissivity, and in hydrogeologies with transmissivity of $0.5m^2/day$ a little over 40% of daily water demand is fulfilled by handpumps and less than 20% by uncovered wells. Appendix figure A3 illustrates sensitivity of the drawdowns to uncertainties in water demand. In contrast, a mechanized well of approximately 150m depth and powered by a 4HP pumpset is expected to fulfill daily domestic water requirements in regions of similar transmissivity.[4] The simulation results indicate that, ceteris paribus, transmissivity of hydrogeologies via drawdowns induce a natural hierarchy among water structures, with machine-pumped sources performing best, followed by handpumps and uncovered wells.

Villages with high concentrations of ST population fare worse for both electricity and machine-pumped water access

Next, we examine the associations between concentrations of ST and SC within villages, and their respective distributions of drinking water supply and lighting sources (figures 5 and 6). We consider three major sources of drinking water - handpumps, uncovered wells, and machine-pumped water (taps and tubewells combined). Together, the three sources accounted for over 91% of village average household water supply in 2011. We find ST villages on average to be associated with 5 percentage points (pp) fewer households using machine-pumped water sources for their drinking needs. The negative penalty for ST villages remains even after controlling for the proportion of households using electricity, suggesting constraints beyond the presence of electricity infrastructure in using machine-pumped water infrastructure. The lower use of machine-pumped water is largely compensated by uncovered wells. On average, approximately 2pp more households report sourcing drinking water from uncovered wells in ST villages. SC villages on the other hand are modestly better off, with roughly 1pp and nearly 2pp more households using machine-pumped water and handpumps respectively. We present the full set of results with all sources of drinking water in appendix table A3.

On average, approximately 27% and 21% of all households at the village-level sourced their drinking water from machine-pumped sources and uncovered wells in 2011 respectively. The coefficients for the ST village binary therefore translate to a 20% decrease in the usage of machine-pumped sources and a 7.5% increase in uncovered wells among households in ST villages on average (appendix table A3, columns (1) and (5)). Our results for ST population are robust to alternate model specifications where we use the proportion of ST and SC populations instead of the binary indicators (appendix table A4). However the small positive association with SC population is lost for households using machine-pumped water. We also run fractional logit regressions since our outcome variables are bounded between 0 and 1 (appendix table A5), test our results by separating out taps from tubewells to account for the ambiguity in the definition of taps (appendix table A6), and include updated numbers from 2020 and 2023 on piped water supply to capture the progress that may have occurred since 2011 (appendix table A7). We find qualitatively similar results in all three cases.

We find ST villages to be penalized to a greater extent in case of lighting. On average, ST villages are associated with 10pp fewer households using electricity, whereas SC villages are modestly better off by approximately 1pp more households using electricity (figure 6). Unsurprisingly, results for kerosene operate in the opposite direction since it is the most popular substitute for fueling lights. On average, 52% households in our sample reported using electricity and 45% households used kerosene in 2011 at the village-level, implying an average loss of 18% households using electricity among ST villages. Our results for both ST and SC villages are robust to alternate fractional logit model specifications (appendix table A9) but only in the case of ST villages

[3] Our simulations are simplified estimations assuming uniform transmissivity values for the entire range of depths of the water infrastructure. In reality, these values vary across depth influencing drawdowns and water productivity. We also do not consider recharge through precipitation, which is likely to be higher in uncovered wells on account of direct capture.

[4] Pumping energy requirements estimated using equations 5, 6, and 7

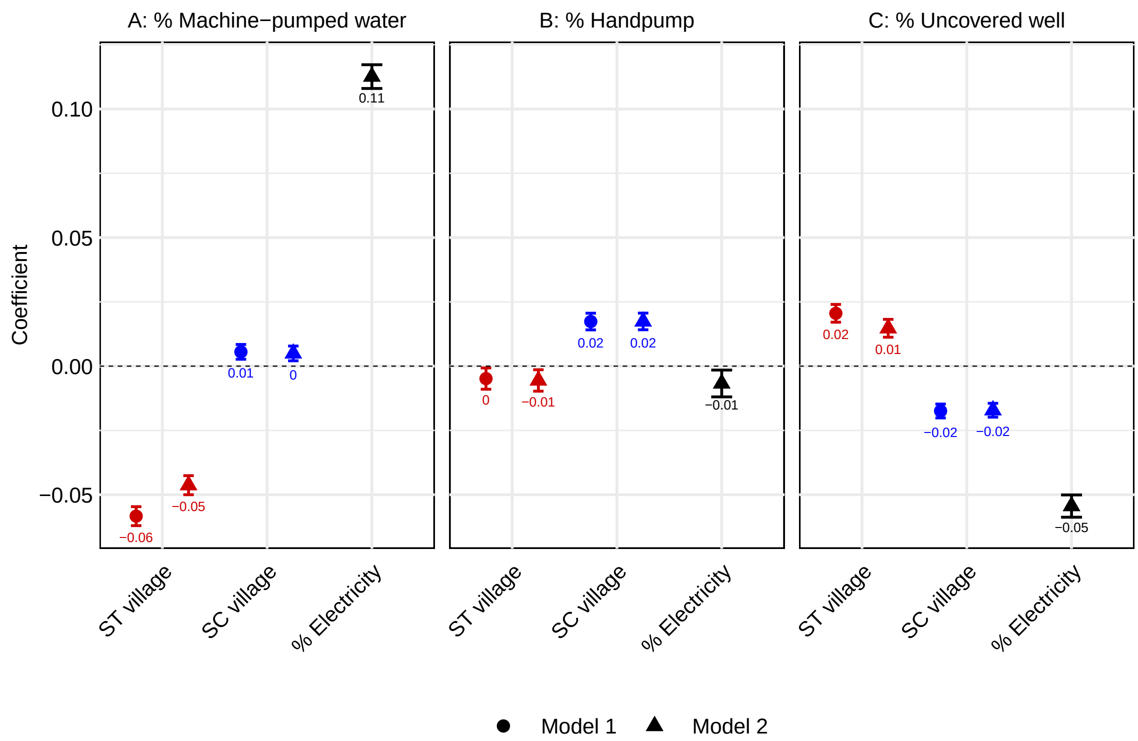


Figure 5: **Coefficient plots of regressions of the proportion of households using different sources of drinking water on ST and SC villages.** Plots show differences in the use of machine-pumped water (Panel A), handpumps (Panel B), and uncovered wells (Panel C) across ST and SC villages. Triangular and circular markers represent models with and without controlling for the proportion of households using electricity as the primary source of lighting respectively. Whiskers show 95% confidence intervals accounting for sub-district level clustering. Coefficient estimates are based on equation 1 and dataset constructed using SHRUG [31] and Census 2011 [52] data (n=233,622). ST village is a binary indicator for $\geq$ 80% ST population and SC village indicates $\geq$ 20% SC population. Full tabular results are in appendix table A3.

when we use the proportion of ST and SC population instead of the binary indicators (appendix table A10). Interestingly, proportion of households using electricity is negatively associated with the proportion of SC population, suggesting the large effect of villages without any SC population in the overall association (appendix figure A4). We also consider night time luminosity captured by the Defense Meteorological Satellite Program (DMSP) Operation Linescan System (OLS) as a way to triangulate data on electricity use. The results are similar qualitatively, where ST villages are associated with approximately 8% decrease and SC villages with approximately 3% increase in night-time luminosities (figure 6c).

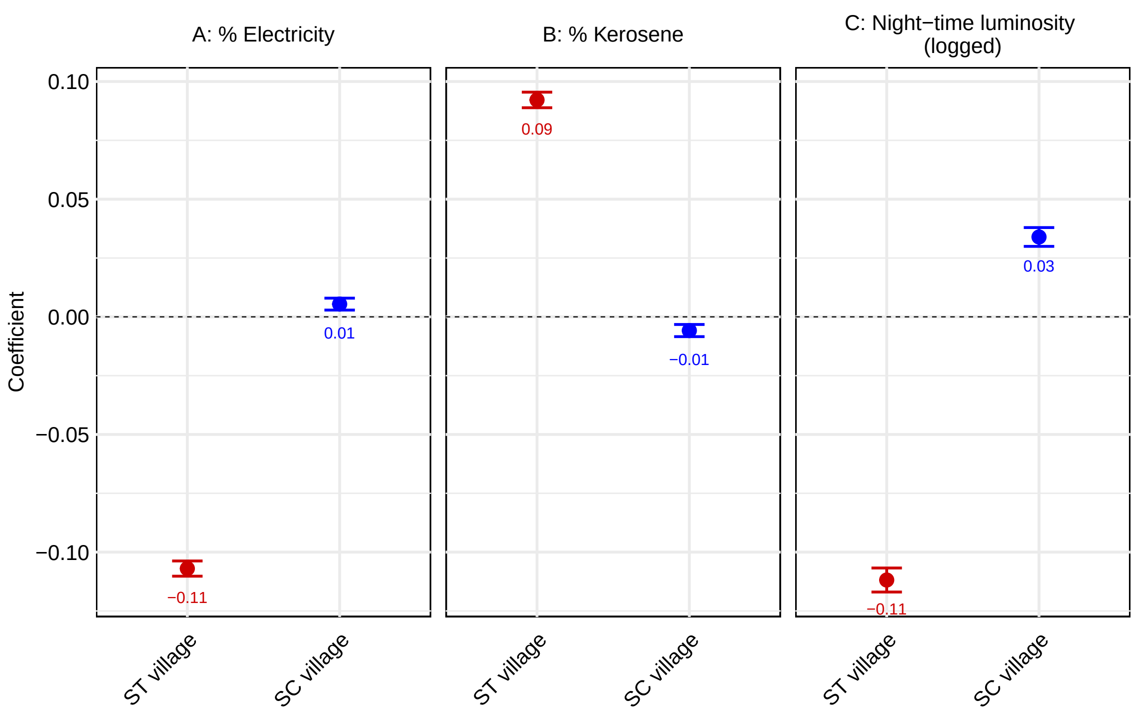


Figure 6: **Coefficient plots of regressions of the proportion of households using different sources of lighting on ST and SC villages.** Plots show differences in the use of electricity (Panel A) and kerosene (Panel B) as primary sources of lighting. Panel C corresponds to logged values of night-time luminosity of villages, averaged over 2010 to 2012. Whiskers show 95% confidence intervals accounting for sub-district level clustering. Coefficient estimates are based on equation 1 and dataset constructed using SHRUG [31] and Census 2011 [52] data (n=233,622). ST village is a binary indicator for $\geq$ 80% ST population and SC village indicates $\geq$ 20% SC population. Full tabular results are in appendix table A8.

Hydrogeologies do not explain the distributional differences between high and low concentrations of marginalized population villages

While hydrogeologies play an important role in determining success of water infrastructure, it is unclear whether they also play a role in determining the types of water structures that get built across villages. We find low transmissivity areas to have greater pumping energy needs owing to the deep drawdowns caused by continuous pumping (figure 4). Therefore, we expect greater proportions of households in low -transmissivity dominated areas to source drinking water from machine-pumped sources than in high transmissivity areas. As both socioeconomic and biophysical factors may shape access to drinking water resources, we examine whether marginalized populations are concentrated within distinct hydrogeologies, and parse out the associations between hydrogeologies and concentration of marginalized populations and access to drinking water and lighting. We classify villages as low or high transmissivity villages using district geological maps in Odisha, and find ST concentration to to be greater in low transmissivity areas (figure 7).

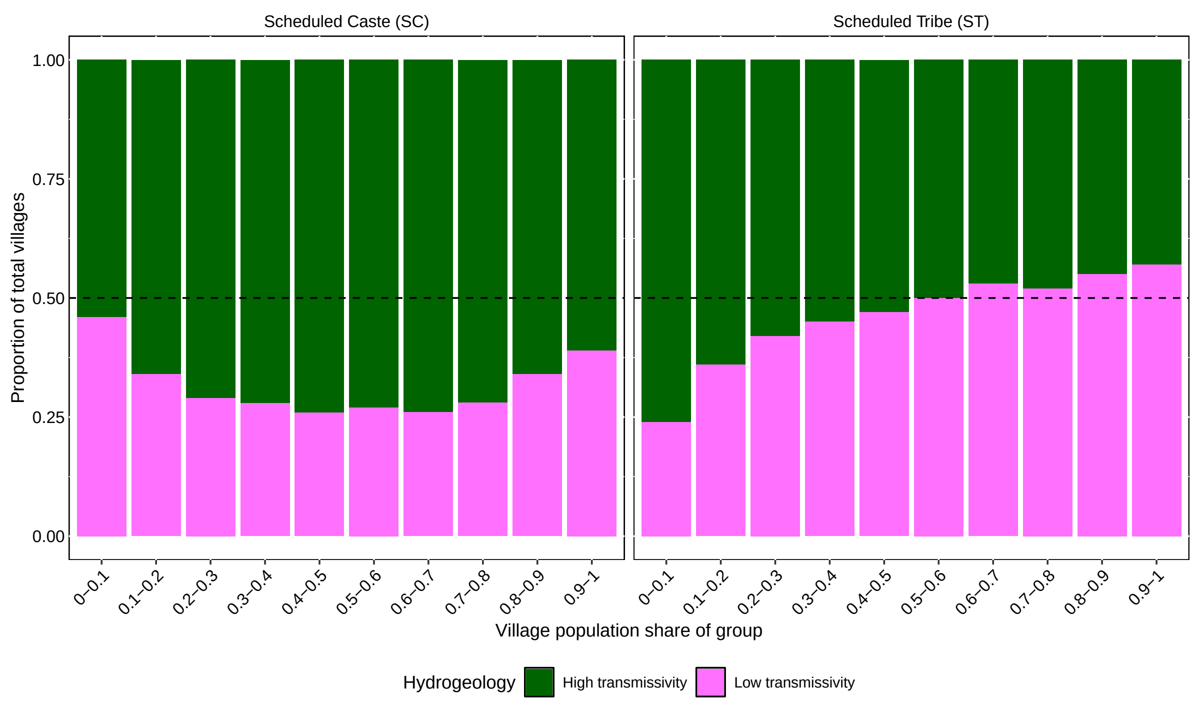


Figure 7: **Village-level distribution of low transmissivity hydrogeologies across deciles of Scheduled Caste (SC) and Scheduled Tribe (ST) share in villages across Odisha.** Each bar represents the proportion of low and high transmissivity villages in the respective village population decile. Villages are classified as low transmissivity if the majority share of hydrogeologies are classified as low transmissivity. Data compiled by authors using Census 2011 [52] and geological maps [66] (n = 43,231), full details can be found in the Methods section.

If hydrogeologies were duly considered while constructing water structures, we would expect ST villages to have greater proportions of households utilising machine-pumped water sources than non-ST villages. However, when we add the binary classification of low and high transmissivity villages in Odisha, we find ST villages are associated with approximately 3pp fewer households using machine-pumped water, and 10pp fewer households using electricity compared to non-ST villages (figure 8). Even after controlling for the proportion of electrified households, ST villages face a penalty of close to 2pp in accessing machine-pumped water. On average, only 25% households reported machine-pumped water access among villages in Odisha translating to a loss of 11% households with machine-pumped water supply among ST villages (appendix table A11, column (1)). Hydrogeologies do not appear to worsen or alleviate ST penalty, implying that planning of drinking water structures does not consider specific hydrogeologies of locations despite the latter's disproportionate influence on water productivity. We observe no similar penalty for SC villages.

Interestingly, low transmissivity villages are associated with 3pp fewer households using electricity than in high transmissivity villages on average, implying the concentration of marginalized population in low transmissivity areas may be driving the difference (figure 8b). The lack of electricity among ST villages in low transmissivity areas also suggests the urgent need for, and additional hurdles to providing machine-pumped water infrastructure to the same villages. We find no additional effect from interacting low transmissivity with ST and SC village binaries (appendix table A11).

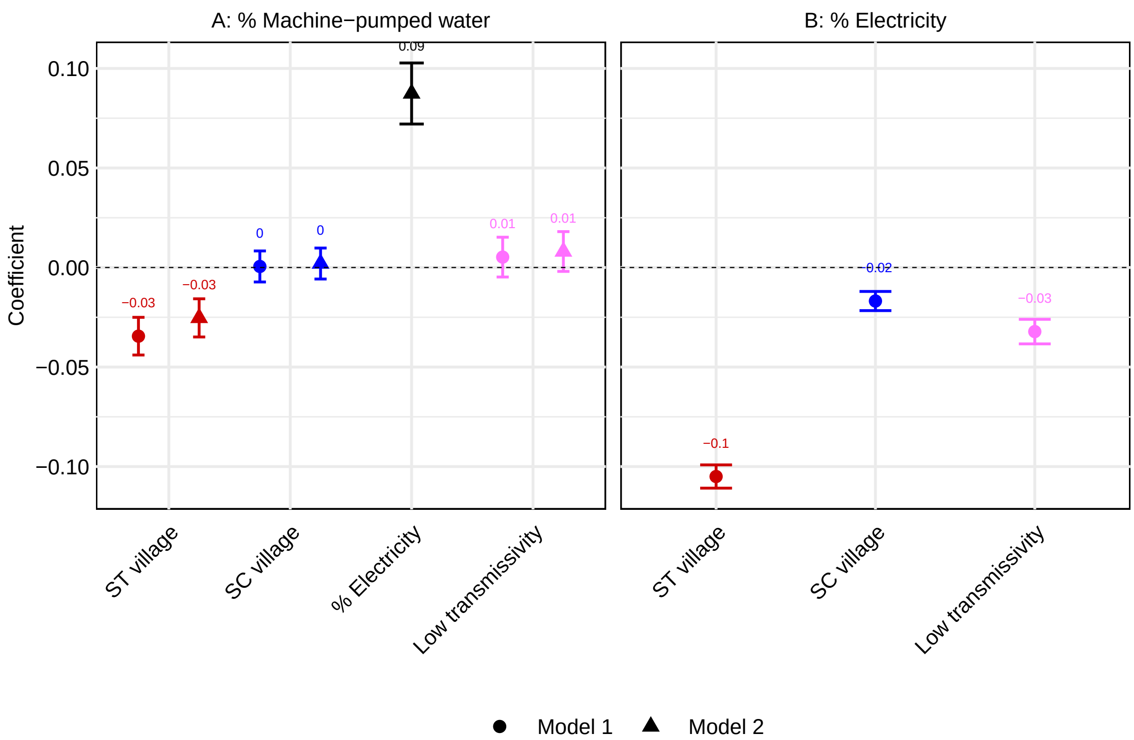
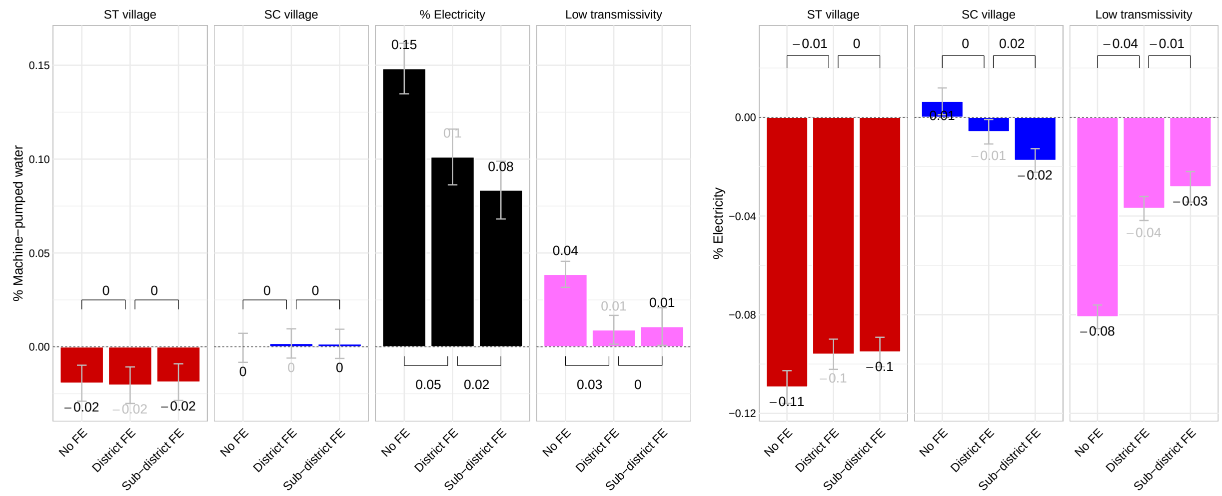


Figure 8: **Coefficient plots of regressions of the proportion of households using machine-pumped water and electricity on ST, SC villages, and low transmissivity villages.** Plots illustrate differences in the proportion of households using machine-pumped water (Panel A) as the primary source of drinking water, and electricity (Panel B) as the primary source of lighting. In panel A, triangular and circular markers represent models with and without controlling for the proportion of households using electricity as the primary source of lighting respectively. Whiskers show 95% confidence intervals accounting for sub-district level clustering. Coefficient estimates are based on equation 8 and dataset constructed using geological maps [66], SHRUG [31], and Census 2011 [52] data (n=43,173). ST village is a binary indicator for $\geq$ 80% ST population, SC village indicates $\geq$ 20% SC population and 'Low transmissivity' corresponds to villages with a majority share of low transmissivity hydrogeologies. Full tabular results are in appendix table A11.

Although hydrogeologies appear to have a weak association with the distribution of village-level water infrastructure, it may be that they are considered at district-level or state-level planning (as is the case in several government programmes targeting lagging districts and states [37, 67]). To account for India's decentralized structure of rural governance, we tease out the varying levels of distributional inequities across different administrative levels by using a method proposed by Asher et al [30] (see Methods). Overall, we find hydrogeologies have little association with the distribution of both machine-pumped water and electricity and largely do not explain the ST village penalty (figure 9). Districts are the only level where we find low transmissivity villages on average to be associated with greater proportions of households sourcing drinking water from machine-pumped structures (figure 9(a)). However, no similar statistically significant difference is observed among sub-districts or villages which have greater or lower presence of low transmissivity hydrogeologies. ST villages are associated with lower outcomes at every administrative level, with the greatest loss experienced among villages for both machine-pumped water and electricity. The loss in electricity observed among ST villages is captured almost entirely at the village-level - on average, ST villages in Odisha were associated with 11pp fewer households using electricity after controlling for the dominant hydrogeology (figure 9(b)) translating to an average loss of 32% electrified households among ST villages. For SC villages the distributional inequity is largely missing across all administrative levels for machine-pumped water, and is small and mixed for access to electricity.



(a) (b)

Figure 9: **Coefficient plots of regressions of the proportion of households using machine-pumped water and electricity on ST, SC villages, and low transmissivity villages with varying scales of fixed effects.** Bars illustrate the the effect sizes after controlling for fixed effects at the sub-district, district levels, and no fixed effects. Whiskers show 95% confidence intervals accounting for the respective levels of clustering. Bar heights represent coefficient estimates from equation 8 and annotated values are calculated using equation 9. Dataset is constructed using geological maps [66], SHRUG [31], and Census 2011 [52] data (n= 44,173). Full tabular results are in appendix table A12 and appendix table A13.

Discussion

The Indian government has been increasing rural drinking water coverage through various national-level programmes with the ultimate aim of equipping every house with piped-water supply. However, the most recent statistic available from 2023 reveals a large gap in ambition and reality, with a mere 34% of rural households with piped-water connections [68]. Rural households in India access drinking water majorly from groundwater based sources through handpumps and wells of varying depths and modes of construction. These water structures greatly differ in terms of physically and economically accessible volumes, and water quality. Large variations exist even for a single water structure across geographies and seasons despite the presence of sufficient water volumes. The influence of biophysical characteristics in determining water productivity is widely accepted, yet their role in perpetuating regional inequities and interaction with socioeconomic dimensions remains largely unexamined. We study the distributive inequities of existing water infrastructure for marginalised populations, and estimate the likelihoods of the water structures in meeting daily drinking water demands by employing a novel village-level dataset including biophysical and socioeconomic data.

We consider the two largest groups of marginalized population in India - Scheduled Castes (SC) and Scheduled Tribes (ST), and focus our analysis on seven Indian states where STs comprise over 10% of the respective state populations. By doing so, we are able to identify the extreme geographic concentration of ST populations within villages, which is in direct contrast to the distribution of the SC population. Nearly 50% of the ST population across the seven states resided in villages where they made up over 80% of the village population, while less than 3% of SC population resided in similarly concentrated villages. Our study addresses an important oversight in the existing literature on the distribution of public resources among marginalized population by revealing this extreme concentration of ST populations. We find ST-concentrated villages to experience an average decrease of 20% households sourcing drinking water from machine-pumped groundwater sources. This decrease is largely compensated by an increase of 7.5% in uncovered wells. Lower levels of machine-pumped water in ST villages could be due to shortcomings in the electricity infrastructure. Indeed, we find ST villages are associated with 10pp fewer electrified households as compared to non-ST villages. Instead, households in ST villages rely on kerosene for their lighting needs implying reduced electricity supply (or the entire absence of electricity infrastructure) also needed to power groundwater-based pumping. Drinking water and electricity are two of numerous public resources distributed through village-level infrastructure. These village-level infrastructure - water structures such as wells and handpumps in case of drinking water and transformers for electricity - are critical determinants of supply adequacy. While our analysis is not causal, we identify ST concentrated villages as a possible lever to enhance poverty targeting of government programmes and help plug subsidy leakages [69, 70].

Overall our results paint a grim picture of the plight of ST villages similar to existing literature. However, we identify a distinct geographic pattern of ST concentration across the central Indian belt unlike SC and Muslim populations [30]. We find this geographic segregation of STs to have biophysical dimensions in addition to socioeconomic ones in Odisha. Both dimensions are crucial determinants of drinking water access. Aquifer properties are widely recognized as important factors determining extractable water volumes, and yet, are largely missing from rural drinking water literature with a few notable exceptions [12, 64, 65]. We study the influence of one such aquifer characteristic of transmissivity and find non-motorized water structures to be largely inadequate in furnishing domestic water needs across low transmissivity villages. Low transmissivity translates to deep drawdowns which make electrified pumping of water essential. Even handpumps which operate up to a maximum depth of 50m are only capable of furnishing a maximum of 85% of the daily volume for a village with an average population of 750. However, we are unable to consider multiple water structures which could be present in one village, since the data are not collected. At transmissivity values of approximately $0.1m^2/day$, which are typical of granitic rocks, handpumps and wells can supply less than 10% of the daily water demand, thereby necessitating machine-pumped sources.

Despite its role in determining water productivity, we find transmissivity to play a small role in the observed distributional pattern of drinking water structures. In our dataset for Odisha, greater proportions of ST populations reside in low transmissivity villages. Yet, the negative association between ST villages and machine-pumped water does not diminish even after considering the presence of low transmissivity hydrogeologies. ST villages are thus doubly penalized - not only are they more likely to be in low transmissivity regions where mechanized pumping is essential due to deep drawdowns, they are also more likely to have less productive water structures such as uncovered wells. On the other hand, wells of approximately 150m depth and electricity capable of supporting a 4HP pump are likely adequate in serving domestic water needs of low transmissivity villages on average. Although we study the role of hydrogeologies only for a single state in our sample, the hydrogeologies we consider are present in over 60% of the country's landmass [71]. Further research is therefore needed to inform whether the patterns of marginalized population concentration in areas of low transmissivity that we observe in Odisha are repeated elsewhere in the country and manifest in similarly negative outcomes for drinking water and other public resources.

Our results are broadly consistent with recent literature that finds negative outcomes for spatially segregated marginalized communities in India, and in contrast to prior literature highlighting the positive (negative) role of social homogeneity (heterogeneity) [26, 29, 30, 48]. In particular, Bharathi et al (2024) find social homogeneity to have negative consequences in villages where SC and ST populations are cumulatively in the majority. In our sample of states, over 70% of the ST population resided in villages where they comprised more than half of the village population compared to only about 11% of SC population in similarly concentrated villages. Therefore, the negative impacts of social homogeneity are largely driven by the concentration of ST populations in our sample of states. However, we are unable to study intra-village segregation which can be particularly relevant for SC populations, in view of the widely documented inter-neighbourhood drinking water discrimination against SC populations [30, 72–74]. Our results also suggest that low political representation is an unlikely explanation for the poor outcomes of ST populations, as we observe similarly poor outcomes in Chhattisgarh and Jharkhand—states that were created expressly to advance ST interests [26].

While we are unable to offer reasons behind the poor performance of ST communities, our results have important implications for remedying their situation. First, we find the greatest distributive inequity to exist at the village-level, making villages important sites for targeting affirmative government policies. Second, it is imperative to consider local hydrogeological characteristics explicitly in drinking water policies and planning. Our results imply that water planning based only on population parameters is unlikely to succeed in low transmissivity villages with deeper drawdowns, greater pumping energy requirements, and with higher probabilities of seasonal failures. Since we find ST populations in Odisha to be concentrated in these regions, they are also more likely to be impacted. Besides identifying the urgent need for machine-pumped water sources among low transmissivity areas, we are unable to provide a comprehensive analysis of water structure requirements in view of seasonal and annual fluctuations in groundwater level, precipitation and uncertainty added by climate change. Aquifer characteristics other than transmissivity such as groundwater storage, presence of faults or fractures, connectivity to deeper aquifers and depth to water also influence extractable volumes of groundwater to varying degrees [44, 63, 64]. Third, electricity in service of pumping groundwater for drinking water supply is another important factor that needs to be brought front and center in water planning. Large parts of the rural developing world rely on groundwater for their drinking water supply. These same regions are also experiencing rapid transitions in their energy infrastructures [14, 15]. Timely incorporation of energy demand for groundwater pumping is therefore critical to ensuring adequate and reliable drinking-water supply before policy and electricity-infrastructure decisions become locked into path-dependent trajectories [1, 75]. While we find ST villages to be doubly penalized with lower access to machine-pumped water and electricity, their concentrations in low transmissivity regions in Odisha and within villages across central India, offers a targeted route to improving their situation. Future policies and research need to consider biophysical elements impacting public resource access, and identify areas of overlap with distributions of marginalized populations. These overlaps, while creating unforeseen poverty traps for marginalized communities, also offer opportunities of targeting and compounding positive impacts of public resources.

Methods

Data Sources

We use four main sources of data for our analyses - Population census data, 2020 Mission Antyodaya survey data, night-time luminosity, forest cover, and ruggedness compiled by Asher et al [31]; Primary sources of drinking water and lighting from the house-listing tables of the most recent population census conducted in 2011 from censusindia.gov.in [52]; 2023 Mission Antyodaya survey data from the Department of Rural Development website [68] and; hydrogeological data for Odisha from the state's Centre of Environmental Information System (ENVIS) [66]. We match all the datasets at the village-level.

SHRUG

Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG) provides data at the "shrid" level, which is a geographic unit with stable boundaries that occasionally combines multiple villages. We limit our analysis to shrids representing single villages and hence, present all results at the village level. We use five types of data from v2.0 of the SHRUG -

1. **Village Amenities modules** - We use the Village Amenities modules from the 1991, 2001, and 2011 Censuses of India to calculate the village ST and SC population shares, and to control for various village characteristics. While there can be occasional discrepancies, for our sample, the population numbers reported in the Primary Census Abstract are identical to those in the Village Amenities module. The census reports the total, ST, and SC population for each village. We calculate the ST and SC share for each village and define ST villages as those that have $\geq$ 80% ST share in at least one of the three years and $>$ 50% ST share in all three years. ST village is thus a time invariant binary variable that identifies villages with at least a majority share of STs over three decades. Similarly, we define SC villages as those that have $\geq$ 20% SC share in at least one of the three years and $>$ 10% SC share in all three years.

   We also define village-level controls using this data - village area (in hectares), non-ST/SC population (total - ST - SC), and total number of households. We control for remoteness of villages in two ways - road (presence of *pucca*/paved or black-topped roads) and straight-line distance to the district headquarters (in km).

2. **Night time luminosity** - We use annual mean night time luminosity as an additional proxy for village electrification. Night time luminosity is captured by the Defense Meteorological Satellite Program Operation Linescan System (DMSP-OLS), expressed as the average brightness of all pixels in the village boundary, and ranges from 0 to 63. We use village-level night time luminosity averaged across 2010, 2011, and 2012.

3. **Forest cover** - We use the forest cover module to identify forested area [76]. Annual forest cover is measured as Vegetation Continuous Fields (VCF) using satellite imagery which predicts primary and secondary forest coverage for 250m pixels. The satellite images are taken from MODIS, a product which measures tree cover, after which crops and plantations are differentiated from forests using a machine learning algorithm based on higher resolution satellite imagery. We use percentage area of the village covered by forests averaged over 2010-2012 for Census 2011 outcomes and 2018-2020 for Mission Antyodaya 2020 and 2023 outcomes due to data limitations.

4. **Ruggedness** - We use the Terrain Ruggedness Index (TRI) constructed using elevation data from the Shuttle Radar Topography Mission [77] to identify ruggedness. The TRI is a relative measure of the difference in elevation between each 1-arc second pixel and its eight adjacent pixels. Full construction details can be found on the SHRUG website.

5. **Mission Antyodaya 2020** - We use the proportion of households with piped water connections, total number of households, and indicators for presence of road and electricity in the village in 2020 to examine robustness of main results to recent data. Mission Antyodaya was a 2017 initiative to reorient a slew of

government schemes around the Gram Panchayat, a village-level governing body designed to be the basic unit for planning. It includes a quasi-annual survey, for which data is available from SHRUG for the 2020 round. Panchayat officials were surveyed about GP and village infrastructure related to drinking water, sanitation, roads, communication, health, and poverty alleviation programmes, among others. Surveyors physically verified their responses where possible and geotagged assets. The data does not contain the population breakdown by caste. Hence, we impute the non-SC/ST population using the growth rate of the total households in the village between 2011 and 2020, assuming that the non-SC/ST population grew at the same rate as total households. As the survey does not contain data on the proprotion of households with electricity, we define a binary indicator 'Village electrified' that equals 1 if the village 'has electricity' and 0 otherwise.

Mission Antyodaya 2023

We obtain data from the 2023 round of the Mission Antyodaya survey from the Department of Rural Development website [68]. We match this to SHRUG data using census village codes. Dropping villages that cannot be matched due to data issues such as duplicated village codes, we are able to use 2023 Mission Antyodaya data for 91.6% of villages in our main sample. Similar to the 2020 data, we impute the non-SC/ST population using the growth rate of the total households in the village between 2011 and 2023. As the survey questionnaire changes over time, we define the 'village electrified' binary as 1 if the village receives non-zero hours of domestic electricity in 2023 and 0 otherwise.

Census house-listing data

The census house-listing module provides proportions of total households accessing the different sources of drinking water (including uncovered wells, handpumps, tube wells, and piped water) for each village [52]. We do this as although the village amenities data provides binary indicators of the presence of various drinking water infrastructures, it does not report how many households rely on each source. The house-listing data also provide similar breakdowns for the main sources of lighting, which are largely kerosene and electricity. We use village crosswalks provided by SHRUG to combine the house-listing tables with the other census modules. In all, we are able to match 99.8% of the villages across census data provided by SHRUG [31] with the villages in the house-listing module (figure 10 panel A). Although more recent data on household access to piped water connections exist, we choose not to include it in our main analysis since the data sources do not cover drinking water sources exhaustively. Mission Antyodaya [68, 78] and the Jal Jeevan Mission [79] capture progress for households with piped water connections within households and we use data from the former to test the robustness of our results.

Hydrogeological characteristics

We identify primary rock formations in each village by digitising and overlaying geological maps from ENVIS onto village base maps from SHRUG [31, 66] for one of our seven states - Odisha. Digitising and overlaying maps is imprecise owing to the mismatch between geological and administrative boundaries of villages. Consequently, in cases where the geological and administrative boundaries of the village are misaligned, we under or over-classify few villages, and drop villages that appear to be $< 50\%$ (the extent of overlap between the geological and administrative area is less than half of the administrative area) or $> 110\%$ (village is constituted by multiple hydrogeologies, which together amount to greater than 110% of the administrative area) covered by rocks (1,212 villages of the original 45,873). We classify the remainingvillages as high transmissivity or low transmissivity - if low transmissivity rocks constitute $> 50\%$ of the village area covered by rocks, i.e. the majority share, we classify it as a low transmissivity village. We use the results from the simulation model in section to classify rocks as low or high transmissivity. We exclude villages that are dominated by rocks that are dominated by rock types that we are unable to classify due to lack of transmissivity values from exploratory wells in Odisha. The full range of rock types and transmissivity values along with their classification can be found in appendix table A2 and figure A2.

On average, rock formations cover geographic area that are in between blocks and villages - for example, the average village area is approximately $1.8km^2$in Chandua sub-district in Mayurbhanj district of Odisha, which has a cumulative granitic area of approximately $10km^2$ and alluvial area of $80km^2$.

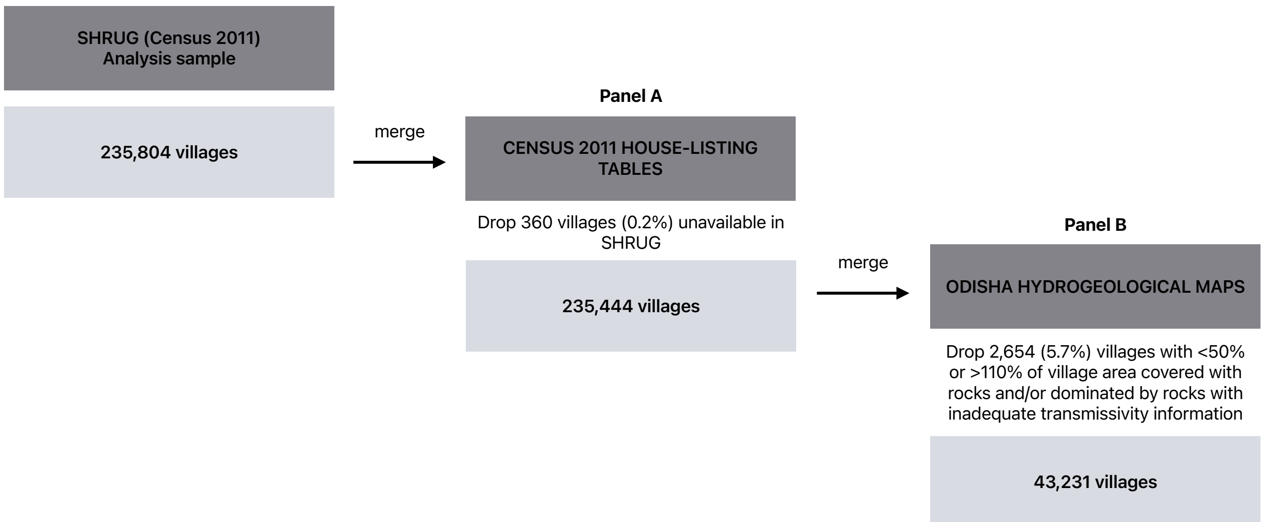


Figure 10: Results from data merges between SHRUG, Census 2011 house-listing tables, and Odisha hydrogeological maps

Estimation Strategy

We use data matched from house-listing tables and SHRUG to study the distribution of drinking water and electricity across ST, SC, and non-ST,SC villages (figure 10 panel A), and data matched across all three sources to study the role of hydrogeology in public resource distribution in Odisha (figure 10 panel B).

SC/ST villages and access to drinking water and lighting

We employ one-way fixed effects model (Equation 1) to estimate the association between ST, SC concentrations, and access to sources of drinking water and lighting.

$$Y_{ij} = \beta_1 \mathrm{ST}_i + \beta_2 \mathrm{SC}_i + X_i + \gamma_j + \epsilon_i \tag{1}$$

Where, Y_{ij} is the proportion of households accessing electricity (or kerosene) and machine-pumped sources of water (or handpumps, uncovered wells) in village i in subdistrict j in 2011. ST_i and SC_i are binary indicators of whether the village is an ST village or an SC village respectively. Our data includes 47,675 ST and 60,227 SC villages. A small proportion of our sample were classified as both ST and SC villages, which we drop from the analysis. Therefore, the comparison group are villages which have ST and SC populations below 80% and 20% respectively.

X_i is a vector of village controls including the non-SC/ST population, total number of households in village i, binary indicator capturing the presence of road (whether the village has black-topped or *pucca*/paved roads), distance from district headquarters (km), village area in hectares, average forest cover in the village, and ruggedness. In the case of distribution of drinking water sources, X_i also includes the proportion of households reporting electricity as the primary source of lighting. γ_j are subdistrict fixed effects - our estimates therefore represent the differences between ST, SC, and non-ST/SC villages *within* sub-districts. Standard errors are clustered at the sub-district level.

Hydrogeology and access to drinking water and lighting

We conduct two sets of analyses to unpack how varying hydrogeological conditions shape access to water. First, we estimate water volumes that can be extracted from the various drinking water structures across hard and non-hard rock geologies. Second, we use one-way fixed effects regression models to test whether hydrogeological conditions influence the types of water structures accessed by varying proportions of marginalized households across villages.

1. **Role of hydrogeologies in determining extracted volumes from different water structures**

   We estimate maximum productivities of different drinking water structures based on their respective depths and hydrogeological characteristics of underlying rocks across villages. We assume a daily water requirement of 41,250 litres for a representative village of 750 people using the national drinking water policy's goal of 55 litres per capita per day. We simulate daily drawdowns from meeting daily water requirement for bore wells (structures that support handpumps) and uncovered wells using equations (2-4) across a range of Transmissivity ($0.001m^2/day - 10m^2/day$) and average Storativity of 5.5×10^{-5} (averaged over 1.0×10^{-4} and 1.0×10^{-5}).

   Drawdowns are the depths to which groundwater levels fall with continuous pumping and subsequently recover once pumping has ceased. Average daily water requirements cannot be met without deepening water structures for aquifers where drawdowns exceed depths. We assume well depths of 12m and 50m for uncovered wells and handpump-based bore wells respectively. For both structures, we run constrained simulations to calculate new daily pumping rates Q^* which would limit the monthly drawdowns to well depths. Q^*/Q is therefore the proportion of daily water requirement that is met.

   The drawdowns at the end of a month are a function of the levels to which water drops at the end of the each day and the subsequent increase in water levels during hours of recovery. We use the principle of superposition of Theis equation to calculate monthly drawdowns [44] -

$$
\begin{aligned}
H_s &= \frac{Q_1}{4\pi T} W(u_1) + \frac{Q_2 - Q_1}{4\pi T} W(u_2) + ... && (2)\\
W(u_n) &= -0.5772 - ln(u_n) + u_n - \frac{u_n^2}{2.2} + \frac{u_n^3}{3.3} + ... && (3)\\
u_n &= \frac{r^2 S}{4\pi T t_n} && (4)
\end{aligned}
$$

   Where, Q_i is the daily pumping rate in month i of $41.25m^3$, H_s is the drawdown at the end of the month, T Transmissivity (m^2/day) and S Storativity of the specific hydrogeology. Radius r is assumed to be 50mm for boreholes and 5m for uncovered wells. t_n represents time elapsed since the start of pumping. We assume that water is extracted for 8 hours each day with a recovery duration of 16 hours during which no water is extracted.

   We also model the energy requirements for pumping water to meet the daily water requirement using the following equations -

   (a) Daily average energy requirement in month i

$$E_i = \frac{Q_i \rho g H}{3.6 \times 10^6 \eta_{pump}} \tag{5}$$

   Where, H is the average depth to groundwater (m), assumed to be 5m at the beginning, to which drawdown from pumping is added. η_{pump} pump-set efficiency and assumed to be 70%, ρ is the density of water (1000 kg/m^3) and g is the gravity constant (9.8 m/s^2).

   (b) Required Pump Capacity -

$$P_{cap} = \frac{max(E_i) \times 1.341}{t_{max}} \tag{6}$$

   Where, $P_{cap,i}$ is the required pump capacity (HP), $maxE_i$ is the maximum energy requirement. t_{max} is the maximum hours of pumping in a day and assumed to be 8 hours.

   (c) Required hours of pumping -

$$t_i = \frac{E_i \times 1.341}{P_{cap}} \tag{7}$$

2. **Role of low transmissivity hydrogeologies in drinking water and lighting access** We use a second set of one-way fixed effects models (Equation 8) to study the role of hydrogeological characteristics in addition to ST, SC concentrations in characterizing access to sources of drinking water.

$$Y_{ijk} = \beta_1 \mathrm{ST}_i + \beta_2 \mathrm{SC}_i + \alpha \mathrm{LowT}_i + \mathbf{X}_i + \gamma + \epsilon_i (8)$$

As before, Y_{ijk} is the proportion of households accessing electricity (or kerosene) and machine-pumped sources of water (or handpumps, uncovered wells) in village i in subdistrict j and district k in Odisha during 2011. ST_i and SC_i are binary indicators of SC and ST villages. In addition to village controls of non-SC/ST population, total households, presence of roads, distance from district headquarters, village area, forest cover, and ruggedness, we add a binary indicator $LowT_i$ to indicate whether the village is dominated by low transmissivity hydrogeologies or not.

To parse the inequalities in access at different administrative levels, we run equation (8) with varying fixed effects γ (sub-district, district and no fixed effects), and use equations (9-12) adapted from [30].

$$\alpha_{District} = \beta_{Total} - \beta_{District} \tag{9}$$

$$\alpha_{Sub-district} = \beta_{District} - \beta_{Sub-district} \tag{10}$$

$$\alpha_{Village} = \beta_{Sub-district} \tag{11}$$

$$\alpha_{District} + \alpha_{Sub-district} + \alpha_{Village} = \beta_{Total} \tag{12}$$

We first estimate a model with no fixed effects giving us β_{Total} as the total association between ST villages and our outcomes of interest. This total effect is an additive composite of the associations across villages, sub-districts and districts. Next, we use a district fixed effects models with $\gamma_{District}$, which allows us to estimate $\beta_{District}$, the association with ST villages within each district. The association with ST villages across districts is calculated by removing the within-district association, $\beta_{District}$, from the total association, β_{Total}. Hence, we calculate $\alpha_{District} = \beta_{Total}$ - $\beta_{District}$ as the across-district association. Similarly, we estimate $\beta_{Sub-district}$ (association within sub-districts) using sub-district fixed effects $\gamma_{Sub-district}$, and calculate $\alpha_{Sub-district} = \beta_{District}$ - $\beta_{Sub-district}$, which is the association across sub-districts within districts. Finally, $\alpha_{Village} = \beta_{Sub-district}$ (association across villages within sub-districts) using sub-district fixed effects. Adding equations (9-11), we get the total association β_{Total} as the sum $\alpha_{District} + \alpha_{Sub-district} + \alpha_{Village}$, equation (12).

References

[1] WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. Data dashboard, 2022.

[2] Werner Aeschbach-Hertig and Tom Gleeson. Regional Strategies for the Accelerating Global Problem of Groundwater Depletion. *Nature Geoscience*, 5(12):853–861, December 2012.

[3] Tom Gleeson, Yoshihide Wada, Marc F. P. Bierkens, and Ludovicus P. H. van Beek. Water Balance of Global Aquifers Revealed by Groundwater Footprint. *Nature*, 488(7410):197–200, August 2012.

[4] K. J. Joy and Veena Srinivasan. Climate Change and Rural Water Security. In Ravi Agarwal and Omita Goyal, editors, *The Crisis of Climate Change: Weather Report*, pages 149–163. Routledge India, London, 1 edition, September 2021.

[5] Katrina Jessoe. Improved Source, Improved Quality? Demand for Drinking Water Quality in Rural India. *Journal of Environmental Economics and Management*, 66(3), 2013.

[6] Sheetal Sekhri. Wells, Water, and Welfare: The Impact of Access to Groundwater on Rural Poverty and Conflict. *American Economic Journal: Applied Economics*, 6(3):76–102, July 2014.

[7] William Easterly and Ross Levine. Africa's Growth Tragedy: Policies and Ethnic Divisions. *The Quarterly Journal of Economics*, 112(4):1203–1250, November 1997.

[8] Alberto Alesina, Reza Baqir, and William Easterly. Public Goods and Ethnic Division. *The Quarterly Journal of Economics*, 114(4):1243–1284, November 1999.

[9] Abhijit Banerjee and Rohini Somanathan. Caste, Community and Collective Action: The Political Economy of Public Good Provision in India. *Working Paper*, 2001.

[10] Yuhki Tajima, Krislert Samphantharak, and Kai Ostwald. Ethnic Segregation and Public Goods: Evidence from Indonesia. *American Political Science Review*, 112(3):637–653, August 2018.

[11] Naveen Bharathi, Deepak Malghan, and Andaleeb Rahman. Ethnic Diversity and Economic Development with Spatial Segregation. *Economics Letters*, 222, 2023.

[12] Sudatta Ray and Sally M. Benson. Energy Models in Service of Aquifer Specific Groundwater Irrigation expansion in India. *Environmental Research Letters*, Accepted Manuscript, August 2024.

[13] Fiona Burlig and Louis Preonas. Out of the Darkness and Into the Light? Development Effects of Rural Electrification. *Energy Institue at Haas Working Paper*, page 54, 2016.

[14] Kenneth Lee, Edward Miguel, and Catherine Wolfram. Does Household Electrification Supercharge Economic Development? *Journal of Economic Perspectives*, 34(1):122–144, February 2020.

[15] Sudatta Ray and Hemant K. Pullabhotla. The changing impact of rural electrification on Indian agriculture. *Nature Communications*, 14(1):6780, October 2023.

[16] Abhijit Banerjee and Lakshmi Iyer. History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. *American Economic Review*, 95(4):24, 2005.

[17] Alice Z. Xu. Segregation and the Spatial Externalities of Inequality: A Theory of Interdependence and Public Goods in Cities. *American Political Science Review*, 118(3):1431–1448, August 2024.

[18] G. S. Ghurye. *Scheduled Tribes*. Popular Book Depot, 1959.

[19] Virginius Xaxa. Transformation of Tribes in India: Terms of Discourse. *Economic and Political Weekly*, 34(24):1519–1524, 1999.

[20] T.K. Oommen. Scheduled Castes, Scheduled Tribes, and the Nation: Situating G. S. Ghurye. 60(2):228–244. Sage Publications.

[21] Virginius Xaxa. Tribes as Indigenous People of India. *Economic and Political Weekly*, 34(51):3589–3595, 1999.

[22] Vinita Damodaran. Environment, Ethnicity and History in Chotanagpur, India, 1850–1970. *Environment and History*, 3(3):273–298, 1997.

[23] Vinita Damodaran. Looking beyond the Apocalypse: Environmental Crisis, Colonial Environmentalism and Eastern India's Tribal Communities. In *The Environmental Apocalypse*. Routledge, 2022. Num Pages: 18.

[24] Vinita Damodaran and Felix Padel. Investment-Induced Displacement in Central India: A Study in Extractive Capitalism. 38(2):396–411, 2018. Publisher: Duke University Press.

[25] Yoko Kijima. Caste and Tribe Inequality: Evidence from India, 1983–1999. *Economic Development and Cultural Change*, 54(2), 2006.

[26] Abhijit Banerjee and Rohini Somanathan. The Political Economy of Public Goods: Some Evidence from India. *Journal of Development Economics*, 82(2), 2007.

[27] Abhijit Banerjee, Lakshmi Iyer, and Rohini Somanathan. Public Action for Public Goods. *NBER Wroking Paper Series*, Working Paper 12911:40, February 2007.

[28] Alberto Alesina and Ekaterina Zhuravskaya. Segregation and the Quality of Government in a Cross Section of Countries. *American Economic Review*, 101(5):1872–1911, 2011.

[29] Naveen Bharathi, Deepak Malghan, Sumit Mishra, and Andaleeb Rahman. Status Inequality and Public Goods. *World Development*, 176, 2024.

[30] Sam Asher, Kritarth Jha, Paul Novosad, Anjali Adukia, and Brandon Tan. Residential Segregation and Unequal Access to Local Public Services in India: Evidence from 1.5m Neighborhoods. *Working Paper*, 2024.

[31] Sam Asher, Tobias Lunt, Ryu Matsuura, and Paul Novosad. Development Research at High Geographic Resolution: An Analysis of Night-Lights, Firms, and Poverty in India Using the SHRUG Open Data Platform. *The World Bank Economic Review*, 35(4), 2021.

[32] World Bank. Overview of Rural Decentralization in India: Volume 1, 2000.

[33] Pranab Bardhan. Decentralization of Governance and Development. *Journal of Economic Perspectives*, 16(4):185–205, December 2002. Accessed via https://www.aeaweb.org/articles?id=10.1257/089533002320951037.

[34] Meera Mahadevan and Ajay Shenoy. The Political Consequences of Resource Scarcity: Targeted Spending in a Water-Stressed Democracy. 220:104842.

[35] Adam Michael Auerbach and Gabrielle Kruks-Wisner. The Geography of Citizenship Practice: How the Poor Engage the State in Rural and Urban India. 18(4):1118–1134.

[36] Comptroller and Auditor General of India (CAG). Performance Audit of Accelerated Rural Water Supply Programme (ARWSP). 2008.

[37] Minstry of Drinking Water and Government of India Sanitation. National Rural Drinking Water Programme (NRDWP) guidelines, 2013.

[38] Harikishan Sharma. "CAG holds first audit of Jal Jeevan Mission, reports expected soon", 06 June 2025. Accessed via https://indianexpress.com/article/india/cag-holds-horizontal-audit-of-jal-jeevan-mission-reports-soon-10050980/.

[39] Partha Behera. "Jal Jeevan Mission under fire: Accusing dy CM Sao of providing inaccurate info, Oppn MLAs walk out". *The Times of India*, 16 July 2025. Accessed via https://timesofindia.indiatimes.com/city/raipur/jal-jeevan-mission-under-fire-accusing-dy-cm-sao-of-providing-inaccurate-info-oppn-mlas-walk-out/articleshow/122529440.cms.

[40] Harikishan Sharma. “Jal Jeevan graft crackdown: Action against officials in 15 states & UTs, mainly from UP”. *The Indian Express*, 10 November 2025. Accessed via https://indianexpress.com/article/india/jal-jeevan-graft-crackdown-action-against-officials-in-15-states-uts-mainly-from-up-10355980/.

[41] Press Trust of India. “Parliament Panel Flags Concerns Over Authenticity of Jal Jeevan Mission Data”. *The Economic Times*, 11 August 2025. Accessed via https://economictimes.indiatimes.com/news/india/parliament-panel-flags-concerns-over-authenticity-of-jal-jeevan-mission-data/articleshow/123237426.cms?from=mdr.

[42] TNM Staff. “K'taka Minister Wants 1-km Deep Borewell Dug for Water, Clueless that it could Worsen the Problem”. *The News Minute*, 29 Apr 2017. Available at: `https://www.thenewsminute.com/karnataka/k-taka-minister-wants-1-km-deep-borewell-dug-water-clueless-it-could-worsen-problem-61185`.

[43] Minor Irrigation Statistics Division. Manual For Data Collection in the Census of Minor Irrigation Schemes 2013-14. Technical report, Ministry of Water Resources, Government of India, New Delhi, 2017.

[44] R. Allan Freeze and John A. Cherry. *Groundwater*. Prentice-Hall, Englewood Cliffs, N.J, 1979.

[45] Veena Srinivasan, N R Lakshmikantha, G Manjunatha, and Ganesh Shinde. Chasing The Water Table - The Impact of Groundwater Depletion on Rural Drinking Water Supply in Peninsular India. *Earth ArXiv Preprint*, 2023.

[46] Rural Domestic Water Supply in India: Progress and Issues. In *Drinking Water Security in Rural India: Dynamics, Influencing Factors, and Improvement Strategy*, Water Resources Development and Management. Springer Nature Singapore, Singapore, 2022.

[47] Aude-Sophie Rodella, Esha Zaveri, and François Bertone. The Hidden Wealth of Nations: A Global Dataset of Aquifer Typologies and Groundwater Ressources.

[48] Divya Balasubramaniam, Santanu Chatterjee, and David B. Mustard. Got Water? Social Divisions and Access to Public Goods in Rural India. *Economica*, 81(321), 2014.

[49] Ashish Kumar Sedai. Who Benefits from Piped Water in the House? Empirical Evidence from a Gendered Analysis in India. ADBI Working Paper Series, 2021.

[50] Pallavi Choudhuri and Sonalde Desai. Lack of Access to Clean Fuel and Piped Water and Children's Educational Outcomes in rural India. *World Development*, 145, 2021.

[51] Jyotsna Jalan and Martin Ravallion. Does Piped Water Reduce Diarrhea for Children in Rural India? *Journal of Econometrics*, 2003.

[52] Office of the Registrar General & Census Commissioner, India. Census of India, 2011. Available at: `https://censusindia.gov.in/census.website/data/census-tables`. Retrieved in August 2024.

[53] Tushaar Shah, O. P. Singh, and Aditi Mukherji. Some aspects of South Asia's groundwater irrigation economy: Analyses from a survey in India, Pakistan, Nepal Terai and Bangladesh. *Hydrogeology Journal*, 14(3):286–309, March 2006.

[54] Tushaar Shah and Shilp Verma. Co-Management of Electricity and Groundwater: An Assessment of Gujarat's Jyotirgram Scheme. *Economic and Political Weekly*, 43(7):59–66, 2008.

[55] Abhijit Mukherjee, editor. *Groundwater of South Asia*. Springer Hydrogeology. Springer Singapore, Singapore, 2018.

[56] Soumya Balasubramanya and David Stifel. Viewpoint: Water, Agriculture & Poverty in an Era of Climate Change: Why Do We Know So Little? *Food Policy*, 93:101905, May 2020.

[57] Reena Badiani and Katrina Jessoe. The Impact of Electricity Subsidies on Groundwater Extraction and Agricultural Production. *Working Paper, University of California at Davis*, page 41, 2011.

[58] Reena Badiani and Katrina Jessoe. Electricity Prices, Groundwater, and Agriculture: The Environmental and Agricultural Impacts of Electricity Subsidies in India. In Wolfram Schlenker, editor, *Agricultural Productivity and Producer Behavior*. University of Chicago Press, 2019.

[59] Nicholas Ryan and Anant Sudarshan. Rationing the Commons. *Journal of Political Economy*, Forthcoming(Mimeo), June 2020.

[60] World Bank. Deep Wells and Prudence: Towards Pragmatic Action for Addressing Groundwater Overexploitation in India, 2010.

[61] UNICEF. Handpump India Mark II Technical Specifications. Taken from: `https://supply.unicef.org/s0009317.html`.

[62] Government of India Ministry of Jal Shakti. Operational guidelines for the implementation of Jal Jeevan Mission (JJM), 2019.

[63] Robyn Horan, Pawan S. Wable, Veena Srinivasan, Helen E. Baron, Virginie J. D. Keller, Kaushal K. Garg, Nathan Rickards, Mike Simpson, Helen A. Houghton-Carr, and H. Gwyn Rees. Modelling Small-Scale Storage Interventions in Semi-Arid India at the Basin Scale. *Sustainability*, 13(11):6129, May 2021.

[64] Saul Arlosoroff, Gerhard Tschannerl, David Grey, William Journey, Andrew Karp, Otto Langenegger, and Robert Roche. Community Water Supply: The Handpump Option. Technical report, The World Bank, Washington DC, May 1987.

[65] D.J. MacAllister, D. Nedaw, S. Kebede, T. Mkandawire, P. Makuluni, C. Shaba, J. Okullo, M. Owor, R. Carter, J. Chilton, V. Casey, H. Fallas, and A.M. MacDonald. Contribution of Physical Factors to Handpump Borehole Functionality in Africa. 851:158343.

[66] ENVIS Centre of Odisha. Blockwise Groundwater Reource and Utilization, 2011. Available at: `https://orienvis.nic.in/index1.aspx?lid=38&mid=1&langid=1&linkid=5927`. Retrieved in September 2020.

[67] Michael Green and Amit Kapoor. An Assessment of Aspirational Districts Programme. 2020.

[68] Government of India Ministry of Rural Development. Mission Antyodaya. 2023. Accessed via `https://missionantyodaya.dord.gov.in/rawData2022.html`.

[69] Prabhat Barnwal. Curbing Leakage in Public Programs: Evidence from India's Direct Benefit Transfer Policy. *American Economic Review*, 114(12):3812–3846, 2024.

[70] Reetika Khera. Trends in Diversion of Grain from the Public Distribution System. *Economic and Political Weekly*, pages 106–114, 2011.

[71] SC Dhiman, SK Sinha, Rajesh Chandra, Anita Kujur, BK Sahoo, D Dhayamalar, Rani VR, SK Swaroop, and JR Verma. Aquifer Systems of India. Government Document, Central Groundwater Board, Ministry of Water Resources, Government of India, New Delhi, 2012.

[72] Deepa Joshi. Caste, Gender and the Rhetoric of Reform in India's Drinking Water Sector. *EPW*, XLVI(18):56–63, 2011.

[73] Farhat Naz. Water, Water Lords, and Caste: A Village Study from Gujarat, India. *Capitalism Nature Socialism*, 26(3), 2015.

[74] Kathleen O'Reilly and Richa Dhanju. Public Taps and Private Connections: The Production of Caste Distinction and Common Sense in a Rajasthan Drinking Water Supply Project. *Trans Inst British Geog*, 39(3):373–386, 2014.

[75] Guillaume Zuffinetti, Simon Meunier, Donald John MacAllister, Peter K. Kitanidis, and Alan M. MacDonald. A Method for Estimating Maximum Safe Installable Power for Groundwater Extraction with Application to Africa. 955:177062.

[76] C Dimiceli, M Carroll, R Sohlberg, D Kim, M Kelly, and J Townshend. MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250 m SIN Grid V006 [Data Set]. *NASA EOSDIS Land Process*, 2015.

[77] Tom G Farr and Mike Kobrick. Shuttle Radar Topography Mission produces a wealth of data. *Eos, Transactions American Geophysical Union*, 81(48):583–585, 2000.

[78] Government of India Ministry of Rural Development. Mission Antyodaya. 2020. Accessed via SHRUG `https://www.devdatalab.org/shrug_download/`, v2.1.pakora.

[79] Department of Drinking Water and Government of India Sanitation, Ministry of Jal Shakti. Jal Jeevan Mission. 2025. Accessed via `https://ejalshakti.gov.in/jjmreport/JJMIndia.aspx`.

[80] Government of India National Sample Survey Office. NSS 76th Round: Drinking Water, Sanitation, Hygiene, and Housing Condition in India. (NSS Report No. 584), 2018.

[81] Government of India National Sample Survey Office. NSS 78th Round: Multiple Indicator Survey of India. (NSS Report No. 589), 2020-2021.

[82] Office of Registrar General and india Census Commissioner. Census 2011 Instruction Manual for Housing and Houselisting Census. 2011.

[83] Ministry of Drinking Water and Sanitation, Government of India. Operation and Maintenance Manual for Rural Water Supplies.

[84] Ministry of Water Resources Government of India. Manual for Data Collection in the Census of Minor Irrigation Schemes 2013-2014.

Appendix

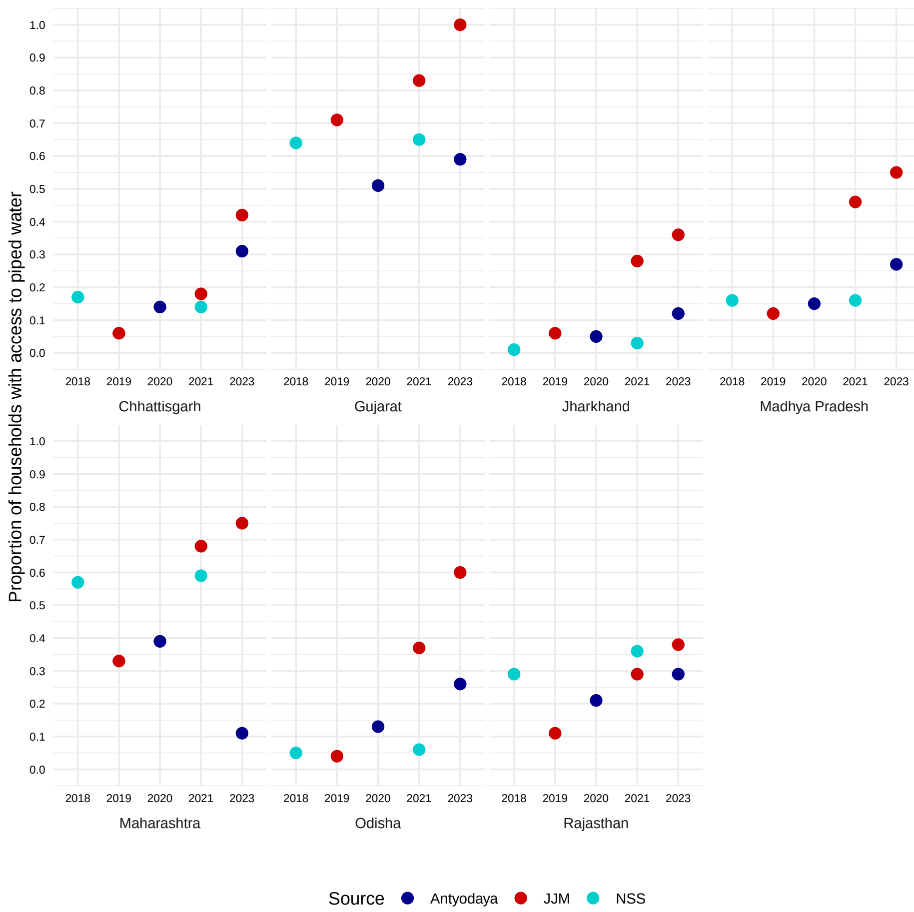


Figure A1: State-wise proportions of households with piped water over time. Each dot represents the number of households with piped water as a proportion of the total number of households surveyed. Light blue dots represent data from reports of the 2018 and 2020 rounds of the National Sample Survey, a district-level representative survey. The 2018 (n=19,924 households) and 2020-21 (n=87,242 households) NSS reports [80, 81] the proportion of households and people, respectively, that report their principal source of drinking water is 'piped water into dwelling, yard or plot' in rural areas. In both years, this is defined as water supplied by a public or private agency to the household premises via pipes that is used as drinking water. Dark blue dots represent data from the 2020 (n=218,408 villages) and 2022-23 (n=213,663 villages) rounds of the Mission Antyodaya survey [68, 78]. In both years, enumerators surveyed village government officials about the number of households with 'piped water connections' as well as the total number of households. Piped water is not specifically defined in the Mission Antyodaya questionnaire. Red dots represent data from annual reports of the Jal Jeevan Mission, downloaded from the e-portal [79]. The annual reports provide the total households and total household piped water connections as of 31 March of 2021 (n=49,379,242 households) and 2023 (n=63,989,931 households). The 2019 JJM data (n=65,857,354 households) is transcribed from the map on the e-portal representing baseline tap water coverage at the launch of mission in August 2019 [79].

Table A1: Census definitions for groundwater sources of drinking water

Source of drinking water	**Census Definition**	**Underlying structure**	**Depth**	**Classification**
Covered well	Well that is covered on the sides, protected from runoff, and covered on the top, protected from animals	Dug well	12m (average of 8–15 $m^{\dagger}$)	Covered well
Uncovered well	Well that is uncovered on the sides, the top, or both	Dug well	12m (average of 8–15 $m^{\dagger}$)	Uncovered well
Hand pump	Hand pump is constructed with a small diameter hole and casing or pipes to protect the source. Groundwater is taken out by manually operating the hand pump.	Borehole	Up to 50 m^{*}	Hand pump
Tube well	Tube well or borehole is constructed with a small diameter hole and casing or pipes to protect the source. Groundwater is taken out using an electric or diesel pump.	Borehole	Depends on the depth of water level in the aquifer*	Machine-pumped water
Tap water from untreated source	Water from a surface or groundwater source provided to households through pipes directly or after pumping to an overhead water tank, or provided by a government body or private agency via pipes within the premises or common taps *without treatment*	Unspecified, depends on whether a surface or groundwater source is used	Depends on source	Machine-pumped water
Tap water from treated source	Water from a surface or groundwater source provided to households by a government body or private agency via pipes within the premises or common taps *after treatment*	Unspecified, depends on whether a surface or groundwater source is used	Depends on source	Machine-pumped water

Notes: We paraphrase census definitions for the sources of drinking water from the instruction manual for the 2011 house listing and housing census[82], from which we also use data for village-level access to various drinking water sources in our analysis. We identify the depths of various structures using the 2013 Operations and Maintenance Manual for Rural Water Supplies*[83] and 2013-14 manual for the Minor Irrigation Census (†)[84]

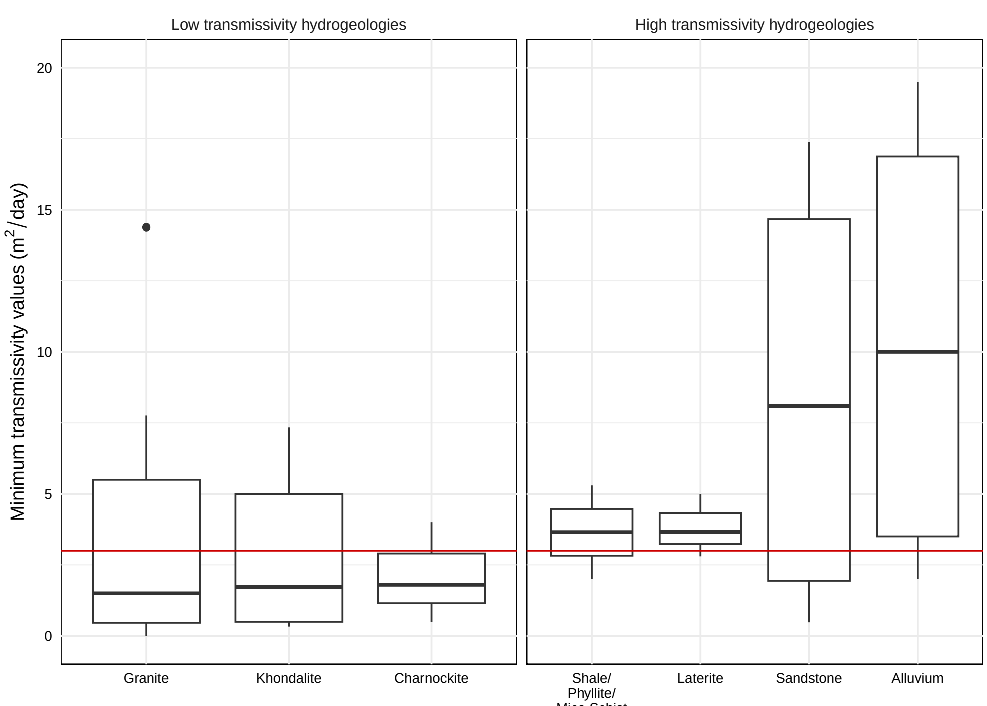
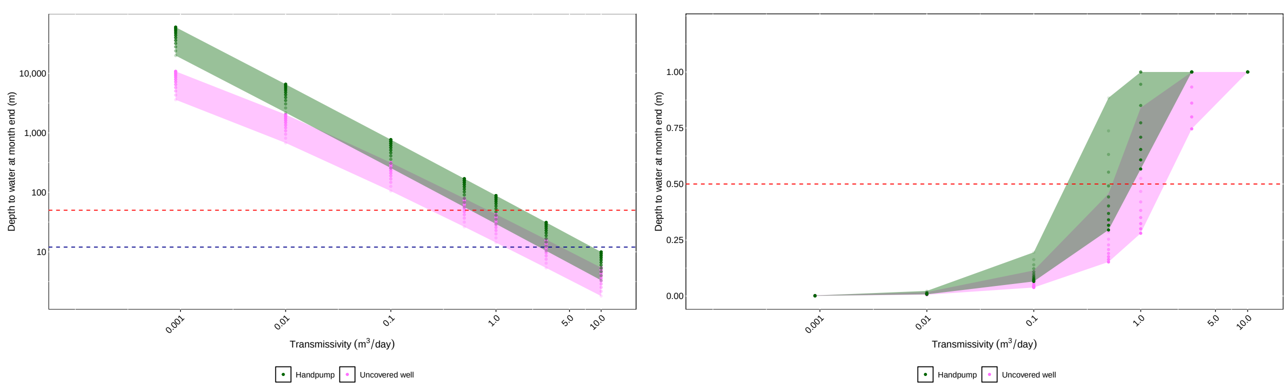


Figure A2: Distribution of minimum tranmissivity values collated from district handbooks of Odisha [66]. The handbooks provide the range of transmissivity values for each hydrogeology in the district from measurements of exploratory wells - we collate the minimums of these. The Y-axis is truncated at 20 due to vast range of T-values for Alluvium.

Table A2: Classification of hydrogeologies in Odisha

	Hydrogeology	Median T-minimum (m^2/day)	Classification
1	BHJ/BCQ	Negligible	Low transmissivity
2	Meta Basic/Meta Volcanic	Negligible	Low transmissivity
3	Granite	1.5	Low transmissivity
4	Khondalite	1.72	Low transmissivity
5	Charnockite	1.8	Low transmissivity
6	Shale/Phyllite/Mica Schist	3.65	High transmissivity
7	Laterite	3.66	High transmissivity
8	Sandstone	15.575	High transmissivity
9	Gneiss	18	High transmissivity
10	Alluvium	57.6	High transmissivity

Notes: We classify hydrogeologies as high or low transmissivity based on transmissivity (T) values from the district geological handbooks of Odisha [66]. The handbooks provide the range of T values for each hydrogeology in the district from measurements of exploratory wells - we collate the minimums of these. As the range can be wide, depending on the number of wells and districts representing each hydrogeology, we only consider hydrogeologies with at least two distinct T values. As such, we drop four hydrogeologies for which we don't have adequate T values - quartzite, granophyre, gabbro-anorthosite, and fruchsite/sericite quarzite. The dropped hydrogeologies constitute $< 5\%$ of the total area under classification. We classify hydrogeologies with a median T value of ≤ 3 as 'low transmissivity' and those with a median of > 3 as 'high transmissivity'. The full range of values can be found in appendix figure A2. For one hydrogeology - gneiss - we were unable to find T values from the district handbooks and instead use the T value provided in the 2012 Aquifer Systems of India atlas [71]



(a) (b)

Figure A3: **Sensitivity to water demand: Simulated drops in water-levels (drawdowns) across transmissivity and water structures, and the latter's adequacy in providing sufficient volumes of drinking water.** (a): Drawdowns at the end of a month based on $\pm 50\%$ of 55 litres per capita requirement for an average-sized village of 750 people. Simulated depths are based on Equation 2 in the Methods section. The blue dotted line denotes an average depth of uncovered wells of 12m and red dotted line denotes average depth of deep handpumps of 50m. (b): Proportions of required daily drinking water met by handpumps and uncovered wells at the end of a month. Simulated volumes are based on Equation 2 in the Methods section. The red dotted line denotes 50% of the total daily requirement.

Table A3: ST villages are less likely to use machine-pumped sources for drinking water

	% Machine-pumped water		% Handpump		% Uncovered Wells		% Covered Wells		% Surface		% Other	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ST village	-0.054***	-0.043***	0.011**	0.009*	0.016***	0.011***	0.001	0.001	0.024***	0.021***	0.002***	0.001**
	(0.004)	(0.004)	(0.005)	(0.005)	(0.003)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)	(0.001)
SC village	0.005**	0.004**	0.017***	0.017***	-0.017***	-0.017***	-0.001***	-0.001***	-0.003***	-0.003***	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Non-ST/SC population	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000	0.000	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total households	0.000***	0.000***	0.000	0.000	0.000	0.000	0.000	0.000	0.000***	0.000***	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Road	0.031***	0.027***	0.000	0.000	-0.009***	-0.007***	-0.002***	-0.002***	-0.020***	-0.018***	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Village area	0.000*	0.000	0.000***	0.000***	0.000**	0.000*	0.000	0.000	0.000**	0.000**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance to district HQ	-0.001***	-0.001***	0.000	0.000	0.000***	0.000***	0.000	0.000	0.000***	0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Forest cover	0.002***	0.002***	-0.005***	-0.005***	-0.001***	-0.001***	0.000	0.000	0.004***	0.004***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ruggedness	-0.005***	-0.004***	-0.004***	-0.004***	0.004***	0.004***	0.000***	0.000***	0.004***	0.004***	0.000*	0.000
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% Electricity		0.109***		-0.015***		-0.051***		0.001		-0.038***		-0.006***
		(0.005)		(0.005)		(0.004)		(0.001)		(0.003)		(0.001)
Sub-district fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable mean	0.269	0.269	0.433	0.433	0.214	0.214	0.020	0.020	0.051	0.051	0.013	0.013
Observations	233,561	233,561	233,561	233,561	233,561	233,561	233,561	233,561	233,561	233,561	233,561	233,561
R^2	0.482	0.487	0.446	0.446	0.339	0.340	0.080	0.080	0.293	0.296	0.242	0.243
Within R^2	0.028	0.037	0.012	0.012	0.007	0.009	0.001	0.001	0.037	0.040	0.001	0.001

Notes: Linear regression of the proportion of households using different sources of drinking water on ST and SC village binaries. Coefficient estimates are based on equation 1 and dataset constructed using SHRUG [31] and Census 2011 [52] data. ST village is a binary indicator for $\geq$ 80% ST population and SC village indicates $\geq$ 20% SC population. Standard errors are clustered at the sub-district levels. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

Table A4: Linear regression of the proportion of households using different sources of drinking water on village-level proportions of ST and SC population

	% Machine-pumped water		% Handpump		% Uncovered Wells	
	(1)	(2)	(3)	(4)	(5)	(6)
ST population share	-0.001***	-0.001***	0.000***	0.000***	0.000***	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SC population share	0.000	0.000	0.001***	0.001***	0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Non-ST/SC population	0.000*	0.000*	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total households	0.000***	0.000***	0.000***	0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Road	0.031***	0.027***	0.001	0.001	-0.009***	-0.007***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Village area	0.000	0.000	0.000***	0.000***	0.000**	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance to district HQ	-0.001***	-0.001***	0.000	0.000	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Forest cover	0.003***	0.003***	-0.005***	-0.005***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ruggedness	-0.005***	-0.004***	-0.004***	-0.004***	0.004***	0.003***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
% Electricity		0.106***		-0.011**		-0.051***
		(0.005)		(0.005)		(0.004)
Sub-district fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable mean	0.269	0.269	0.433	0.433	0.214	0.214
Observations	233,561	233,561	233,561	233,561	233,561	233,561
R^2	0.483	0.487	0.447	0.447	0.339	0.340
Within R^2	0.029	0.037	0.013	0.013	0.007	0.009

Notes: Linear regression of the proportion of households using different sources of drinking water on ST and SC village-level population shared. Coefficient estimates are based on dataset constructed from SHRUG [31] and Census 2011 [52] data. Standard errors are clustered at the sub-district levels. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

Table A5: Fractional logitregression of proportion of households using different sources of drinking water on ST and SC village binaries

	% Machine-pumped water		% Handpump		% Uncovered Wells	
	(1)	(2)	(3)	(4)	(5)	(6)
ST village	-0.529***	-0.436***	0.024	0.017	0.086***	0.056***
	(0.037)	(0.036)	(0.023)	(0.023)	(0.020)	(0.020)
SC village	0.023	0.023	0.082***	0.082***	-0.133***	-0.130***
	(0.014)	(0.014)	(0.012)	(0.012)	(0.014)	(0.014)
Non-ST/SC population	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total households	0.000***	0.000***	0.000*	0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Road	0.237***	0.208***	0.023	0.026*	-0.064***	-0.052***
	(0.021)	(0.020)	(0.016)	(0.016)	(0.016)	(0.016)
Village area	0.000*	0.000	0.000***	0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance to district HQ	-0.006***	-0.005***	0.000	0.000	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Forest cover	0.015***	0.015***	-0.026***	-0.026***	-0.010***	-0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Ruggedness	-0.032***	-0.026***	-0.019***	-0.019***	0.026***	0.024***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
% Electricity		0.987***		-0.076***		-0.326***
		(0.035)		(0.025)		(0.027)
Sub-district fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable mean	0.269	0.269	0.433	0.433	0.214	0.214
Observations	233,557	233,557	233,539	233,539	233,548	233,548

Notes: Coefficient estimates are based on Fractional logitregressions of the proportion of households using different sources of drinking water on ST and SC village binaries. Expressed as marginal effects, after controlling for the availability of electricity, households in ST villages are 1.4% less likely (Model 2) to use machine-pumped water than non-SC/ST villages and 1.3% more likely (Model 6) to use uncovered wells. On the other hand, SC villages are 0.1% more likely (Model 2) to use machine-pumped water than non-SC/ST villages and 2.8% less likely (Model 6) to use uncovered wells. Dataset constructed using geological maps [66], SHRUG [31], and Census 2011 [52] data. Standard errors are clustered at the sub-district level. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

Table A6: ST villages are less likely to use machine-pumped sources for drinking water

	% Machine-pumped water		% Taps		% Tubewells	
	(1)	(2)	(3)	(4)	(5)	(6)
ST village	-0.054***	-0.043***	-0.041***	-0.033***	-0.013***	-0.011***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)
SC village	0.005**	0.004**	0.009***	0.009***	-0.005***	-0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Non-ST/SC population	0.000***	0.000***	0.000***	0.000***	0.000**	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total households	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Road	0.031***	0.027***	0.033***	0.030***	-0.002	-0.003
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Village area	0.000*	0.000	0.000**	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance to district HQ	-0.001***	-0.001***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Forest cover	0.002***	0.002***	0.003***	0.003***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ruggedness	-0.005***	-0.004***	-0.003***	-0.003***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% Electricity		0.109***		0.087***		0.023***
		(0.005)		(0.004)		(0.003)
Sub-district fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable mean	0.269	0.269	0.174	0.174	0.094	0.094
Observations	233,561	233,561	233,561	233,561	233,561	233,561
R^2	0.482	0.487	0.539	0.543	0.253	0.253
Within R^2	0.028	0.037	0.033	0.041	0.003	0.004

Notes: Linear regression of the proportion of households using taps and tubewells on ST and SC village binaries. Coefficient estimates are based on equation 1 and dataset constructed using SHRUG [31] and Census 2011 [52] data. ST village is a binary indicator for $\geq$ 80% ST population and SC village indicates $\geq$ 20% SC population. Standard errors are clustered at the sub-district levels. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

Table A7: ST villages are less likely to have piped water connections

	Model 1			Model 2		
	Census 2011	Antyodaya 2020	Antyodaya 2023	Census 2011	Antyodaya 2020	Antyodaya 2023
	% Pumped water	% Piped water		% Pumped water	% Piped water	
	(1)	(2)	(3)	(4)	(5)	(6)
ST village	-0.054***	-0.034***	-0.046***	-0.043***	-0.034***	-0.045***
	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
SC village	0.005**	0.005***	0.011***	0.004**	0.005***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Non-ST/SC population	0.000***	0.000***	0.000	0.000***	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total households	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Road binary	0.031***	0.034***	0.058***	0.027***	0.032***	0.056***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
Village area (ha)	0.000*	0.000***	0.000***	0.000	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance to district HQ (km)	-0.001***	0.000***	-0.001***	-0.001***	0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Forest cover	0.002***	0.001***	0.000	0.002***	0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ruggedness	-0.005***	-0.002***	-0.002***	-0.004***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% Electricity				0.109***		
				(0.005)		
Village Electrified binary					0.040***	0.058***
					(0.005)	(0.007)
Sub-district fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable mean	0.269	0.194	0.244	0.269	0.194	0.244
Observations	233,561	216,987	212,211	233,561	216,987	212,211
R^2	0.482	0.382	0.306	0.487	0.382	0.307
Within R^2	0.028	0.008	0.014	0.037	0.009	0.015

Notes: Models 2-3 and 5-6 are linear regressions of the proportion of households with piped water connections on ST and SC village binaries. Coefficient estimates are based on equation 1 and dataset constructed using SHRUG [31], Mission Antyodaya 2020 [78], and Mission Antyodaya 2023 [68] data. ST village is a binary indicator for $\geq$ 80% ST population and SC village indicates $\geq$ 20% SC population. Non-ST/SC population is imputed using 2011 population and the growth in the number of households between 2011 and 2020, and 2011 and 2023 respectively. Road is a binary indicator of all-weather road. Village electrified is a binary indicator of whether the village has electricity in 2020 and whether the village receives non-zero hours of domestic electricity in 2023. Forest cover is averaged over 2018-2020 due to lack of data. Models 1 and 2 are the main results presented in Panel A of figure 5 and columns (1) and (2) of appendix table A3, for comparison. Standard errors are clustered at the sub-district levels. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

Table A8: ST villages are less likely to use electricity for lighting

	% Electricity	% Kerosene	Night-time luminosity (logged)
	(1)	(2)	(3)
ST village	-0.095***	0.085***	-0.083***
	(0.004)	(0.004)	(0.006)
SC village	0.004**	-0.005***	0.031***
	(0.002)	(0.002)	(0.004)
Non-ST/SC population	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Total households	0.000**	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Road	0.037***	-0.027***	0.036***
	(0.003)	(0.003)	(0.005)
Village area	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Distance to district HQ	-0.002***	0.001***	-0.005***
	(0.000)	(0.000)	(0.000)
Forest cover	0.000	-0.001***	0.003***
	(0.000)	(0.000)	(0.001)
Ruggedness	-0.006***	0.005***	-0.019***
	(0.000)	(0.000)	(0.001)
Sub-district fixed effects	Yes	Yes	Yes
Dependent variable mean	0.524	0.453	1.377
Observations	233,561	233,561	233,829
R^2	0.589	0.565	0.681
Within R^2	0.065	0.046	0.129

Notes: Linear regression of the proportion of households using different sources of lighting on ST and SC village binaries. Coefficient estimates are based on equation 1. Models 1 and 2 employ dataset constructed using SHRUG [31] and Census 2011 [52] data. Model 3 uses night-time luminosity of villages averaged over 2010-2012 captured by the Defense Meteorological Satellite Program (DMSP) Operation Linescan System (OLS) and available in SHRUG [31]. ST village is a binary indicator for $\geq$ 80% ST population and SC village indicates $\geq$ 20% SC population. Standard errors are clustered at the sub-district levels. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

Table A9: Fractional logitregression of proportion of households using different sources of lighting on ST and SC village binaries

	Electricity (1)	Kerosene (2)
ST village	-0.538***	0.463***
	(0.024)	(0.023)
SC village	0.026**	-0.031***
	(0.010)	(0.010)
Non-ST/SC population	0.000***	0.000***
	(0.000)	(0.000)
Total households	0.000	0.000***
	(0.000)	(0.000)
Road	0.183***	-0.123***
	(0.014)	(0.014)
Village area	0.000***	0.000***
	(0.000)	(0.000)
Distance to district HQ	-0.009***	0.007***
	(0.000)	(0.000)
Forest cover	-0.009***	-0.006**
	(0.003)	(0.003)
Ruggedness	-0.042***	0.032***
	(0.003)	(0.003)
Sub-district fixed effects	Yes	Yes
Dependent variable mean	0.524	0.453
Observations	233,559	233,559

Notes: Coefficient estimates are based on Fractional logitregressions of the proportion of households using different sources of lighting on ST and SC village binaries. Expressed as marginal effects, households in ST villages are 11.1% less likely to electricity than non-SC/ST villages and 10.4% more likely to kerosene. On the other hand, SC villages are 0.6% more likely to use electricity than non-SC/ST villages and 0.7% less likely to use kerosene. Dataset constructed using geological maps [66], SHRUG [31], and Census 2011 [52] data. Standard errors are clustered at the sub-district level. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

Table A10: Linear regression of the proportion of households using different sources of lighting on village-level proportions of ST and SC population

	Electricity (1)	Kerosene (2)	Night-time luminosity (3)
ST population share	-0.002***	0.002***	-0.002***
	(0.000)	(0.000)	(0.000)
SC population share	-0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Non-ST/SC population	0.000*	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Total households	0.000***	0.000**	0.000***
	(0.000)	(0.000)	(0.000)
Road	0.035***	-0.025***	0.034***
	(0.003)	(0.003)	(0.005)
Village area	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Distance to district HQ	-0.002***	0.001***	-0.005***
	(0.000)	(0.000)	(0.000)
Forest cover	0.000	-0.002***	0.003***
	(0.000)	(0.000)	(0.001)
Ruggedness	-0.006***	0.005***	-0.018***
	(0.000)	(0.000)	(0.001)
Sub-district fixed effects	Yes	Yes	Yes
Dependent variable mean	0.524	0.453	1.377
Observations	233,561	233,561	233,829
R^2	0.593	0.569	0.683
Within R^2	0.074	0.054	0.134

Notes: Linear regression of the proportion of households using different sources of lighting on ST and SC village-level population shares. Models 1 and 2 employ dataset constructed using SHRUG [31] and Census 2011 [52] data. Model 3 uses night-time luminosity of villages averaged over 2010-2012 captured by the Defense Meteorological Satellite Program (DMSP) Operation Linescan System (OLS) and available in SHRUG [31]. Standard errors are clustered at the sub-district levels. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

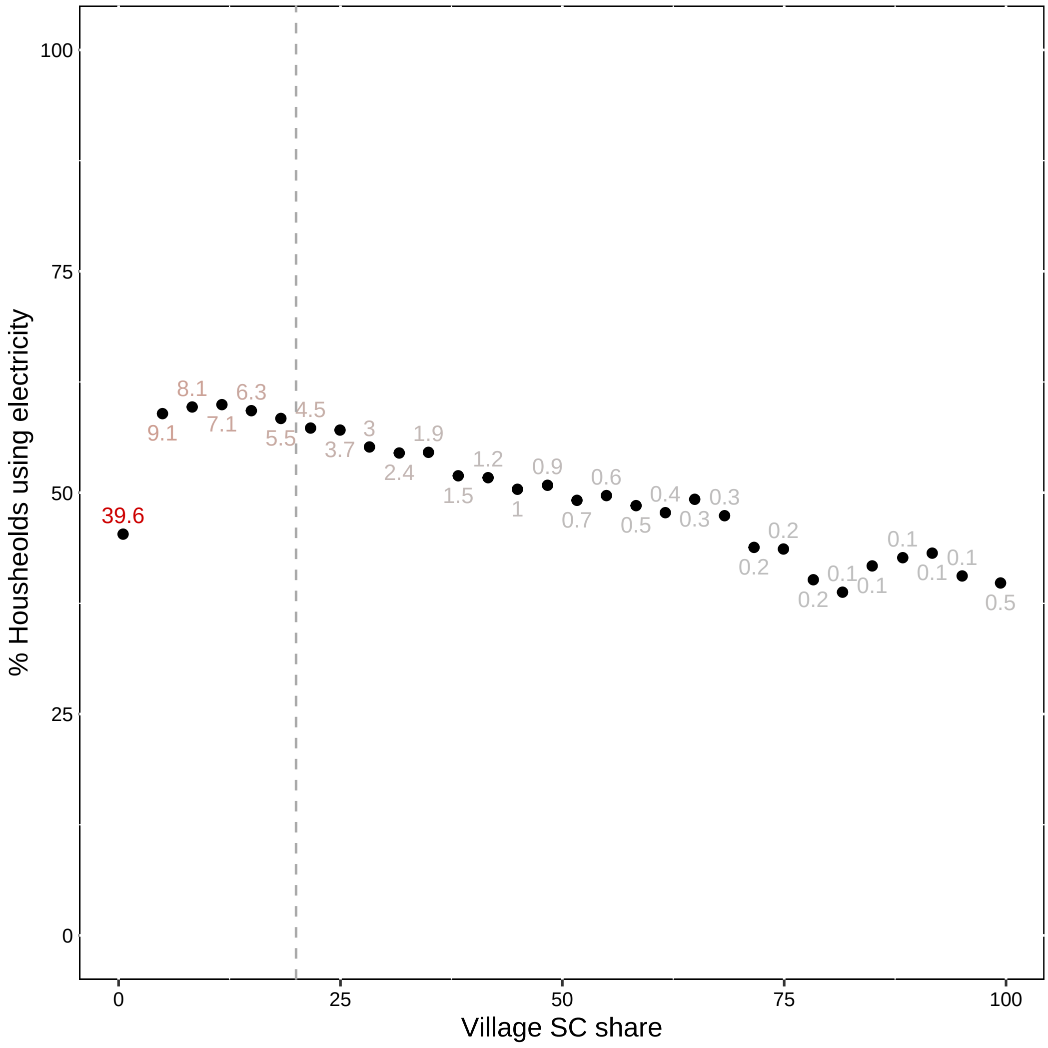


Figure A4: Binned scatter-plot of village-level proportions of SC population and households using electricity using SHRUG [31] and Census 2011 [52] data. Data labels indicate the proportion of total villages represented by the point.

Table A11: Presence of low transmissivity hydrogeologies does not explain the ST village penalty in machine-pumped sources for drinking water or electricity for lighting

	% Machine-pumped water				% Electricity	
	(1)	(2)	(3)	(4)	(5)	(6)
Low transmissivity	0.008	0.011	0.006	0.009	-0.028***	-0.033***
	(0.007)	(0.007)	(0.009)	(0.009)	(0.006)	(0.007)
ST village	-0.027***	-0.019***	-0.028***	-0.020**	-0.095***	-0.095***
	(0.006)	(0.006)	(0.009)	(0.009)	(0.006)	(0.010)
SC village	0.000	0.002	-0.001	0.001	-0.017***	-0.023***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.003)	(0.004)
Low transmissivity x ST village			0.003	0.003		0.003
			(0.011)	(0.011)		(0.012)
Low transmissivity x SC village			0.005	0.003		0.017***
			(0.009)	(0.009)		(0.006)
Non-ST/SC population	0.000	0.000**	0.000	0.000**	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total households	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Village area	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Road	0.014**	0.012*	0.014**	0.012*	0.033***	0.032***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Distance to district HQ	-0.001***	0.000***	-0.001***	0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Forest cover	-0.001**	-0.001*	-0.001**	-0.001*	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Ruggedness	-0.001	-0.001	-0.001	-0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
% Electricity		0.083***		0.083***		
		(0.012)		(0.012)		
Sub-district fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable mean	0.245	0.245	0.245	0.245	0.301	0.301
Observations	43,173	43,173	43,173	43,173	43,173	43,173
R^2	0.137	0.139	0.137	0.139	0.515	0.515
Within R^2	0.008	0.011	0.008	0.011	0.083	0.083

Notes: Linear regression of the proportion of households using machine-pumped water and electricity on ST village and SC village, and low transmissivity binaries. Coefficient estimates are based on equation 8 and dataset constructed using geological maps [66], SHRUG [31], and Census 2011 [52] data. ST village is a binary indicator for $\geq$ 80% ST population and SC village indicates $\geq$ 20% SC population. Low transmissivity corresponds to villages with a majority share under low transmissivity hydrogeologies. Standard errors are clustered at the sub-district levels. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

Table A12: ST villages are less likely to use machine-pumped water for drinking water across all administrative levels

	% Machine-pumped water					
	OLS			Logit		
	No FE	District FE	Sub-district FE	No FE	District FE	Sub-district FE
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.163***			-1.578***		
	(0.007)			(0.037)		
ST village	-0.019***	-0.020**	-0.019***	-0.126***	-0.132***	-0.124***
	(0.005)	(0.008)	(0.006)	(0.028)	(0.045)	(0.038)
SC village	-0.001	0.002	0.002	-0.003	0.009	0.007
	(0.004)	(0.006)	(0.004)	(0.021)	(0.034)	(0.025)
Low transmissivity	0.039***	0.009	0.011	0.212***	0.051	0.067
	(0.004)	(0.009)	(0.007)	(0.020)	(0.050)	(0.044)
% Electricity	0.148***	0.101***	0.083***	0.793***	0.531***	0.453***
	(0.007)	(0.017)	(0.012)	(0.037)	(0.086)	(0.064)
Non-ST/SC population	0.000	0.000	0.000**	0.000	0.000**	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total households	0.000***	0.000***	0.000***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Village area	0.000	0.000	0.000***	0.000	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Road	0.011**	0.016*	0.012*	0.066***	0.098*	0.076*
	(0.004)	(0.008)	(0.006)	(0.025)	(0.049)	(0.040)
Distance to district HQ	0.000	0.000	0.000***	0.000	-0.001	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Forest cover	0.000	0.000	-0.001*	0.000	-0.001	-0.007**
	(0.000)	(0.001)	(0.001)	(0.002)	(0.005)	(0.004)
Ruggedness	-0.001	-0.001	-0.001	-0.003	-0.007	-0.006
	(0.000)	(0.001)	(0.001)	(0.002)	(0.007)	(0.005)
District fixed effects		Yes			Yes	
Sub-district fixed effects			Yes			Yes
Dependent variable mean	0.245	0.245	0.245	0.245	0.245	0.245
Observations	43,173	43,173	43,173	43,173	43,173	43,173
R^2	0.023	0.066	0.139			
Within R^2		0.013	0.011			

Notes: OLS and fractional logit regressions of the proportion of households using machine-pumped water on ST village, SC village, and low transmissivity binaries with different levels of fixed effects. Models 1 and 4 have no fixed effects, models 2 and 5 include district fixed effects, and models 3 and 6 include sub-district fixed effects. Models 1-3 are linear regressions estimated using equation 8 and dataset constructed using geological maps [66], SHRUG [31], and Census 2011 [52] data. Models 4-6 are corresponding Fractional logitregressions. ST village is a binary indicator for $\geq$ 80% ST population and SC village indicates $\geq$ 20% SC population. low transmissivity corresponds to villages with a majority share under low transmissivity hydrogeologies. Standard errors are clustered at the same level as fixed effects. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

Table A13: ST villages are less likely electricity for lighting across all administrative levels

	% Electricity					
	OLS			Logit		
	No FE	District FE	Sub-district FE	No FE	District FE	Sub-district FE
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.412***			-0.236***		
	(0.004)			(0.038)		
ST village	-0.109***	-0.096***	-0.095***	-0.791***	-0.808***	-0.819***
	(0.003)	(0.014)	(0.006)	(0.035)	(0.118)	(0.060)
SC village	0.007**	-0.006	-0.017***	0.015	-0.049	-0.091***
	(0.003)	(0.007)	(0.003)	(0.023)	(0.032)	(0.017)
Low transmissivity	-0.081***	-0.037***	-0.028***	-0.408***	-0.218***	-0.193***
	(0.002)	(0.011)	(0.006)	(0.022)	(0.070)	(0.040)
Non-ST/SC population	0.000***	0.000***	0.000***	0.001***	0.000**	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total households	0.000***	0.000	0.000***	-0.001***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Village area	0.000***	0.000***	0.000***	-0.001***	-0.001***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Road	0.038***	0.037***	0.033***	0.220***	0.232***	0.211***
	(0.003)	(0.008)	(0.005)	(0.030)	(0.056)	(0.034)
Distance to district HQ	-0.001***	-0.001***	-0.001***	-0.005***	-0.005***	-0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Forest cover	0.002***	-0.002***	-0.002***	0.014***	-0.016***	-0.015***
	(0.000)	(0.001)	(0.000)	(0.002)	(0.004)	(0.003)
Ruggedness	-0.006***	-0.001	0.000	-0.042***	-0.011	-0.009
	(0.000)	(0.001)	(0.001)	(0.003)	(0.008)	(0.005)
District fixed effects		Yes			Yes	
Sub-district fixed effects			Yes			Yes
Dependent variable mean	0.301	0.301	0.301	0.301	0.301	0.301
Observations	43,173	43,173	43,173	43,173	43,173	43,171
R^2	0.287	0.440	0.515			
Within R^2		0.121	0.083			

Notes: OLS and fractional logit regressions of the proportion of households using electricity on ST village, SC village, and low transmissivity binaries with different levels of fixed effects. Models 1 and 4 have no fixed effects, models 2 and 5 include district fixed effects, and models 3 and 6 include sub-district fixed effects. Models 1-3 are linear regressions estimated using equation 8 and dataset constructed using geological maps [66], SHRUG [31], and Census 2011 [52] data. Models 4-6 are corresponding Fractional logitregressions. ST village is a binary indicator for $\geq$ 80% ST population and SC village indicates $\geq$ 20% SC population. low transmissivity corresponds to villages with a majority share under low transmissivity hydrogeologies. Standard errors are clustered at the same level as fixed effects. Significance codes: ***: 0.01, **: 0.05, *: 0.1.