

Assessing Reliability of Electricity Grid Services from Space: The Case of Uttar Pradesh

—Supplementary Information—

Eugenie Dugoua*
London School of Economics Ryan Kennedy
University of Houston Myriam Shiran
University of Houston
Johannes Urpelainen
Johns Hopkins SAIS

March 12, 2022

Contents

SI1 Data Description	SI-2
SI1.1 Plotting the Prayas Data	SI-2
SI1.2 Summary Statistics	SI-5
SI2 Main Analysis	SI-7
SI2.1 Regression Tree	SI-7
SI2.2 Median Bootstrap Predictions	SI-9
SI3 Alternative measures of reliability	SI-13
SI3.1 Prediction Errors	SI-16

*Corresponding author. Email: E.Dugoua@lse.ac.uk

SI1 Data Description

SI1.1 Plotting the Prayas Data

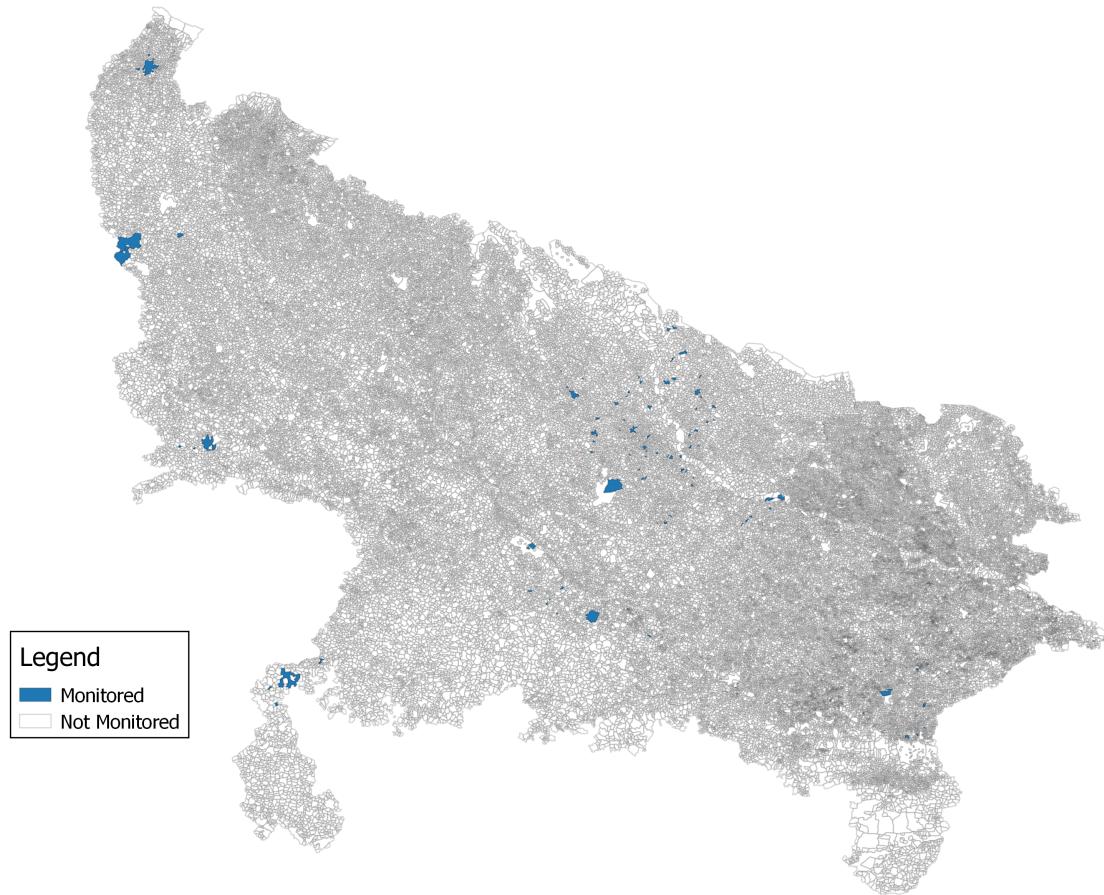


Figure SI1: Areas With Prayas ESM monitors. Map shows the villages with EMS monitors set up by Prayas during the time period of the study. Regions in blue have been monitored for at least part of the study period, while those not shaded were not.

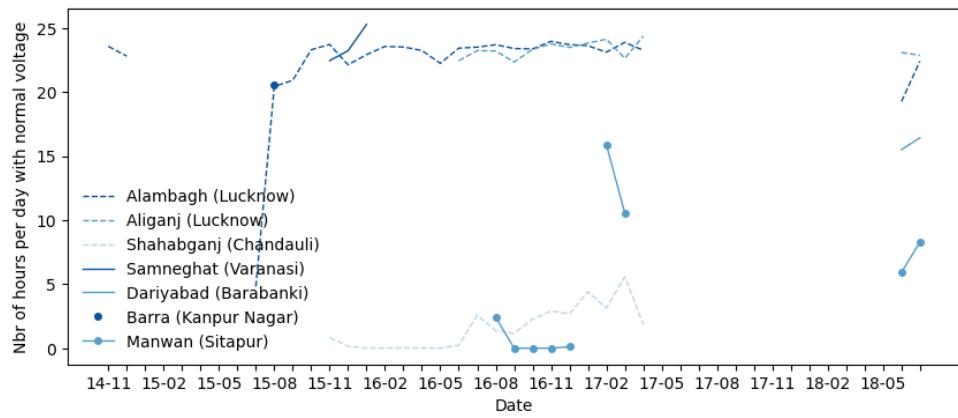


Figure SI2: Hours per Day with Electricity: Monthly Average Over Time For Selected Locations Chart shows number of hours per day with normal voltage as recorded by Prayas ESM monitors for seven sample areas during the study period. The results show both substantial variety in the hours of normal voltage, as well as the issue discussed in the main manuscript with missing readings due to monitor failure. When monitors fail, these are treated as missing data (see main manuscript).

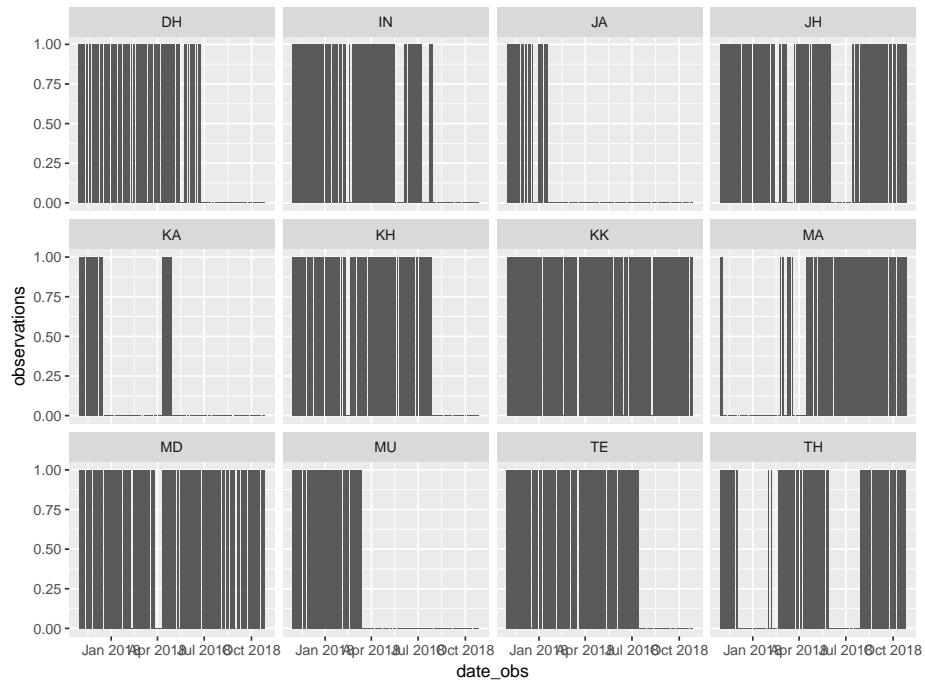


Figure SI3: Monitor outages for 12 Prayas locations. This figure shows the outages across 12 locations during the study period. Blank areas indicate no readings for that period. Such outages were significant and indicate the need for even better ground truth data for future studies.

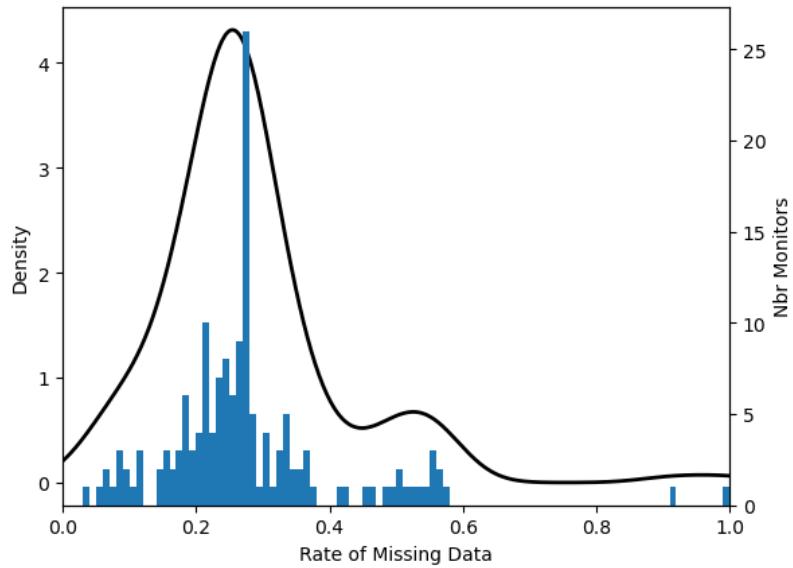


Figure SI4: Rate of Missing Data For Each Monitor. To construct this graph, we first identify the first and last day when a monitor recording is available for each monitor. We then calculate the number of hours spanning this first and last date. This corresponds to the number of observations that we could expect to have data for. Next, we look at in practice how many hours we have data for each monitor. We then take the ratio between those two things.

SI1.2 Summary Statistics

	count	mean	sd	min	max
Area	96	179.3	174.3	0.9	986.5
Population Density (x 1000)	96	2.4	5.0	0.0	24.2
Number of households (x 1000)	96	18.7	72.2	0.0	538.1
Population (x 1000)	96	99.4	378.2	0.2	2817.1
Schedule caste population (%)	96	20.4	15.1	0.0	79.5
Schedule tribe population (%)	96	0.1	0.4	0.0	2.8
Literate population (%)	96	56.2	13.2	14.6	81.0
If rural (1), or urban (0)	96	0.7	0.5	0.0	1.0
Sum of nightlight	96	896.3	3412.7	1.4	21640.9
Sum of nightlight per area	96	19.7	67.2	0.0	599.5
Sum of nightlight per population	96	0.0	0.0	0.0	0.2
Sum of nightlight per household	96	0.1	0.1	0.0	1.0
Number of hours with no voltage (24h day)	96	9.0	4.3	0.9	19.3
Number of hours with no voltage during the day (6AM-8PM)	96	6.1	2.8	0.6	12.5
Number of hours with no voltage at night (8PM-6AM)	95	2.8	1.7	0.2	8.0
Number of hours with low voltage (24h day)	96	3.6	3.5	0.0	12.0
Number of hours with low voltage during the day (6AM-8PM)	96	1.7	1.6	0.0	5.5
Number of hours with low voltage at night (8PM-6AM)	95	2.0	2.1	0.0	7.7
Number of hours with normal voltage (24h day)	96	11.2	6.5	0.0	22.9
Number of hours with normal voltage during the day (6AM-8PM)	96	6.1	3.7	0.0	13.3
Number of hours with normal voltage at night (8PM-6AM)	95	5.2	3.1	0.0	9.7
Number of hours with high voltage (24h day)	96	0.2	0.6	0.0	5.5
Number of hours with high voltage during the day (6AM-8PM)	96	0.1	0.4	0.0	2.8
Number of hours with high voltage at night (8PM-6AM)	95	0.1	0.3	0.0	2.7
Monthly voltage (24h day)	96	140.3	47.3	32.5	246.6
Monthly voltage at night (8PM-6AM)	96	158.2	47.5	49.6	250.5

Table SI1: Summary statistics at the village level. The table provides key Census data, as well as Prayas and nightlight data at the village level for all areas of the study across all time periods.

	count	mean	sd	min	max
Sum of nightlight	600	2424.5	6361.1	0.0	39802.1
Sum of nightlight per area	600	38.9	121.3	0.0	1105.4
Sum of nightlight per population	600	0.0	0.0	0.0	0.4
Sum of nightlight per household	600	0.1	0.2	0.0	1.9
Number of hours with no voltage (24h day)	597	8.2	5.2	0.0	24.0
Number of hours with no voltage at night (8PM-6AM)	599	2.6	2.0	0.0	10.0
Number of hours with low voltage (24h day)	597	3.2	4.0	0.0	21.4
Number of hours with low voltage at night (8PM-6AM)	599	1.7	2.2	0.0	9.3
Number of hours with normal voltage (24h day)	597	12.5	7.6	0.0	24.0
Number of hours with normal voltage at night (8PM-6AM)	599	5.7	3.4	0.0	10.0
Number of hours with high voltage (24h day)	597	0.1	0.5	0.0	5.8
Number of hours with high voltage at night (8PM-6AM)	599	0.0	0.2	0.0	2.9
Monthly voltage (24h day)	600	149.3	56.4	0.0	252.6
Monthly voltage at night (8PM-6AM)	600	166.3	55.6	0.0	260.0

Table SI2: Summary statistics at the month-village level. The table provides key summary statistics at the month-village level (the level of analysis for the paper) for all areas included in the study during the study time period.

SI2 Main Analysis

SI2.1 Regression Tree

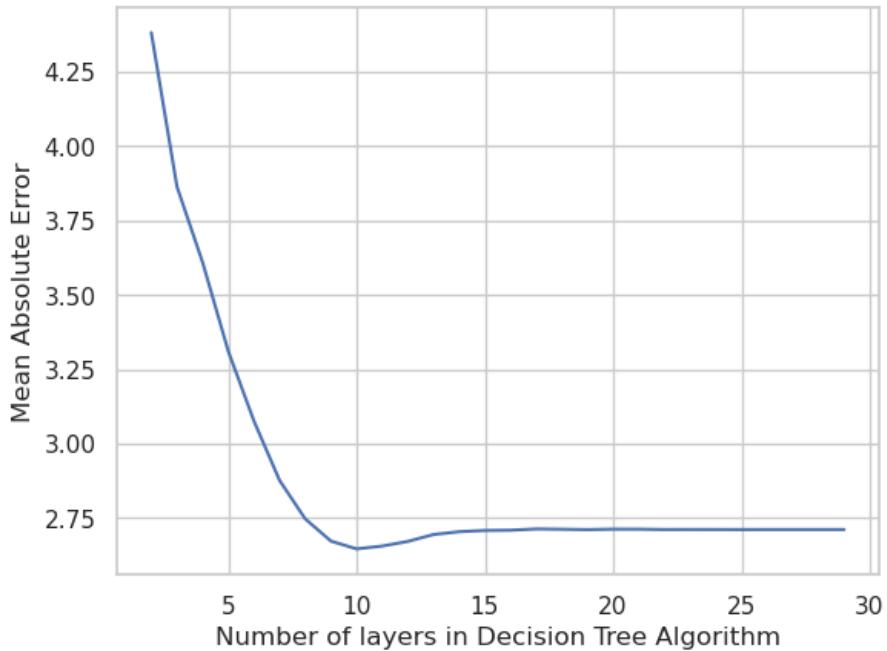


Figure SI5: **Assessing the performance of layers in cross-validation.** The plot shows the average mean absolute error (MAE) in a five-fold cross-validation exercise repeated 1000 times. The results suggest that the minimum MAE was achieved at a depth of 10 layers.

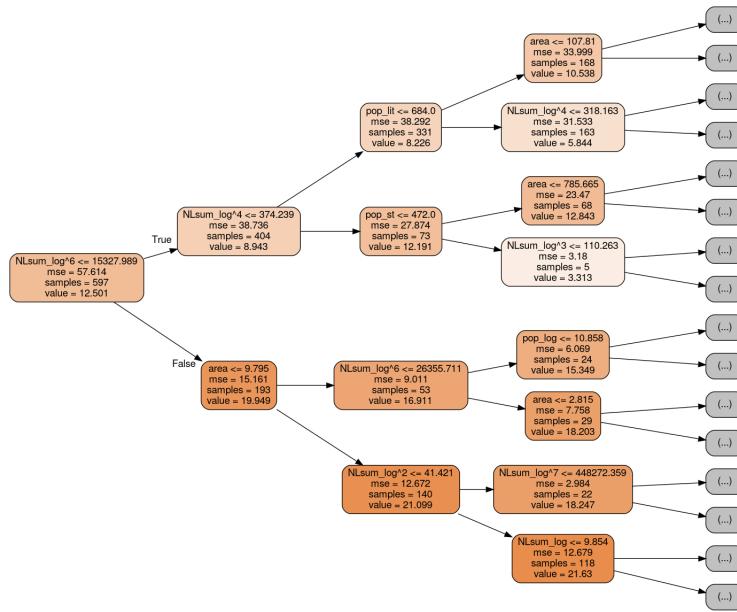


Figure SI6: First three layers of the ten-layer tree. Figure shows the first three layers of the full tree to illustrate how the decision tree functions. It also illustrates the importance of the nightlights data, in addition to the village Census data, as the nightlight data enters very early in the regression tree.

SI2.2 Median Bootstrap Predictions

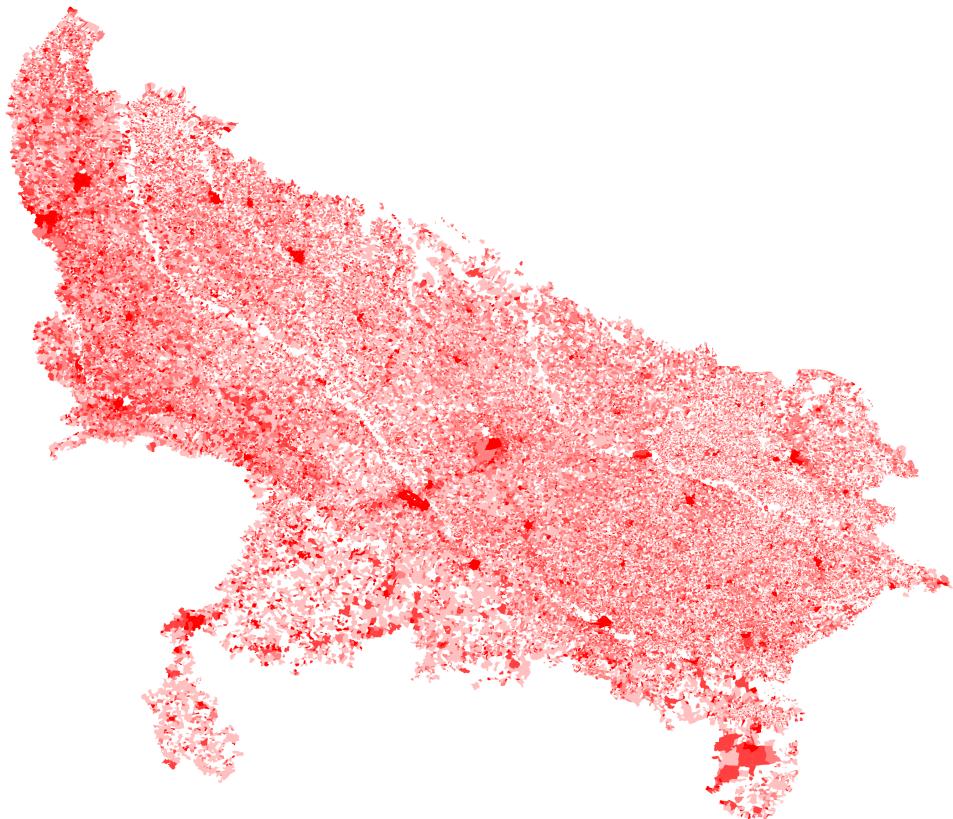


Figure SI7: Mapping the median bootstrap prediction for January 2018 Map illustrates the predicted number of hours with normal voltage across all survey areas. Darker red indicates a higher number of hours, while lighter areas indicate lower hours.

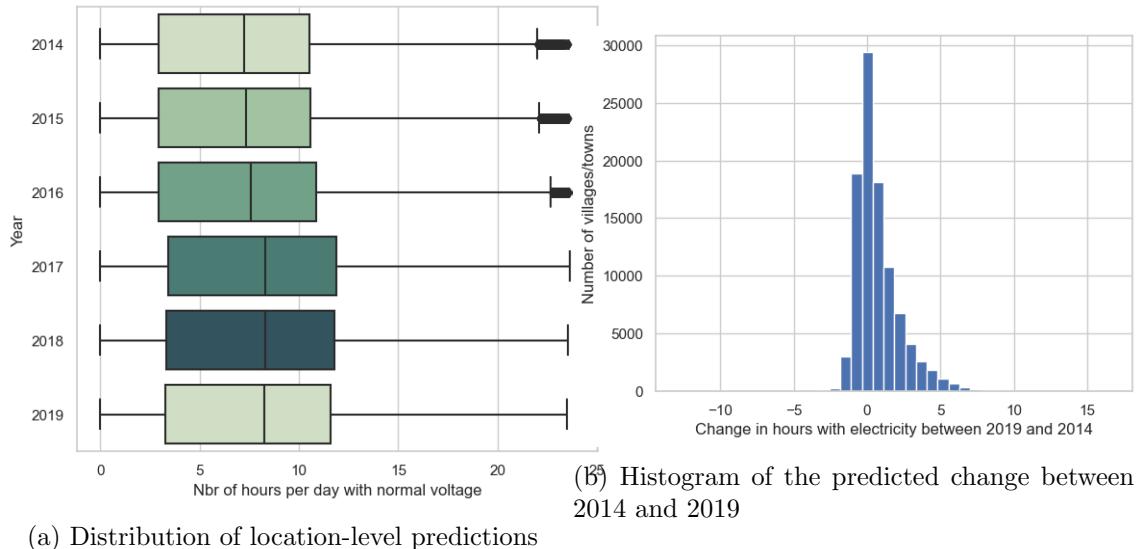


Figure SI8: Median bootstrap predictions for each location. Subfigure (a) shows boxplots of the distribution of predictions for each year in each location. Subfigure (b) shows the predicted change in the median level of hours per day with normal voltage in each location from 2014-2019 across bootstrap intervals to show illustrate the level of uncertainty.

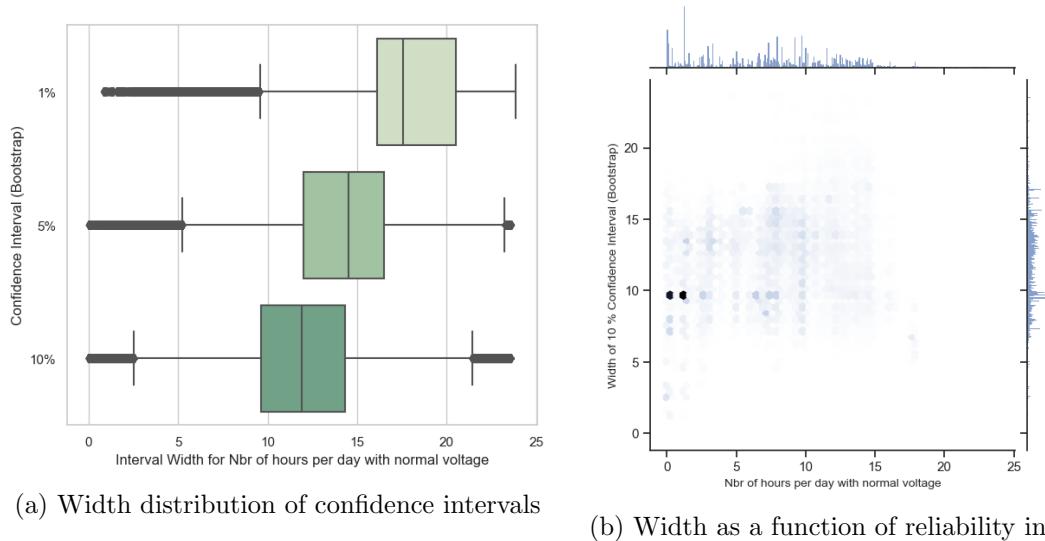


Figure SI9: Confidence intervals of median predictions for each location. Subfigure (a) shows the width of confidence intervals under 1%, 5% and 10% confidence, with confidence intervals calculated from bootstrap predictions. The results show significant uncertainty for individual locations, feeding the argument in the main paper that the method we are illustrating is not highly accurate for individual locations. Subfigure (b) shows the width of errors as a function of overall reliability. We note no particularly clear pattern in this data. The intensity of color in each hexagonal bin represents the number of village-month observations that fall within the range of the shape.

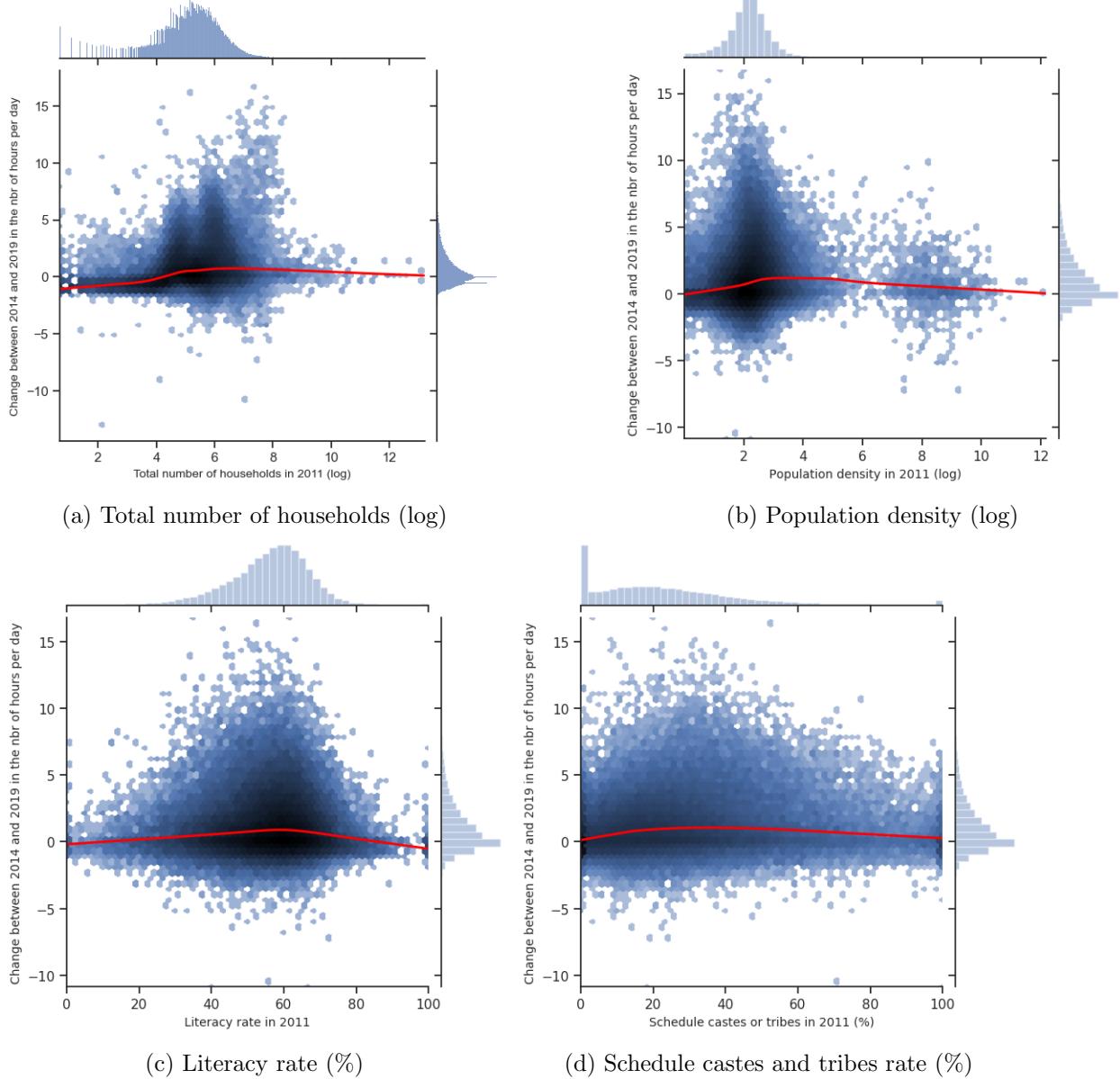


Figure SI10: Median predicted change between 2014 and 2019 as a function of covariates for each location. Subfigures (a) shows the relationship between the log total number of households in the village from the 2011 Census and the expected change in normal hours per day from 2014 to 2019. Subfigure (b) does the same for log population density. Subfigure (c) does this for village literacy rate. And subfigure (d) does this for percent scheduled caste or tribe. The results show no clear relationship between village characteristics and the change in the number of hours per day with normal electricity. The intensity of color in each hexagonal bin represents the number of village-month observations that fall within the range of the shape.

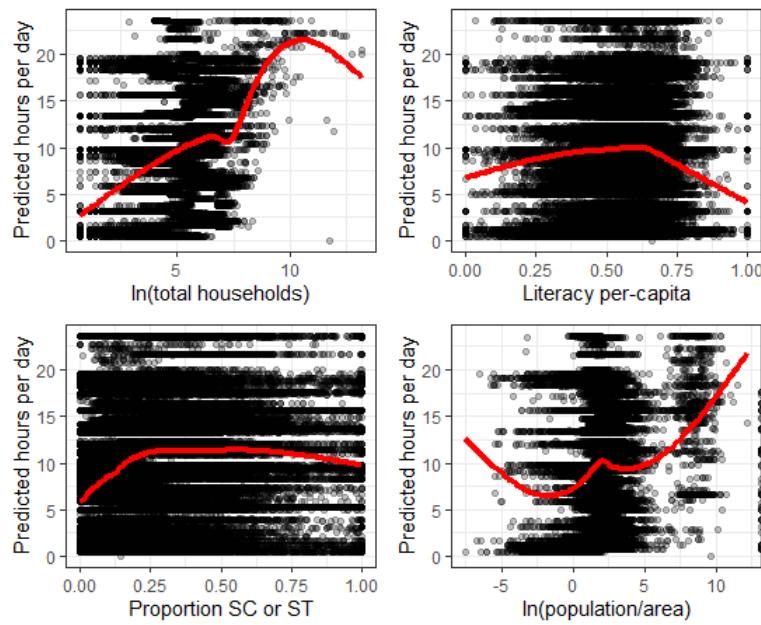
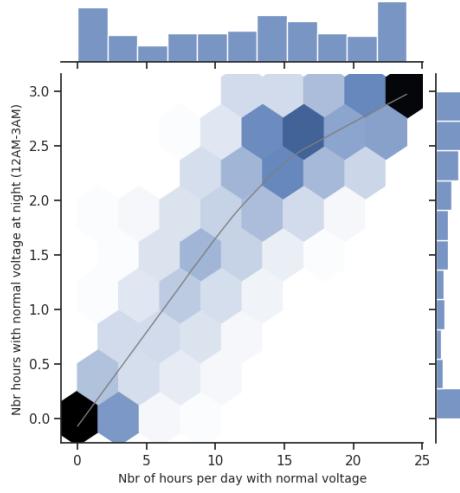
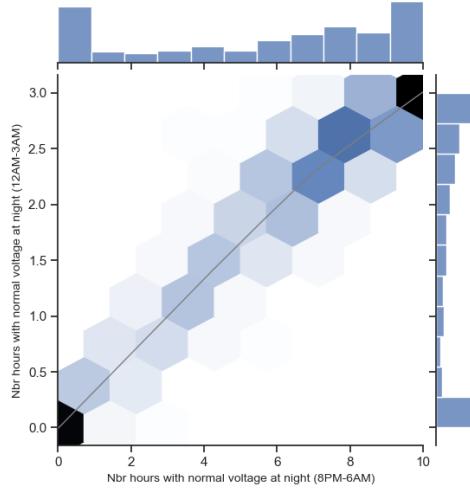


Figure SI11: Median predicted reliability as a function of covariates for each location.
 Figures show the relationship between the predicted hour per day and village characteristics as recorded in the 2011 Census. In particular, log of total households and log of total population density appear to have a substantial relationship with the reliability of electricity.

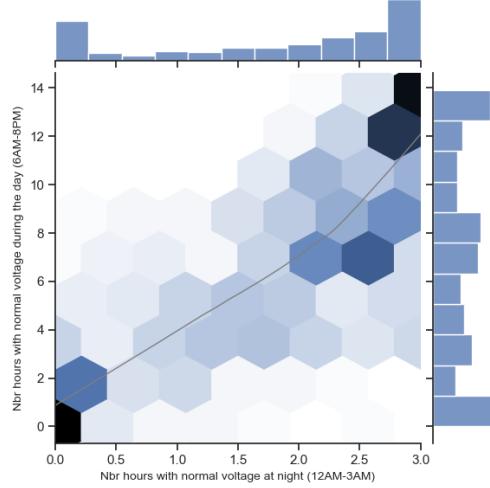
SI3 Alternative measures of reliability



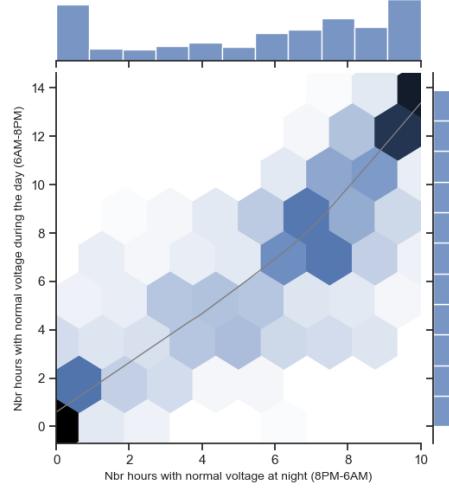
(a) Middle of the Night vs Whole Day



(b) Middle of the Night vs Nighttime



(c) Middle of the Night vs Daytime



(d) Nighttime vs Daytime

Figure SI12: Measures of Reliability for Different Time Windows. Figures show a strong relationship between nighttime hours of “normal” voltage and daytime hours, as well as between different periods during the night. Subfigure (a) shows the relationship between measures from 8PM to 6AM and measures taken from 6AM to 8PM. Subfigure (b) shows the relationship between measures taken from 8 PM to 6 AM and those taken from 12 AM to 3 AM. Subfigure (c) shows the relationship between measures taken from 12 AM to 3 PM and those from 6 AM to 8 PM. In all of these cases, there is a nearly linear and positive relationship, with greater error observed in Subfigure (c) than in Subfigure (a), and greater error in Subfigure (a) than Subfigure (b). The intensity of color in each hexagonal bin represents the number of village-month observations that fall within the range of the shape.

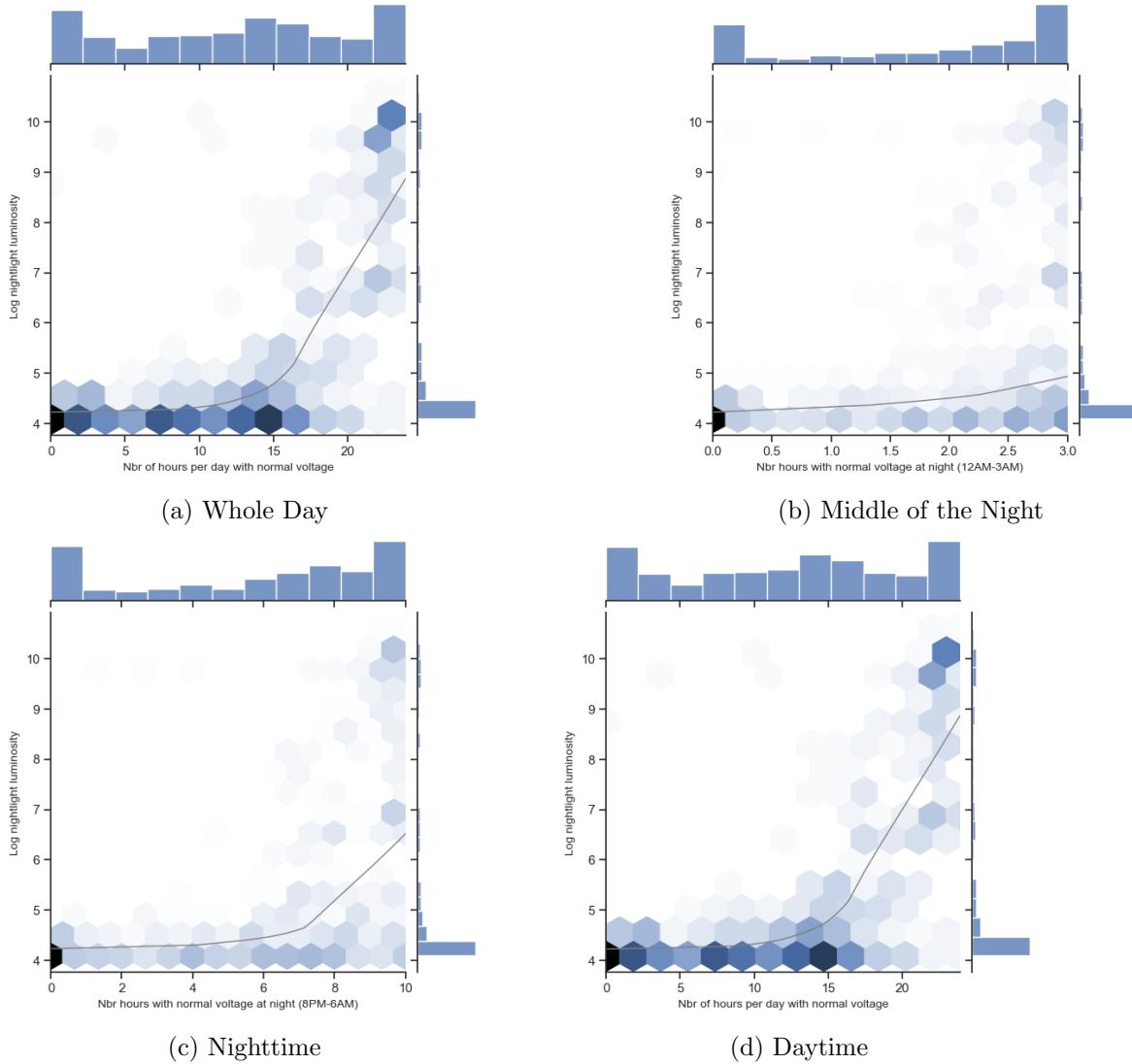
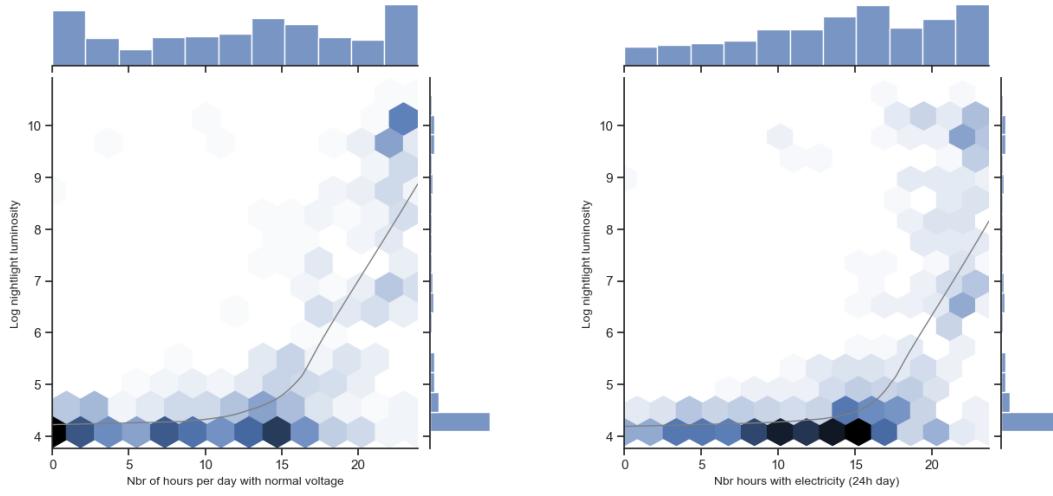


Figure SI13: Correlation between Nightlights and Reliability Measures

The intensity of color in each hexagonal bin represents the number of village-month observations that fall within the range of the shape.



(a) Nightlights vs Nbr Hours with Voltage between 205-270V throughout the day
 (b) Nightlights vs Nbr Hours with Voltage higher than 175V throughout the day

Figure SI14: Two measures of reliability based on different voltage thresholds, throughout the day. Figures compare the relationship between the number of hours with normal voltage against the log of nighttime luminosity at two different voltage thresholds for defining “normal” as measured over the full 24 hour period of monitoring. Figures (a) uses a threshold of 205-270V, while Figure (b) uses a threshold of 175V. Little change is noted based on the differences in these thresholds. The intensity of color in each hexagonal bin represents the number of village-month observations that fall within the range of the shape.

SI3.1 Prediction Errors

	(1) Normal 24h	(2) Normal 3AM	(3) Normal Night	(4) Normal Day
Area (X1000 Ha)	0.60 (1.20)	0.25 (0.25)	0.34 (0.77)	0.10 (0.66)
Number of households (log)	-0.36 (0.58)	-0.21 (0.13)	-0.47 (0.42)	0.01 (0.43)
Population (log)	0.42 (0.59)	0.21 (0.14)	0.41 (0.42)	0.02 (0.44)
Schedule caste population (x1000)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Schedule tribe population (x1000)	-0.88*** (0.31)	-0.08 (0.05)	-0.26 (0.17)	-0.51*** (0.18)
Literate population (million)	2.90*** (0.69)	0.24*** (0.09)	1.16*** (0.29)	1.47*** (0.41)
If rural (1), or urban (0)	0.02 (0.74)	-0.16 (0.13)	-0.46 (0.39)	0.06 (0.45)
Nbr of Monitors	-0.07 (0.04)	-0.03*** (0.01)	-0.10*** (0.03)	-0.03 (0.03)
Rate Missing Data	2.32 (2.46)	-0.53 (0.33)	-1.19 (0.93)	1.29 (0.99)
Number of Voltage Observations (X1000)	-0.05 (0.07)	0.03*** (0.01)	0.08** (0.03)	-0.04 (0.03)
Sum of nightlight (log)	0.07 (0.19)	0.02 (0.03)	0.06 (0.11)	0.03 (0.12)
Village Clustered SE	Yes	Yes	Yes	Yes
R-squared	0.02	0.02	0.02	0.01
Observations	608	608	608	608

Standard errors in parentheses

Bootstrapped standard errors in parentheses.

Dependent variable: Mean Prediction Error.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table SI3: **Correlations between errors and covariates at 205-270V threshold.** We note that errors tend to be larger for places with large literate population (i.e., richer towns) and smaller for places with a larger schedule tribe population. The number of monitors is also marginally significant indicating that places with more voltage monitors tend to have smaller errors.

	(1) 24h	(2) Daytime	(3) Nighttime	(4) 12PM-3AM
Nightlights (sum, log)	3.65*** (0.83)	1.63*** (0.39)	1.96*** (0.50)	0.67*** (0.18)
Village/Town FE	Yes	Yes	Yes	Yes
R-squared	0.87	0.88	0.81	0.77
Observations	604	604	672	672

Standard errors in parentheses

Standard errors in parentheses. Clustered at the village/town level.

Dependent variable: Number of Hours with Normal Electricity.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table SI4: Correlations between Reliability and Nightlights Within Village/Town Over Time.

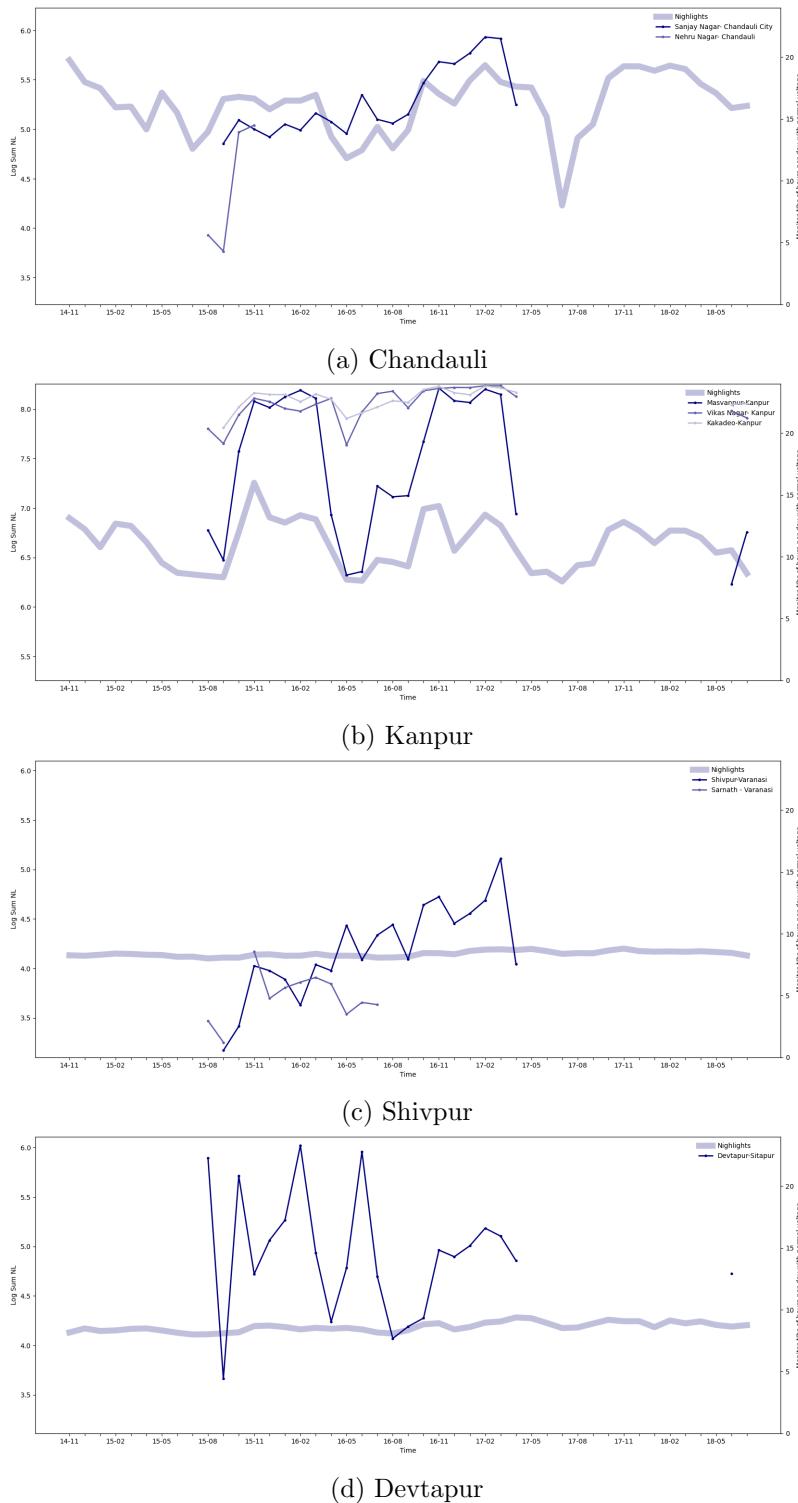


Figure SI15: Four Selected Locations Over Time and their Nightlights.