Energy Policy 145 (2020) 111746

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

Crowdsourcing data on the reliability of electricity service: Evidence from a telephone survey in Uttar Pradesh, India[☆]

Ryan Kennedy[a,*], Aseem Mahajan[b], Johannes Urpelainen[c]

[a] University of Houston, United States of America
[b] Harvard University, United States of America
[c] Johns Hopkins SAIS, United States of America

ARTICLE INFO

Keywords:
Electrification
Energy access
India
Crowdsourcing
Mobile phones

ABSTRACT

Measuring energy access in developing countries involves much more than simply recording whether or not households are connected to the grid. Both international organizations and scholars now recognize the importance of reliable electricity supply for achieving positive development outcomes. Yet, measuring reliability is much more difficult than measuring the existence of connections. We propose an economical croudsourcing method for measuring reliability, and compare this method to energy monitor data for 122 households over 12 months. The results suggest that, while far from perfect, crowdsourcing provides a reasonably accurate method for monitoring the reliability of access over time, especially when modeled as a non-linear relationship. We apply these findings to model energy reliability in a broader group of villages across Uttar Pradesh, India, demonstrating the existence of disparities between urban and rural reliability and seasonal fluctuations in reliability. The system laid out in this study can be utilized by government and non-government organizations to quickly and cheaply monitor energy reliability.

1. Introduction

Efforts to measure progress in energy access play a critical role in the quest for sustainable energy for all. In the field of rural electrification, almost one billion people globally remain without electricity at home (IEA, 2018). What is more, the reliability of electricity service remains poor across many developing countries, reducing the social and economic benefits of electric connections (Chakravorty et al., 2014; Aklin et al., 2016b; Allcott et al., 2016). In India, there have been massive gains in electricity access over the past decade, culminating with the Indian government announcing 100% village electrification in 2018.[1] But scholars worry that these gains may be subject to backsliding, especially for low-income groups (Aklin and Urpelainen, 2020), and the reliability of these new connections may be suspect (Phadke et al., 2019).

Unfortunately, measuring reliability can be quite challenging. While household surveys can be useful for estimating household connections, they offer only limited insights into the reliability of service. Cross-sectional surveys reflect recent situation, but fail to give an adequate understanding of reliability of service across seasons and over longer time spans. On the other hand, technical measurement of reliability is expensive and difficult in remote rural areas. In the developing world, few governments collect and publish data on electricity service at the local level, and the data that is available from utilities can be misleading.

To address this problem, we explore the potential of crowdsourcing to measure the reliability of electricity service. While the concept of crowdsourcing – gathering information electronically from disparate individuals or groups – has been applied in a range of fields, from tracking illness outbreaks (Smolinski et al., 2015; Baltrusaitis et al., 2017) to preventing civil conflict (Heinzelman et al., 2011), there have been few examples of this type of work for measuring public service provision in the developing world (Post et al., 2018). Leveraging modern telecommunications technology, we are now able to quickly and repeatedly survey a large number of households about the reliability of their electricity services. Moreover, we can compare these reports to real-time monitoring of reliability to test the degree to which such crowd-sourced measures correspond to ground-truth.

We report results from two studies for measuring the reliability of rural electricity service in the state of Uttar Pradesh, India. In the first

☆ The authors would like the thank Michael Aklin and Maryam Shiran, as well as the participants in the Initiative for Sustainable Energy Policy (ISEP) virtual workshop, for their comments on previous drafts. All errors are the authors'.

* Correspondence to: 447 Philip G. Hoffman (PGH) Hall, Houston, TX 77204, United States of America.
E-mail address: rkennedy@uh.edu (R. Kennedy).

[1] See https://saubhagya.gov.in/ for details.

https://doi.org/10.1016/j.enpol.2020.111746
Received 10 February 2020; Received in revised form 6 June 2020; Accepted 5 July 2020
Available online 16 July 2020

study, we conduct short telephone interviews on a weekly basis with a sample of 122 households from 12 villages in the Sitapur district over a period of 12 months starting in October 2017. In these mobile-phone-based surveys, we ask respondents to describe their electricity supply over the past 24 h. Responses are compared with actual electricity service data from the Prayas Energy Group's Energy Supply Monitoring Initiative (ESMI), which records voltage by the minute using Electricity Supply Monitors (ESMs) in households, farms, and small commercial establishments all over India.

We find that respondents have difficulty assessing the reliability of electricity service over the previous 24 h when daily hours of power are low. In the low- to mid-levels of electricity availability, respondents tend to underestimate the power that is available. But as daily hours of power increase, respondents begin to notice a clear difference in the reliability of service. Overall, the correlation between reported and true hours of supply is about 0.194, which is not strong, but is still quite informative, especially at distinguishing those areas with reliable power from those with less-reliable power. We find that this can be improved substantially by explicitly incorporating the information about the non-linear pattern between the survey responses and the actual supply.

Building on these findings, we survey a representative sample of 1800 households across 96 villages and 96 urban wards in 24 districts of Uttar Pradesh. Using both the linear and non-linear models developed on the Sitapur study sample, we estimate the reliability of power across seasons and compare the differences between rural and urban areas. We detect significantly fewer problems with power reliability in rural areas across the districts, as well as decreased reliability during monsoon seasons. These reliability problems are much less pronounced in urban areas, and we do not see the same dropoff in reliability during the monsoon season. Both of these findings support the concerns of scholars that rural areas suffer from a large disparity in electricity reliability that is especially bad during the monsoon season, when high demand and poorly planned load-shedding increase power outages (Alam, 2014; Sharma et al., 2018; Conevska and Urpelainen, 2020). Moreover, once we account for the non-linear pattern of reporting observed in Sitapur, these disparities become even more severe.

On balance, the results are encouraging insofar as they show that crowdsourcing can be used to detect general variation in reliability of electricity service. While not perfect, we managed to collect reasonably useful data on the reliability of electricity service over an entire year in ten districts at a very low cost. While respondents were able to assess the difference between high and low daily hours, they had difficulty gauging exact values when supply was low. Conversely, the results are discouraging insofar as they suggest that the reliability of rural electrification may lag even further behind urban supply than previous surveys indicate.

More broadly, the results offer insights into the potential that crowdsourcing has in data collection. Our results suggest that, even for something that is imperfectly observed, crowdsourcing can be used as an economical method for estimating the general outlines of the policy challenge, while they also show the utility of having ground truth data against which to develop models of crowdsourcing biases.

This study proceeds in four parts. First, we will discuss the importance of and some of the challenges in measuring reliability. Next we will present the design of this research, including a description of the sampling and analysis methods for both our crowdsourcing exercise and our attempt to generalize those results. Third, we will interpret the outcomes of both analyses. Finally, we will conclude with a discussion about the opportunities and challenges of extending the methods analyzed herein, as well as the policy implications.

2. Measuring the reliability of electricity service

As global efforts to eliminate energy poverty have expanded, policy-makers have begun to recognize that energy access is multidimensional. The Multi-Tier Framework (MTF) for monitoring and evaluating energy access, defines energy access as "the ability to obtain energy that is adequate, available when needed, reliable, of good quality, convenient, affordable, legal, healthy, and safe for all required energy services" (Bhatia and Angelou, 2014, 3).

This presents a challenge for scholars and policy-makers attempting to track expanding energy access. It is quite easy to measure the prevalence of household electricity connections (Aklin et al., 2016b). Surveys using representative samples can give an accurate understanding of the percentage of rural and urban households that have electricity connections, be it through the electric grid or using decentralized energy. Similarly, government statistics on connections are more likely to be accurate, since the incentives of governments and electricity companies are relatively aligned when it comes to reporting. Governments are interested in showing expanding access and companies are interested in tracking their consumers.

Measuring the reliability of electricity service is far more difficult. Most measurements of electricity reliability come in the form of surveys. Some surveys ask about the number of blackouts within a particular timeframe (Gibson and Olivia, 2010). Others ask respondents to recall approximately how many hours a day they usually have electricity or how much is available during nighttime hours (Kennedy et al., 2019). Such survey measurements are difficult to take at face value. The limits of human memory are well-known, and more recent information tends to cloud the general memory of events ("recency bias"). This makes asking about general reliability difficult, as transitory events are likely to drive results. Not surprisingly, for example, simply surveying customers on their planned usage has proven to be a relatively inaccurate method for predicting customer behavior compared to drawing inferences from the behavior of existing customers (Blodgett et al., 2017).

Government and company data is also more difficult to find and may not be accurate. Governments are less inclined to report reliability issues, especially when the expansion of connections has been a key political promise of an incumbent. Politicians often ignore, or even encourage, practices that undermine reliability, like electricity theft, to improve their near-term electoral prospects (Min and Golden, 2014; Baskaran et al., 2015). Companies also have few incentives to publicize problems in electricity delivery, especially poorly planned load-shedding, which might lead to more coordinated demands for greater investment in poorer, less profitable areas. For example, in 2017, large disparities were noted between official government statistics in Uttar Pradesh, India and the information collected by Prayas Energy Group, an NGO that had set up Electricity Supply Monitors (ESMs) throughout the region. While government data showed nearly no power cuts in the region, the Prayas monitors indicated regular power cuts that lasted between two to nine hours a day.[2] More recent data from Prayas similarly suggests that rural households in Uttar Pradesh had just over 14 h of daily service on average.[3]

And yet, reliability of electricity service is critically important. Existing scholarship suggests that poor reliability can negatively impact demand for electricity (Kemmler, 2007), household satisfaction with electricity (Aklin et al., 2016b), gains in non-agricultural income from electrification (Chakravorty et al., 2014), and willingness to pay for electricity (Kennedy et al., 2019).

[2] Debjoy Sengupta, 23 February 2017, "Wide power deficit discrepancy between NGO and govt data in Uttar Pradesh", *The Economic Times*, https://economictimes.indiatimes.com/industry/energy/power/wide-power-deficit-discrepancy-between-ngo-and-govt-data-in-uttar-pradesh/articleshow/57311992.cms?from=mdr.

[3] Daily average hours of supply in Uttar Pradesh, Electricity Supply Monitoring Initiative, Prayas Energy Group, https://www.watchyourpower.org/analysis_dashboard.php (accessed June 29, 2019).

In this study, we explore the feasibility of using crowdsourcing for measuring reliability. While crowdsourcing has not been used extensively in studying electricity service reliability, it has been applied to a variety of other domains, and all of us have likely interacted with a crowdsourcing application. Online question answering forums, like Quora and StackOverflow rely on crowdsourcing to answer questions on general knowledge and computer programming respectively. The commercial driving application, Waze, uses crowdsourcing to track traffic speeds and report accidents. Perhaps the best-known application of crowdsourcing for tracking issues of public concern is the "Flu Near You" project, where people can report their symptoms in real time using a simple weekly survey to track flu outbreaks.[4] This work has been extensively studied (Smolinski et al., 2015; Baltrusaitis et al., 2017), and is used by the U.S. Centers for Disease Control (CDC), among others, to more quickly respond to illness outbreaks. Similar efforts have been made to label fake news on the internet (Warkentin et al., 2010; Tschiatschek et al., 2018) and track outbreaks of civil violence and unrest (Van der Windt and Humphreys, 2016), among other applications. In developing countries, Post et al. (2018) show how crowdsourced data can be utilized to better understand the politics of urban water delivery.

In this article, we contribute to the study of crowdsourcing methods by exploring how they can be applied in the case of electricity service reliability. We utilize economical cell-phone based surveys to track power reliability over time. Allowing for responses to be collected on a daily basis dramatically reduces the reliance on recollection of respondents. We then compare this against ground truth data collected by electricity monitors. These results allow for a cheap and detailed overview of the state of energy reliability. This data, if it proves to be even moderately accurate, would provide a method for public policy advocates and governments to discover where issues of reliability remain problematic and when such issues are most severe. If measuring a problem is the first step to addressing it, this could provide the necessary information for ensuring that the reliability requirements of the MTF are being met.

3. Research design

We conduct this research in two parts. First, we attempt to build a model that links self-reported energy reliability to actual reliability. To do this, we conduct an intensive survey of residents in 12 villages in the Sitapur district of Uttar Pradesh. This involved conducting weekly surveys over an extended period of time, along with a baseline study of the respondents' characteristics. By comparing the results of the surveys against monitoring data, we create a strong over-time comparison set on which we can test the correspondence between reported and actual reliability. We can also test for heterogeneity based on time and user characteristics.

Second, starting from these models, we leverage a much larger, but less information-rich, survey conducted by the International Institute for Sustainable Development (IISD), with assistance from Columbia University's Center for Global Energy Policy (CGEP). Applying the models developed in Sitapur, we are able to produce reliability estimates across 96 villages and 96 urban wards in 24 districts of Uttar Pradesh.

While the exact cost of technical monitoring and crowdsourcing depends on context, in this project the difference was substantial. Installing each monitor cost us approximately USD 250 that included the technology, data server, and fieldwork. In addition, maintenance of the monitors for the 12 villages cost over USD 1000 for a year. In contrast, the entire crowdsourcing exercise for 1280 households across 12 districts of Uttar Pradesh cost us only about USD 3000 for 12 months. Although average monitoring costs would decrease with scale, the difference would be measured in orders of magnitude.

3.1. Comparing self-reported reliability data to technical measures

We begin this analysis by looking at the validity of crowdsourcing measures of electricity reliability in a relatively small subset of villages over time. To measure the actual reliability of electricity supply in each village, we rely on monitors installed by the Prayas Energy Group (Prayas), a non-governmental and non-profit organization operating out of Pune, India. Through ESMI, Prayas installed ESM plug-in devices that record voltages on an hourly basis. To compare participants' self-reported assessments of the reliability of electricity supply to actual reliability in Uttar Pradesh's Sitapur district, shown in Fig. 1, we focus on 12 villages covered by ESMI.[5]

Between September 29 and October 6, 2017, enumerators from Morsel India, an Uttar Pradesh-based survey company, recruited a random sample of 10 households with grid electricity from each village. Each of the 120 households was given INR 100 (approximately USD 1.50) for participating in a 15-minute in-person baseline survey, which collected respondents' mobile phone numbers, socioeconomic and demographic information, and data on households' sources of electricity, lighting and their reliability.

To assign the dates on which households would receive calls to administer the mobile phone survey, we first randomly assigned each of the 120 households into two groups, with one receiving calls in odd weeks and the other in even weeks. With data collected over 48 weeks, each group comprised 60 × 24 household-week combinations. To balance out the days of the week when each household would participate, which would likely affect their responses, we stratified the sample by households and then randomly assigned each day of the week, between Monday and Saturday, to four sets of the 24 weeks in which the household would be surveyed. The assignment of dates to households is illustrated in Fig. 2.

Between October 30, 2017 and September 29, 2018, enumerators administered a six-question survey to assigned households collecting data on hours of grid electricity used, hours available, voltage levels, voltage fluctuations, and household satisfaction over the 24 h prior to the call. To increase the likelihood of reaching a member of the household, enumerators placed calls at night after working hours, or in the early morning following the date of the assigned treatment.

Once the survey data and the Prayas data are combined, we have 1673 complete observations.[6] These observations are nested within 122 surveyed households. Because we expect households to demonstrate heterogeneity in their estimates of electricity reliability, we model this as a hierarchical (multilevel) process, with the actual number of hours available being a function of the reported hours available, along with a random component by household (Gelman and Hill, 2006).[7] Put more formally, we model this process as

$$y_{ij} = \beta_{1j} + \beta_2(Reported\ Hours\ Available)_i + \zeta_j \quad (1)$$

where y_{ij} is the actual hours available according to the Prayas monitors and ζ_j is the random intercept based on the household j. By assumption, $\zeta_j \sim N(0, \phi)$. Our interest is in the fixed component, which is the general relationship between the reported hours and the monitored hours, once the random household heterogeneity is accounted for.

We recognize, however, that the main threat to valid inference between reported hours and actual hours available, may be systematic, rather than random by household. There may be systematic nonlinearities that make the linear estimates from the hierarchical model

[4] https://flunearyou.org/.

[5] The 12 villages were Dharampur, Inchauli, Jhauwa Khurd, Jyotishah Alampur, Kahmria Kathura, Kankari, Khindaura, Manwan, Mukimpur, Muradpur, Tedwa Deeh, and Thangaon.

[6] In some cases, monitor data from Prayas was not available because of monitor failure.

[7] We also tested a random component for the village, but this did not capture significant variance, and, thus, was excluded from our final model.

Fig. 1. Uttar Pradesh's Sitapur district. This is the area in which households' self-reported assessment of electricity reliability was compared to data collected from Prayas ESMs, installed in 12 villages.

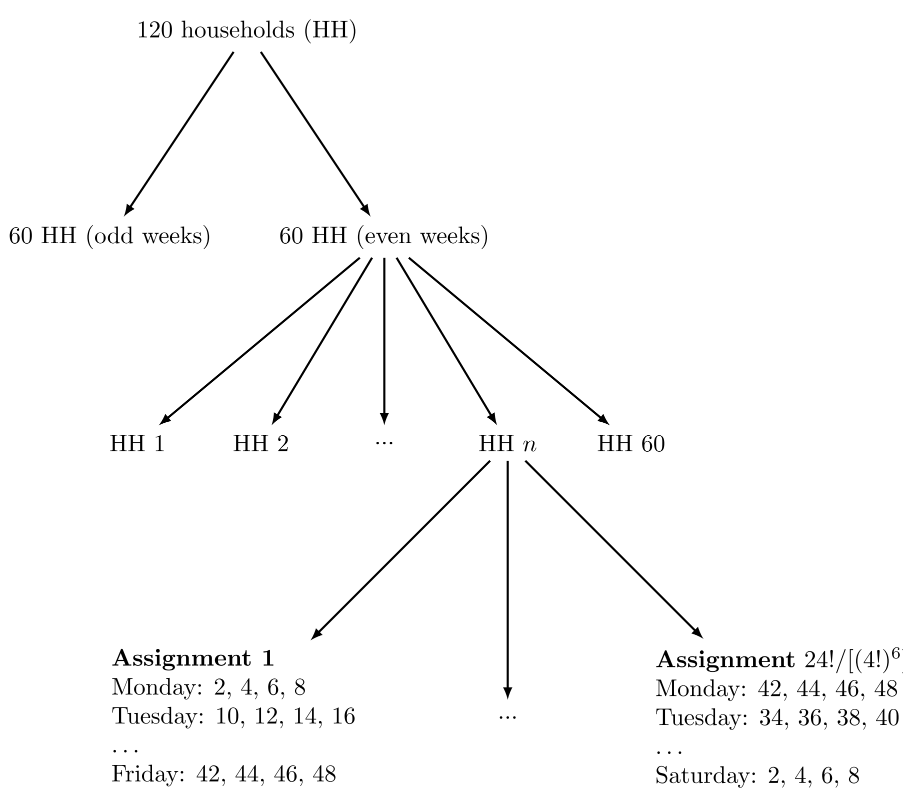

Fig. 2. Illustration of randomized block design assigning dates to households in Sitapur, with household as a blocking factor.

inaccurate. To test for this, we also test a local regression model (LOESS) (Cleveland and Devlin, 1988), where the linear model is estimated locally over a fraction of the dataset determined by the smoothing parameter α. We experimented with several different levels of α, settling on a value of $\alpha = 0.75$.[8]

Combining these two analyses allows us to compare assumptions about the data generating process from our surveys, i.e. whether respondents are accurate across the range of reliability, with inaccuracies primarily being randomly derived across households, or whether there are systematic nonlinearities in the relationship between respondents' forecasts and the actual hours available.

3.2. Extending the model to a broader sample

We extend the insights drawn from the survey conducted in Sitapur to a larger and more representative sample of households across Uttar Pradesh. Our sampling frame is a set of households surveyed by the IISD, with assistance from Columbia University's CGEP, as part of a study on electricity sector reform in Uttar Pradesh (Sharma et al., 2018).

To obtain a representative set of households, the 75 districts in Uttar Pradesh were separated into four groups of approximately equal population corresponding to the north, south, east, and west of the state. Three districts were then randomly selected from each group for a total of 12 districts, shown in Fig. 3. The villages in each district were then classified as "small" or "large" and then divided into two groups of approximately equal total population, with the "small" group composed of many small villages and the large group composed of fewer but larger ones. From each of the two groups within the 12 districts, four villages were randomly selected producing a sample of 96 rural villages (4×3×2×4). From each of the 96 villages, the enumerators surveyed 10 households for a total of 960 rural households.

Urban villages were sampled from the 12 chosen districts. Within each district, 120 wards were randomly selected from urban areas in proportion to the size of the urban areas. Within each urban area, 96 wards were then selected, with the probability of inclusion determined by the in-town population of the ward population over the total town-wide population. Within each sampled ward, enumerators selected households using systematic random sampling.[9] In total, 957 urban households were selected across the 96 urban wards.

Of the 1917 households in the sampling frame, we eliminated those that did not provide phone numbers in the original survey, reducing the sampling frame to 1280 households. Dividing households on the basis of their districts and their classification as urban or rural (e.g., District 133 — Rural) produced 24 *district-type blocks*, summarized in Table 1, of which nine included districts with fewer than 40 households. All households in each of these district-type blocks were included in the sample, resulting in 210 urban households and 25 rural households.

From each of the 11 rural and four urban district-type blocks with at least 40 households, a random sample of 40 was drawn, increasing the total to $440 + 25 = 465$ rural and $160 + 210 = 370$ urban households. Then, from the 11 districts with more than 40 households and the four with more than 40 urban households, 15 additional rural households and 110 additional urban ones were selected.[10] The final distribution of rural and urban households by district-type blocks is given in Table 2 and Fig. 4 summarizes the sampling procedure.

[8] As the reader will see below, there is one main area of nonlinearity in the results, which makes the LOESS relatively invariant to settings of α. This process is not, however, well-represented by a simple quadratic regression, as it tends to underestimate the degree of inflection.

[9] In some cases, the enumerators chose the sixth household from a random starting point, and in others, they sought permission to survey odd households.

[10] One rural household was chosen from each of seven districts and two were chosen from the remaining four districts; and 27 urban households were selected from each of two districts and 27 from the remaining two districts.

Table 1
Summary of total households and sampled households from district-type blocks. *Total* indicates the number of urban and rural households with phone numbers and grid electricity in each district and *Within* indicates the number of households sampled from within the district. In cases where district-type blocks were composed of fewer than 40 households (italicized in *Total*), all households in the block were included and additional households were drawn from corresponding urban/rural district-type blocks with more than 40 households to ensure a total sample size of 480 rural households and 480 urban households. The number of households drawn from outside a given district to reach the requisite sample side is given in the column *Outside*.

District	Rural			Urban		
	Total	Within	Outside	Total	Within	Outside
133	61	40	0	143	40	0
142	48	40	0	93	40	0
144	48	40	0	*39*	39	1
148	55	40	0	*21*	21	19
166	53	40	0	80	40	0
175	42	40	0	155	40	0
177	55	40	0	*33*	33	7
180	*25*	25	15	*26*	26	14
184	50	40	0	*10*	10	30
191	61	40	0	*32*	32	8
195	57	40	0	*29*	29	11
200	44	40	0	*20*	20	20
	599	465	15	681	370	110

Table 2
Distribution of urban and rural households by district.

District	Rural	Urban
133	41	68
142	41	68
144	42	39
148	41	21
166	41	67
175	41	67
177	42	33
180	25	26
184	42	10
191	42	32
195	41	29
200	41	20

The set of unsampled households – 119 rural ones and 201 urban ones – were employed as alternates in cases where responses from sampled households were missing (e.g., cases where the sampled household was unavailable or chose not to participate in the survey). Missing observations from a given district-type block were first replaced with those obtained from alternates in the same district-type block. Upon using all alternates from a given urban or rural district, alternates were chosen from the urban or rural district with the greatest number of remaining alternates available. Alternate households were then contacted in all subsequent weeks over the duration of the survey.

To assign the dates on which households would be called, households in each of the 24 district-type blocks were first randomly assigned to one of the eight weeks over which the first calls would be made. Because households were contacted every eight weeks, the week of their first assigned call determined the weeks of their subsequent five calls. Then, blocking by week and type, households were assigned days, ranging between Monday and Saturday, for each of their given weeks over the period between Monday, October 30, 2017 and Saturday September 29, 2018. Using the weeks as blocking factors to assign days allowed for calls to be evenly distributed across days over each week.

Our sampling strategy for the broader sample has several advantages. First, because it begins with a statistically representative sample of both rural and urban households from a previous project, we can have confidence in any state-level inferences drawn from the analysis. Second, even though we have phone numbers for only 1280 households, we can easily verify that these are spread across the districts under study, so that the geographic scope of the sample remains

Fig. 3. Map of 12 districts from which households rural and urban households were sampled for participation in the survey. The sampling frame is a subset of 1280 households, chosen from 1917 that Sharma et al. (2018) sampled from the 12 districts.

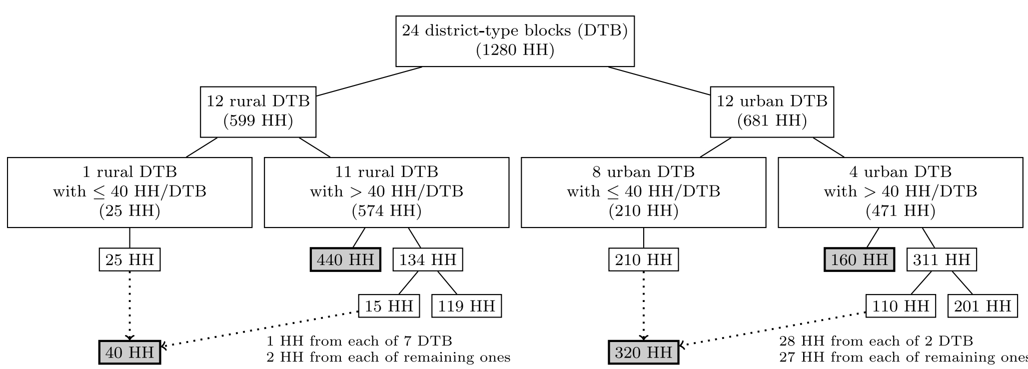

Fig. 4. Illustration of sampling procedure for broader sample. This procedure was used to select 480 urban households and 480 rural households from the sampling frame of 1280 households that provided phone numbers as part of the survey conducted by Sharma et al. (2018), who relied on a sample of 1917 households. Groups of sampled households are shown in the gray-shaded boxes. District-type blocks (DTB) are used to describe urban or rural households in a particular district. Because the sampling frame included households across 12 districts, households are sampled from 24 DTB (urban or rural households, dispersed across 12 districts).

unaffected. As the number of households remains relatively large and we survey them multiple times, our statistical estimates are reasonably precise. The primary challenge we face is that households may not answer the phone calls. Evaluating the implications of such missing data is an important part of our study design, and we pay careful attention to the confidence intervals of our estimates.

4. Results

The results reveal two main patterns. We note a substantial amount of nonlinearity in the relationship between reported and observed voltage reliability over a 24-hour period, and very little that conditions this relationship. Extending this to the broader sample, we note a substantial gap in reliability between urban and rural areas, which becomes even more clear when we take into account the nonlinearity of the relationship between reported and actual reliability.

4.1. Comparing self-reported reliability data to technical measures

We begin by analyzing the comparison between our crowdsourced observations in Sitapur and the Prayas monitor observations. The left-hand portion of Fig. 5 shows the fixed element of the hierarchical model, plotted against a scatterplot of the reported versus observed hours of normal voltage, with 95% confidence intervals, based on 100 simulations from the posterior distribution of the model, in blue. The 45 degree line shows where the observations would be if respondents reported voltage with perfect accuracy.

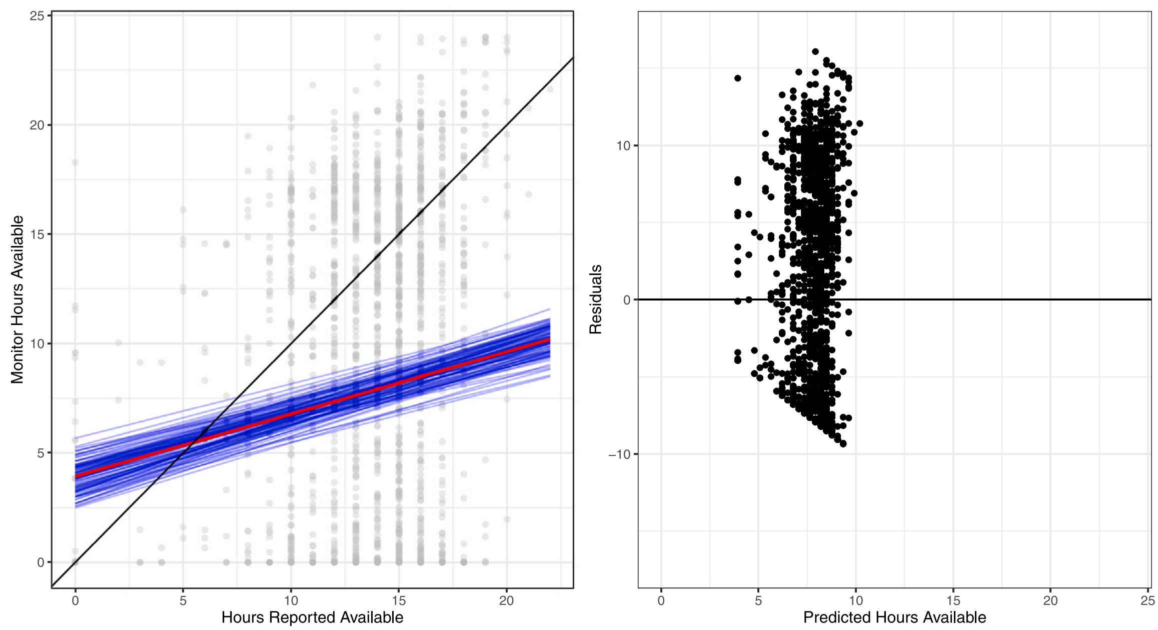

Fig. 5. Comparison of monitor reporting to crowd-sourced estimates. Left-hand plot shows the hours of normal voltage electricity available according to Prayas monitors, plotted against the hours reported by survey participants. The 45 degree line shows the expected relationship if the reports were equivalent. The red line shows the estimated regression relationship (see equation 1) and the blue lines show the ranges from 100 simulations from the posterior distribution of the model. The intercept shows that, on average, the monitors detect 5 more hours of normal voltage than what is reported, while the slope indicates that, for each additional hour of normal voltage detected by the monitors, the number of hours reported by our respondents increases by about 12 minutes. The right-hand plot shows the residuals from the model plotted against the predicted values. This chart demonstrates the issues of limited prediction range and large errors.

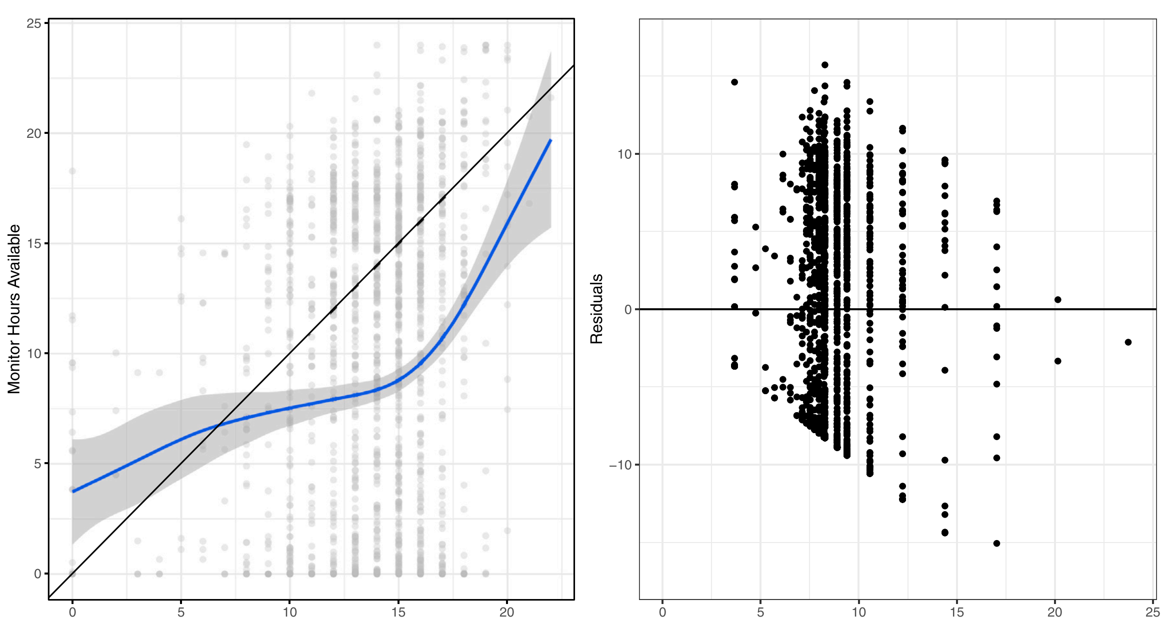

Fig. 6. Non-linear comparison of monitor reporting to crowd-sourced estimates. The left-hand plot shows the hours of normal voltage electricity available according to Prayas monitors, plotted against the hours reported by survey participants. The blue line shows a LOESS smoothed estimate for the relationship between reported and monitored hours. The results suggest that respondents were able to distinguish very low and very high levels of normal voltage relatively well, but had greater difficulty distinguishing the middle levels of availability, about 7 to 15 h of availability. The right-hand plot shows the residuals plotted against the predicted values. These results show that the prediction range is now much broader and the errors begin to decrease in predictions at the low and high end of the scale.

It is easy to see that individual recall is not very accurate. Specifically, after a certain point, respondents seem to substantially underestimate the amount of normal voltage that they receive. This can also be seen in the right-hand portion of the chart, which shows the residuals of the model plotted against the predicted values. The results show that forecasts were made over a relatively narrow range of values, with substantial error in both directions. In other words, the linear model is not very good at capturing differences between users with relatively high levels of electricity versus those with low levels because of the systematic underestimation in the middle-range of the distribution.

We check for a range of conditioning factors, looking at whether some respondents are better at estimating the number of hours without electricity or low voltage. The results in the SI indicate that this is not the case. We also check in the SI whether some of our respondents

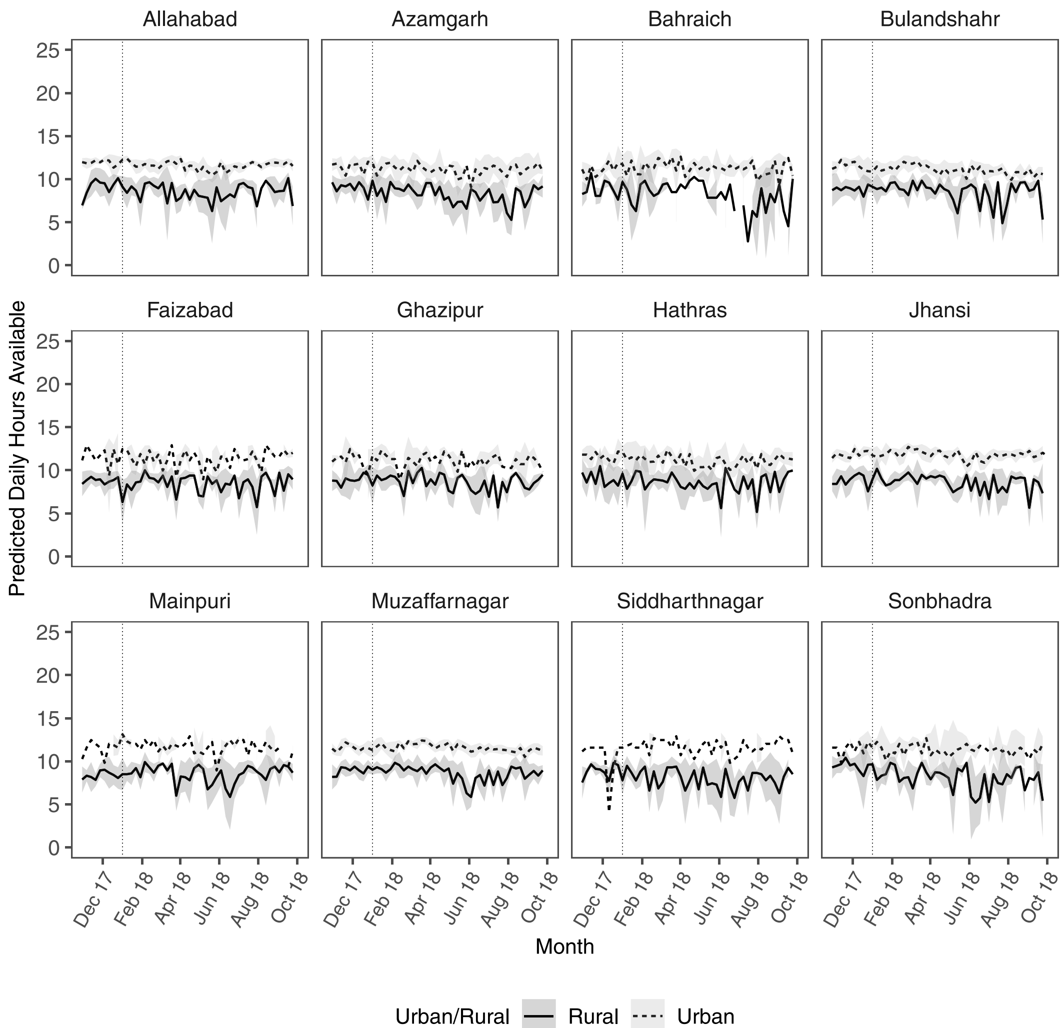

Fig. 7. Between-district comparison of linear predictions of monitored hours of normal voltage electricity over time. Using a linear estimate of the relationship between hours reported in the Prayas survey and those monitored by Prayas, households' hours of normal voltage electricity are predicted based on their self-reported hours in the CGEP survey. Predicted hours for each household are aggregated by households' districts and their classification as urban or rural. Aggregated averages are then plotted against each of the 12 months over which the study was conducted. The dotted vertical line at January 2018 demarcates predictions made using self-reported data in 2017 from those made using that in 2018.

might be particularly good or bad at reporting the reliability of their electricity based on demographic characteristics like education, ownership of alternative energy sources, income, and reported hours utilized. Again, we do not find that these characteristics interacted with reported hours available in a manner that improves the accuracy of reports.

Looking closer at Fig. 5, it is clear that the results are being driven by a mass of observations in the middle-range of observed voltage. Respondents in this area consistently under-report the amount of voltage that they are receiving. This seems to reveal a particular pattern of voltage observation that is quite understandable. For those whose electricity is irregular, there is a tendency to simply understand that their reliability is poor. Trying to distinguish levels of moderate reliability, it appears, is quite difficult for most respondents. Meanwhile, the model is missing those at the very high end, who seem to be closer to the 45 degree line.

To more directly model this, we turn to the LOESS model in Fig. 6. As anticipated, the left-hand panel shows that there is a substantial discontinuity toward the middle of the reported hours. The results demonstrate that those who report high levels of normal voltage, do seem to actually have high levels, while those in the middle range are likely to underestimate their access. The right-hand panel, again plots the residuals, and, while it still has difficulty in the middle-range of the distribution, the predicted values now cover nearly the full range of access and the errors tend to decrease as access becomes more reliable.

Overall, these results suggest that crowdsourcing reports about reliability does provide some signal about actual reliability, but the relationship is highly nonlinear — those who have moderate reliability tend to underestimate the amount of time in which they have normal voltage, which causes a problem for linear models. Armed with this information, we now move into a broader examination of electricity reliability in several areas of Uttar Pradesh.

4.2. Extending the model to a broader sample

Having fitted a prediction model to the sample of crowdsourced responses in Sitapur and the Prayas monitor observations, we then draw predictions of monitored hours using crowdsourced data collected from a broader and more representative sample. Self-reported average daily hours of electricity among rural households in our representative

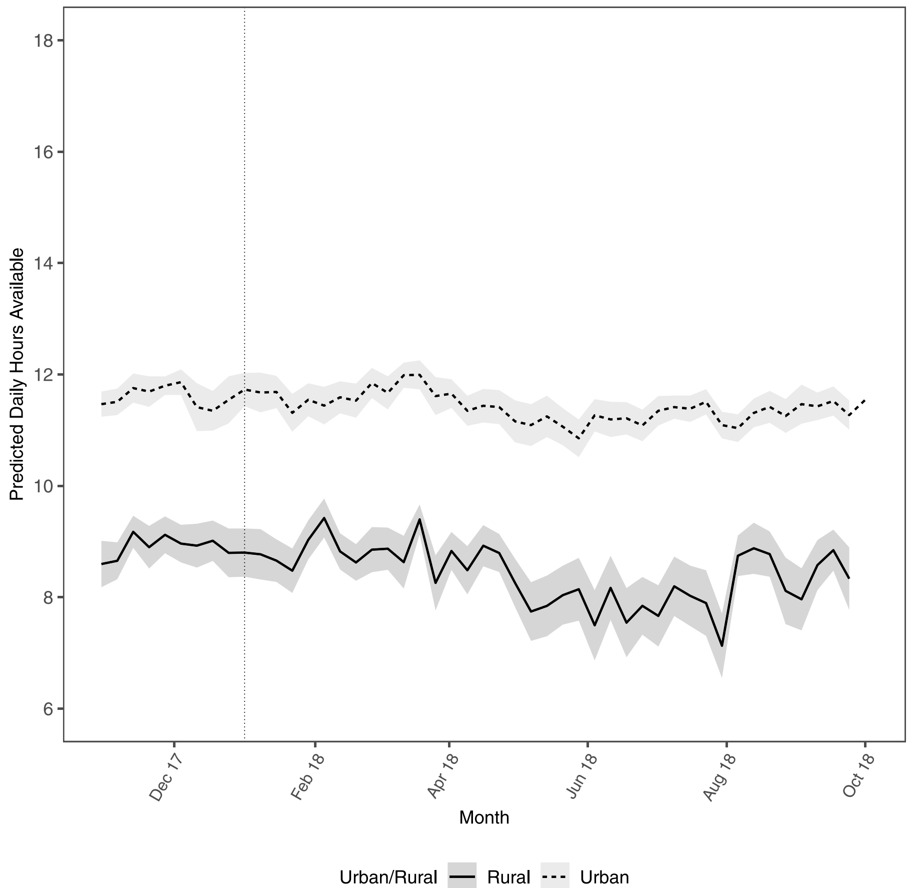

Fig. 8. Urban-rural comparison of linear predictions of monitored hours of normal voltage electricity over time. Using a linear estimate of the relationship between hours reported in the Prayas survey and those monitored by Prayas, households' hours of normal voltage electricity are predicted based on their self-reported hours in the CGEP survey. Predicted hours for each household are aggregated based on their classification as urban or rural. Aggregated averages are then plotted against each of the 12 months over which the study was conducted. The dotted vertical line at January 2018 demarcates predictions made using self-reported data in 2017 from those made using that in 2018.

sample were similar to those reported in other surveys conducted around the same time across the state (Jain et al., 2018).[11]

Fig. 7 compares predicted daily hours of electricity available across districts, segmented on the basis of rural and urban households. Across all districts, predicted daily hours among rural households lag behind their urban counterparts. This finding comports with existing research on the challenges of rural electrification (Joseph, 2010; Phadke et al., 2019; Aklin and Urpelainen, 2020). Fig. 8, which compares urban and rural predicted hours aggregated across all districts, further reinforces these observations. Fig. 7 also illustrates a precipitous decline, in rural districts' predicted hours between June and September. This corresponds to India's monsoon season, when high electricity demand and poorly planned load-shedding can increase the incidence of power outages (Alam, 2014; Sharma et al., 2018; Conevska and Urpelainen, 2020). Given the negligible decline in predicted hours among urban households, rural shortages may be attributable to poor electricity infrastructure and management and to comparably high agricultural demand for electricity, which is heavily subsidized by the federal government (Tongia, 2007; Badiani-Magnusson and Jessoe, 2019).

[11] The average of district-aggregated self-reported responses in our representative sample was 13.00 h. In a 2018 survey by Jain et al. (2018) of rural households across 18 districts in Uttar Pradesh, the average daily hours of electricity, aggregated by district was 12.75 h (Jain et al., 2018). A previous version of the survey conducted in 2015 finds a self-reported district-aggregated average hours of 8.80 (Aklin et al., 2016a). In comparison to these figures, the average of district-aggregated daily hours, using the linear and LOESS models, were 8.50 and 7.98 respectively.

Figs. 9 and 10 replicate Figs. 7 and 8, using predictions generated from the LOESS model. In Fig. 9, predicted daily hours exhibit greater temporal and geographic variability that those in Fig. 7. Additionally, the difference between urban and rural electricity access is greater in Fig. 10 than in Fig. 8. Both observations are attributable to non-linearities in the relationship between crowdsourced and monitored hours. As discussed in Section 4.1, subjects with moderate access to electricity systematically underestimate their daily hours compared to subjects with poor access. In capturing this pattern and correcting for the bias in moderate households' access, predictions from the non-linear model draw a starker contrast – over time, across districts, and across urban and rural areas — between moderate and poor access to electricity.

5. Conclusion and policy implications

Crowdsourcing holds promise as a low-cost way to collect data in challenging settings. We have found that very short telephone surveys can offer useful insights into the reliability of electricity service over time at a low cost. While the correlation between the reported and actual reliability of electricity is not perfect, it does function well at obtaining key aspects of energy reliability (i.e., whether access is reliable or not). Moreover, they seem to generalize relatively well over a variety of areas in Uttar Pradesh, demonstrating that there is a substantial difference in reliability between urban and rural areas, and showing that this reliability gap increases during the high demand monsoon season. These findings, in themselves, provide important input for policy-makers, as they reveal that broad gains in electricity access, in

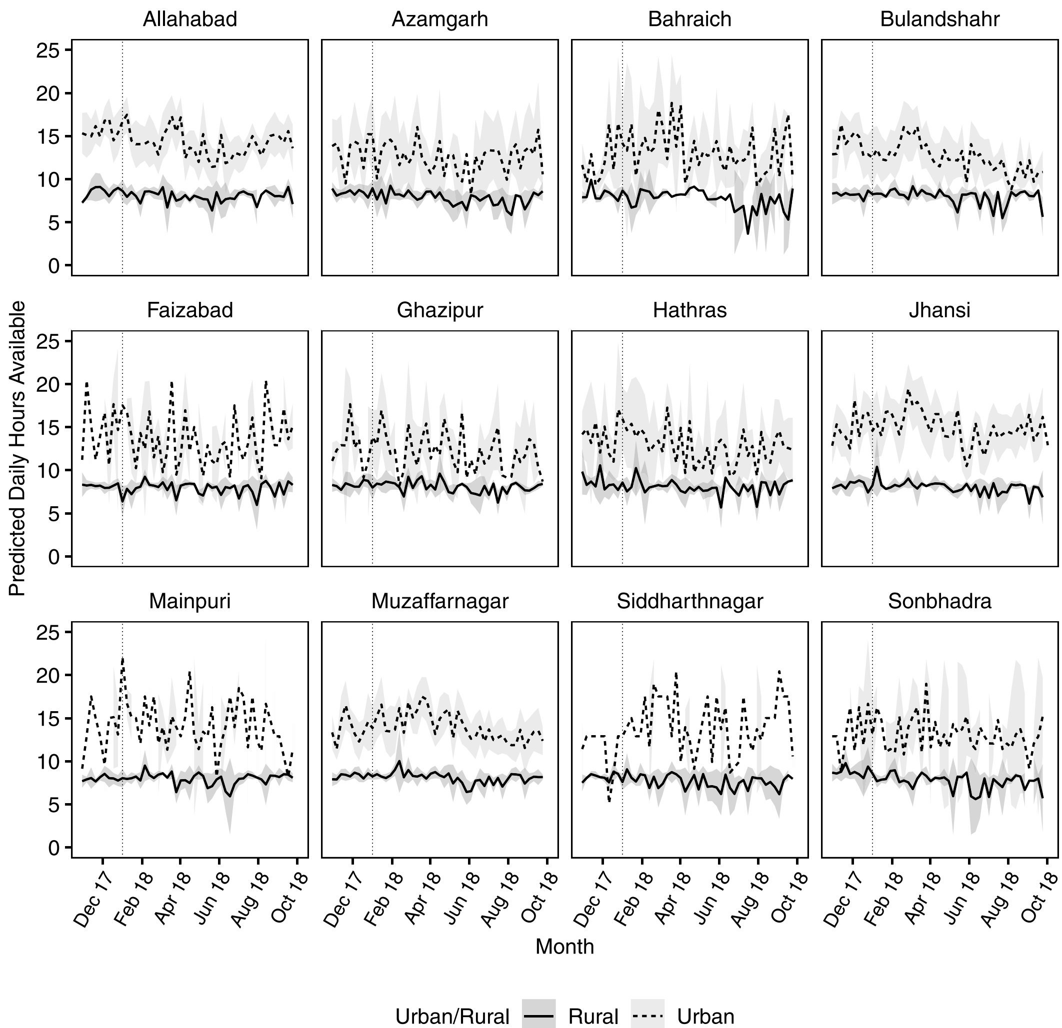

Fig. 9. Between-district comparison of LOESS predictions of monitored hours of normal voltage electricity over time. Using a LOESS smoothed estimate for the relationship between hours reported in the Prayas survey and those monitored by Prayas, households' hours of normal voltage electricity are predicted based on their self-reported hours in the CGEP survey. Predicted hours are aggregated on the basis of households' districts and their classification as urban or rural households. Aggregated averages are then plotted against each of the 12 months over which the study was conducted. The dotted vertical line at January 2018 demarcates predictions made using self-reported data in 2017 from those made using that in 2018.

terms of having a connection, still fall well short of the requirements of the MTF. While the average self-reported hours of electricity in our representative sample are similar to those reported in other recent surveys of Uttar Pradesh (Jain et al., 2018), the district-aggregated average hours predicted by the linear and LOESS models are substantially lower.[12] Increases in self-reported hours between 2015 and 2018 may overstate the impact of government electrification programs,[13] with upward bias increasing with the hours of electricity available to households.

Pending the creation of nationwide monitoring systems, short telephone surveys can help government, private sector, and civil society agents assess the reliability of electricity service. From the results above, we argue that this will be an improvement over baseline methods that rely on individuals self-reporting reliability problems, as these have well-established biases toward wealthier and better-connected communities, who are more likely to notice and report issues (O'Brien et al., 2015), or on estimates from cross-sectional surveys. As long as these agents have access to phone numbers, they can either make phone calls, use even cheaper interactive voice response surveys, or collect responses through text messages. This is a critical tool for advocates and policy-makers as we move the discussion from simple connections to reliable electricity access.

While further research will be necessary to determine the spatial and temporal limitations of this approach, other research suggests that this approach can be maintained over the long-term and be generalized to the rest of India. In terms of long-term prospects, citizen science programs have demonstrated that lose networks of individuals can produce large-scale crowdsourcing data that lasts for decades (e.g. Butcher et al., 1990), but the longevity of these programs depends on continued

[12] The average of district-aggregated self-reported responses in our representative sample was 13.00 h. In a 2018 survey by Jain et al. (2018) of rural households across 18 districts in Uttar Pradesh, the average daily hours of electricity, aggregated by district was 12.75 h (Jain et al., 2018). A previous version of the survey conducted in 2015 finds a self-reported district-aggregated average hours of 8.80 (Aklin et al., 2016a). In comparison to these figures, the average of district-aggregated daily hours, using the linear and LOESS models, were 8.50 and 7.98 respectively.

[13] In April, 2018, the Government of India announced that its *Pradhan Mantri Gramodaya Yojana* initiative had achieved 100% village electrification. Numerous surveys and experiments seeking to measure the impacts of the program on electricity access and reliability rely on self-reported data (Kennedy et al., 2019; Blankenship et al., 2019).

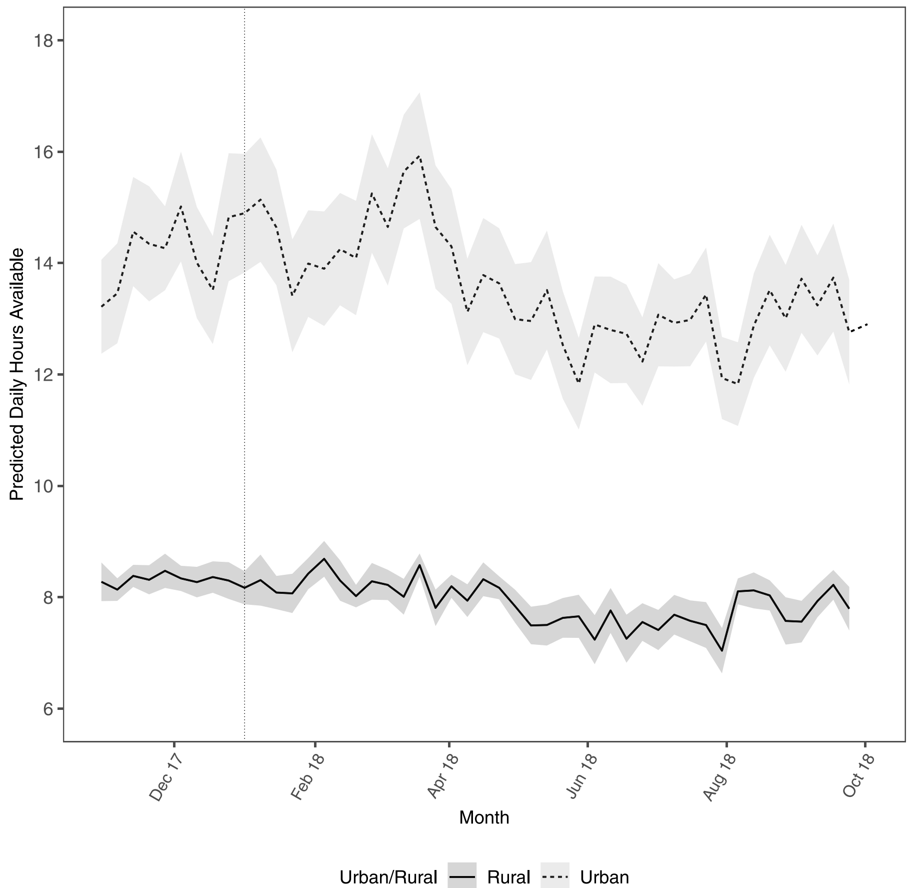

Fig. 10. Urban-rural comparison of LOESS predictions of monitored hours of normal voltage electricity over time. Using a LOESS smoothed estimate for the relationship between hours reported in the Prayas survey and those monitored by Prayas, households' hours of normal voltage electricity are predicted based on their self-reported hours in the CGEP survey. Predicted hours for each household are aggregated based on their classification as urban or rural. Aggregated averages are then plotted against each of the 12 months over which the study was conducted. The dotted vertical line at January 2018 demarcates predictions made using self-reported data in 2017 from those made using that in 2018.

publicity and the development of strong social networks (Lowry et al., 2019). Alternatively, continued monitoring by government agencies or NGOs can provide incentives to participants, as we did, at a relatively low cost, but this relies on continuing interest of the organization to produce this data.

In terms of generalizability, access to cell phones is widespread in India, with an estimated 60% of individuals owning a mobile phone and 30% of the population owning smartphones in 2020.[14] Our results suggest that the characteristics of the household make no significant difference in the accuracy of their reporting. Perhaps the biggest challenge to expansion is developing a representative sample for larger geographies in India, but there are good models for this from previous surveys of energy access (e.g. Aklin et al., 2015).

The strength of crowdsourced data lies not only in its low cost, but also in its ability to record fluctuation over time. While crowdsourcing cannot replace technical measurements as a planning tool, it can highlight clear reductions or improvements in the reliability of service. For example, crowdsourcing can alert the government, media, and public to sudden changes in the reliability of electricity service and motivate further investigation with technical monitoring. Crowdsourcing also need not depend on government, so it can be used to hold electric utilities accountable for their performance and verify claims about said performance.

To be sure, our approach is but a small step toward fully exploiting crowdsourcing in this area. Future research should look at the process of automating these types of push surveys, encouraging online participation, and using other methods to collect this crowdsourced data. We will likely need a myriad of methods for tracking and understanding energy behaviors, of which this is only a start.

[14] Projected numbers from www.statista.com.

CRediT authorship contribution statement

Ryan Kennedy: Data curation, Methodology, Investigation, Visualization, Writing - original draft, Writing - review & editing. **Aseem Mahajan:** Methodology, Investigation, Visualization, Writing - original draft, Writing - review & editing. **Johannes Urpelainen:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Supervision, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aklin, Michaël, Bayer, Patrick, Harish, S.P., Urpelainen, Johannes, 2015. The political economy of energy access: Survey evidence from India on state intervention and public opinion. Energy Res. Soc. Sci. 10, 250–258.

Aklin, Michaël, Cheng, Chao-yo, Ganesan, Karthik, Jain, Abhishek, Urpelainen, Johannes, et al., 2016a. Access to clean cooking energy and electricity: Survey of states in India (ACCESS). In: Harvard Dataverse, Version 1.

Aklin, Michaël, Cheng, Chao-yo, Urpelainen, Johannes, Ganesan, Karthik, Jain, Abhishek, 2016b. Factors affecting household satisfaction with electricity supply in rural India. Nat. Energy 1 (16170).

Aklin, Michaël, Urpelainen, Johannes, 2020. Trials and tribulations: Lost energy access gains in rural India. Energy Sustain. Dev. 55, 190–200.

Alam, Muneeza Mehmood, 2014. Coping with Blackouts: Power Outages and Firm Choices. Yale University.

Allcott, Hunt, Collard-Wexler, Allan, O'Connell, Stephen D., 2016. How do electricity shortages affect industry? Evidence from India. Amer. Econ. Rev. 106 (3), 587–624.

Badiani-Magnusson, Reena, Jessoe, Katrina, 2019. Electricity prices, groundwater and agriculture: The environmental and agricultural impacts of electricity subsidies in India. In: Agricultural Productivity and Producer Behavior. University of Chicago Press.

Baltrusaitis, Kristin, Santillana, Mauricio, Crawley, Adam W., Chunara, Rumi, Smolinski, Mark, Brownstein, John S., 2017. Determinants of participants' follow-up and characterization of representativeness in flu near you, a participatory disease surveillance system. JMIR Publ. Health Surveill. 3 (2), e18.

Baskaran, Thushyanthan, Min, Brian, Uppal, Yogesh, 2015. Election cycles and electricity provision: Evidence from a quasi-experiment with Indian special elections. J. Publ. Econ. 126, 64–73.

Bhatia, Mikul, Angelou, Nicolina, 2014. Capturing the multi-dimensionality of energy access. LiveWire (World Bank) URL: http://bit.ly/30imd0m.

Blankenship, Brian, Wong, Jason Chun Yu, Urpelainen, Johannes, 2019. Explaining willingness to pay for pricing reforms that improve electricity service in India. Energy Policy 128, 459–469.

Blodgett, Courtney, Dauenhauer, Peter, Louie, Henry, Kickham, Lauren, 2017. Accuracy of energy-use surveys in predicting rural mini-grid user consumption. Energy Sustain. Dev. 41, 88–105.

Butcher, Gregory S., Fuller, Mark R., McAllister, Lynne S., Geissler, Paul H., 1990. An evaluation of the christmas bird count for monitoring population trends of selected species. Wildlife Soc. Bull. (1973-2006) 18 (2), 129–134.

Chakravorty, Ujjayant, Pelli, Martino, Marchand, Beyza Ural, 2014. Does the quality of electricity matter? Evidence from rural India. J. Econ. Behav. Organ. 107, 228–247.

Cleveland, William S., Devlin, Susan J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. J. Am. Statist. Assoc. 83 (403), 596–610.

Conevska, Aleksandra, Urpelainen, Johannes, 2020. Weathering electricity demand? Seasonal variation in electricity consumption among off-grid households in rural India. Energy Res. Soc. Sci. 65, 101444.

Gelman, Andrew, Hill, Jennifer, 2006. Data Analysis Using Regression and Multilevel/hierarchical Models. Cambridge University Press.

Gibson, John, Olivia, Susan, 2010. The effect of infrastructure access and quality on non-farm enterprises in rural Indonesia. World Dev. 38 (5), 717–726.

Heinzelman, Jessica, Brown, Rachel, Meier, Patrick, 2011. Mobile technology, crowdsourcing and peace mapping: New theory and applications for conflict management. In: Mobile Technologies for Conflict Management. Springer, pp. 39–53.

IEA, 2018. World energy outlook. International Energy Agency, Paris.

Jain, Abhishek, Tripathi, Saurabh, Mani, Sunil, Patnaik, Sasmita, Shahidi, Tauseef, Ganesan, Karthik, 2018. Access to clean cooking energy and electricity: Survey of states 2018. https://www.ceew.in/publications/access-clean-cooking-energy-and-electricity, (Accessed 13 August 2019).

Joseph, Kelli L., 2010. The politics of power: electricity reform in India. Energy Policy 38 (1), 503–511.

Kemmler, Andreas, 2007. Factors influencing household access to electricity in India. Energy Sustain. Dev. 11 (4), 13–20.

Kennedy, Ryan, Mahajan, Aseem, Urpelainen, Johannes, 2019. Quality of service predicts willingness to pay for household electricity connections in rural India. Energy policy 129, 319–326.

Lowry, Christopher S., Fienen, Michael N., Hall, Damon M., Stepenuck, Kristine F., 2019. Growing pains of crowdsourced stream stage monitoring using mobile phones: The development of crowdhydrology. Front. Earth Sci. 7, 128.

Min, Brian, Golden, Miriam, 2014. Electoral cycles in electricity losses in India. Energy Policy 65, 619–625.

O'Brien, Daniel Tumminelli, Sampson, Robert J., Winship, Christopher, 2015. Ecometrics in the age of big data: Measuring and assessing broken windows using large-scale administrative records. Sociol. Methodol. 45 (1), 101–147.

Phadke, Amol, Park, Won Young, Abhyankar, Nikit, 2019. Providing reliable and financially sustainable electricity access in India using super-efficient appliances. Energy Policy 132, 1163–1175.

Post, Alison E., Agnihotri, Anustubh, Hyun, Christopher, 2018. Using crowd-sourced data to study public services: Lessons from urban India. Stud. Comp. Int. Dev. 53 (3), 324–342.

Sharma, Shruti, Moerenhout, Tom, Garg, Vibhuti, 2018. Electricity Sector Reform in Uttar Pradesh: Analysis fo tariff adjustments and the Ujwal Discom Assurance Yojana Plan (UDAY). Global Subsidies Initiative Report, International Institute for Sustainable Development.

Smolinski, Mark S., Crawley, Adam W., Baltrusaitis, Kristin, Chunara, Rumi, Olsen, Jennifer M., Wójcik, Oktawia, Santillana, Mauricio, Nguyen, Andre, Brownstein, John S., 2015. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. Am. J. Publ. Health 105 (10), 2124–2130.

Tongia, Rahul, 2007. The political economy of power sector reform in india. In: Victor, D., Hellter, T. (Eds.), The Political Economy and Institutions of Power Sector Reform in the Developing World. Cambridge University Press.

Tschiatschek, Sebastian, Singla, Adish, Gomez Rodriguez, Manuel, Merchant, Arpit, Krause, Andreas, 2018. Fake news detection in social networks via crowd signals. In: Companion Proceedings of the The Web Conference 2018. International World Wide Web Conferences Steering Committee, pp. 517–524.

Van der Windt, Peter, Humphreys, Macartan, 2016. Crowdseeding in eastern congo: Using cell phones to collect conflict events data in real time. J. Confl. Resolut. 60 (4), 748–781.

Warkentin, Darcy, Woodworth, Michael, Hancock, Jeffrey T., Cormier, Nicole, 2010. Warrants and deception in computer mediated communication. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work. Technical Report, ACM, pp. 9–12.