Monitoring Grid Voltage Fluctuations in Rural Schools in Maharashtra

Anhad Singh W.
Department of Mechanical Engineering, Imperial College London

Supervisor: Dr Adrià Junyent-Ferré
Dept. Electrical and Electronics Engineering, Imperial College London

Date: October 31, 2025
Word count: *(word count placeholder)*

Acknowledgements

I would like to thank Jagjeet Singh and Ravinder Singh Dhaliwal for their tireless assistance in locating field sites, procuring and sorting wiring and hardware and for the many hours of practical help during the deployment. I acknowledge Dr Adrià Junyent-Ferré for supervision and guidance, the Laidlaw Scholars Programme for funding, and the participating schools and local contacts in Nanded for hosting the equipment.

Abstract

This pilot study deploys compact, low-cost mains-powered voltage loggers at school distribution points in the Nanded district of Maharashtra to characterise short-duration voltage depressions and baseline variability. Devices record time-stamped line voltage at approximately one sample every 85 seconds to SD-card CSV files. Three units were deployed; one dataset was validated, one file is under recovery, and one unit is pending data retrieval. Preliminary inspection indicates repeated short-duration voltage depressions ("notches") that produce a left-tail in the voltage distribution. This report documents the hardware and assembly, deployment protocol, data quality-control procedures, and the analysis plan for event detection and aggregate metrics. Results from the validated dataset are presented as preliminary; full multi-site conclusions await recovery and QA of the remaining data. The cleaned dataset and code will be published after post-processing.

Keywords: Voltage monitoring; power quality; rural electrification; data logger; Maharashtra.

Contents

1 Introduction

Reliable electrical supply is a fundamental prerequisite for modern education, healthcare and economic activity. However, in many parts of rural India the availability and quality of electricity remain far from dependable: distribution networks are often aged, poorly financed and subject to significant technical and non-technical losses, and many households and public services continue to experience frequent outages and low voltages [3].

These systemic issues manifest at multiple scales. At the national scale, episodic large-area blackouts have revealed structural fragilities in grid planning, transmission capacity and operational coordination, with consequences for both service continuity and resilience. Analyses of India's massive past blackouts highlight how limited transmission capacity, fragmented governance and chronic congestion can allow local disturbances to cascade into wide-area failures [2]. At the distribution level, overloaded or poorly maintained feeders produce persistent voltage deviations and frequent temporary disconnections, phenomena that are not captured by simple electrification metrics [4].

Voltage quality matters because low voltage and short-duration depressions directly disrupt essential services in rural schools and clinics (lighting, fans, medical devices) and increase reliance on costly backup generation or curtailed operation. Empirical work from Maharashtra and comparable contexts links outage frequency and duration to measurable adverse service impacts, underscoring the social as well as technical costs of unreliable supply [1].

Despite this importance, detailed, openly available, time-resolved voltage traces from distribution-level connection points in rural India are scarce. Public datasets typically report aggregated outage counts or energy-access statistics but rarely provide high-resolution voltage series that would allow engineers and data scientists to characterise event statistics, validate event-detection algorithms, and evaluate whether single-point voltage observations can be used to infer impending overloads or disconnections. This gap constrains measurement-driven interventions (for example, microgrid control, targeted reactive support, or local protection strategies) and prevents the wider community from benchmarking resilience-improving techniques under realistic field conditions [3, 4].

This project addresses that gap by deploying compact mains-powered voltage loggers at rural school distribution points in Maharashtra to collect time-stamped voltage traces together with contextual metadata. The aims are twofold. First, to quantify short-term voltage variability and the occurrence statistics of short-duration voltage depressions and interruptions that disrupt local loads. Second, to create an academically curated, open dataset of distribution-level voltage time series for use by engineers, modellers and policy researchers. The dataset is intended to support (i) statistical characterisation of events (rates, depths, durations), (ii) development and evaluation of near-term predictors of local supply collapse, and (iii) investigation of correlations between voltage observations and contextual factors such as network loading patterns or local generation [2, 1].

The remainder of this report documents the measurement hardware and deployment strategy, the dataset status and preliminary observations from the first site, and a proposed analysis pipeline to be applied once a larger multi-site dataset is complete. By publishing both the measurement methods and the raw traces we aim to enable reproducible research on rural power quality and lower the barrier for low-cost, data-driven solutions to improve electricity reliability in resource-constrained settings [4, 3].

2 Objectives

- Deploy low-cost, bench-tested voltage logging devices at rural school distribution points in Nanded.
- Collect and curate time-stamped voltage data suitable for event detection and statistical

characterisation.

- Define and detect voltage depressions ("notches") and compute per-site metrics (frequency, duration, depth, area).
- Provide a reproducible processing pipeline and an anonymised dataset for publication.

3 Methods

3.1 Device design and components

Each deployed unit consists of:

- Fused mains plug (230 VAC input).
- 240 VAC to 24 VAC step-down transformer (insulated secondary).
- Bridge rectifier (full-wave).
- Smoothing electrolytic capacitor (470 μF, 35 V recommended).
- Voltage divider and small series resistor feeding ADC input.
- Pre-built ADC/data-logger (RTC + SD card) configured to log CSV rows.

Note that the reason for the voltage divider was to ensure compatibility with the pre-built data-logger which had an operating max voltage of 12V. A simplified wiring schematic is shown in Figure 1.

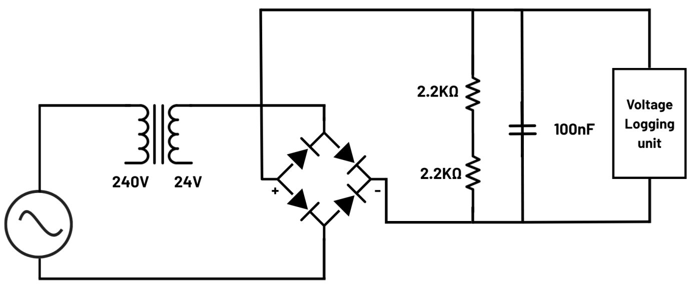

Figure 1: Simplified wiring schematic: mains → step-down transformer → bridge rectifier → smoothing capacitor → voltage divider → Logger (VIN/GND).

The final assembled circuit is shown in Figure 2.

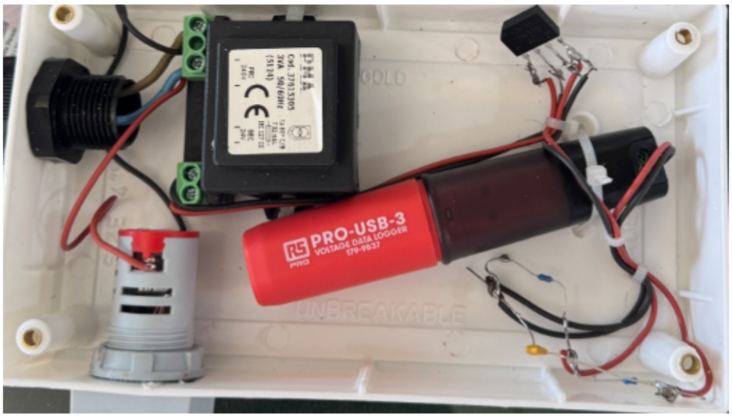

Figure 2: Final assembled circuit

3.2 Assembly and bench testing

All units were assembled inside IP-rated enclosures. Mechanical fixation (standoffs, hot-glue) was used to prevent movement inside the box. Additionally, screw holes were drilled into place to prevent accidental openings of the logger circuit. Bench testing included:

1. DC measurement across smoothing capacitor under no-load and light-load conditions.
2. Verification of the ADC reading vs. calibrated multimeter across a range of voltages.
3. 48-hour bench run to test logging integrity and timestamp stability.

Units were labelled with device IDs and build dates.

Figure 3: Photo showing the tight screws on the box

Figure 4: Photo showing the sealing of the screwed lid

3.3 Deployment protocol

Sites were selected in Nanded city prioritising accessibility and school cooperation. Before deployment:

- Host consent form signed (copy archived).
- Device ID, approximate site description and district-level GPS recorded (exact coordinates stored separately and anonymised before public release).
- Initial clock synchronisation performed on the logger RTC.

3.4 Sampling and data format

Loggers were configured for roughly one sample every 60 s (32,000 sample capacity). CSV format:

```
timestamp_ISO, deviceID, voltage_V, status_flag
2025-08-24T10:12:00+05:30, A, 239.1, OK
```

3.5 Quality assurance (QA)

Automated checks:

- Timestamp parseability and monotonicity.
- Sample interval distribution and median sampling interval.
- Long gaps detection (gap $>$ 5$\times$ median interval).
- Spike detection (isolated samples outside a robust threshold).
- File integrity checks (CSV parse errors flagged for manual inspection).

Corrupt files are documented, attempts at recovery performed, and missing/invalid files are recorded in a dataset status table.

3.6 Event detection

In the manuscript we avoid informal labels and adopt standard power-quality terminology. Two classes of short-duration abnormal behaviour are distinguished: (i) *voltage dips* (also called short-duration voltage depressions or sags), where the measured voltage falls below the local operating level but remains substantially above zero, and (ii) *voltage interruptions*, where the measured voltage falls to near-zero for a finite interval.

Practically, we detect dips and interruptions relative to a local rolling baseline. The baseline is defined as the centred rolling median of the voltage trace computed over a 1-hour window. A **voltage dip** is flagged when the instantaneous voltage satisfies

$$V(t) < \alpha \times V_{\text{baseline}}(t),$$

with a primary threshold $\alpha = 0.9$ used in this study. For sensitivity analysis we test α in the range 0.85–0.95. We reserve the term **voltage interruption** for events where the instantaneous voltage falls below a small fraction of the baseline (for example $V(t) < 0.1 \times V_{\text{baseline}}(t)$), i.e. where the supply is effectively absent at the measurement point for the duration of the event. (Threshold values are configurable and reported alongside results.)

Contiguous samples satisfying the dip or interruption criteria are grouped into discrete events. For each event we compute the following metrics: start and end timestamps, duration (seconds), minimum recorded voltage (V), depth (defined as the difference between the baseline and the event minimum, in V), and the integrated deficit (the time integral of baseline minus measured voltage, reported in $\text{V} \cdot \text{s}$).

4 Data and preliminary results

4.1 Dataset status and availability

Table 1 summarises the current dataset status.

Table 1: Dataset status (preliminary)

Site	Data received	Data integrity
A	Yes	Intact (validated)
B	Yes	Corrupt (under recovery)
C	Yes	Intact (validated)

4.2 Preliminary observations (Unit A)

Three logging units (A, B and C) were deployed in Nanded between 30/09/2025 and 16/10/2025. At the time of writing, Unit A yielded a validated CSV, Unit B produced a corrupt file currently under attempted recovery, and Unit C's SD-card is pending retrieval. Table 1 summarises the current status.

Preliminary inspection of Unit A's time series indicates an otherwise near-steady DC level after rectification with repeated short-duration depressions ("notches") superimposed. These notches appear as abrupt downward excursions relative to a local rolling median baseline and produce a measurable left-hand tail in the distribution of recorded voltages. Basic descriptive statistics for Unit A are: mean voltage = `194.9` V, minimum = `0` V, maximum = `260` V (placeholders to be updated after final data cleaning).

Given the limited multi-site coverage at present, the results presented here are explicitly preliminary. The planned full analysis will aggregate event statistics (frequency, duration, depth, integrated deficit) across sites and perform sensitivity testing on baseline and threshold parameters. The current single-site observations motivate extended deployments and comparative studies planned under the India Connect programme.

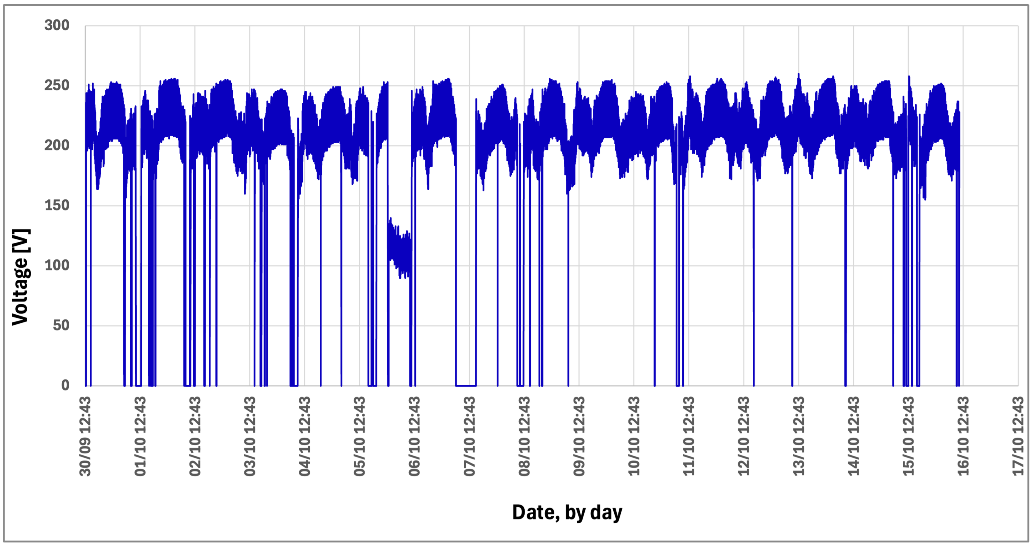

Figure 5: Example voltage time series (Unit A). Shaded regions indicate detected notches (preliminary).

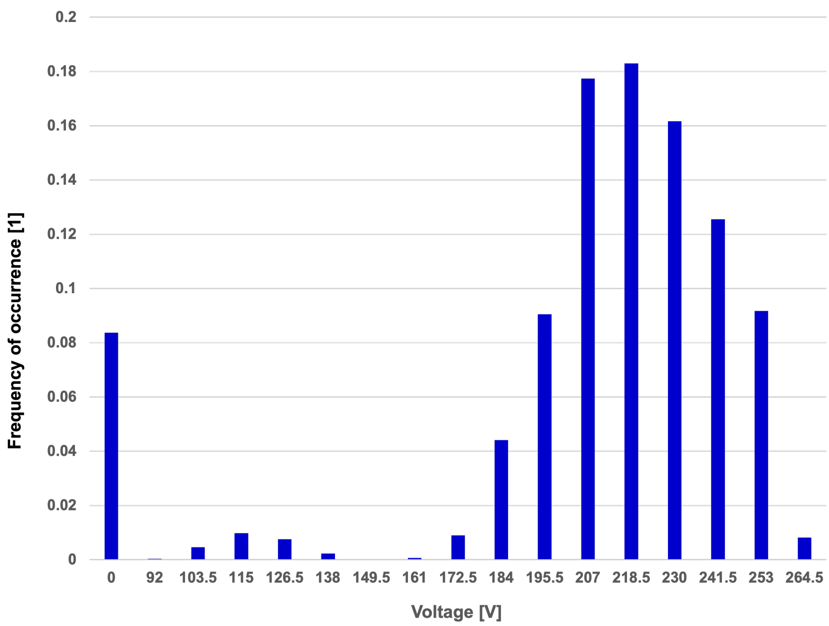

Figure 6: Normalized histogram of recorded voltage values (Unit A). Dashed line: 0.9×baseline (preliminary).

5 Analysis plan / Proposed analyses

This section describes the processing pipeline and the statistical and predictive analyses that will be performed once the larger, cleaned multi-site dataset is available. The primary aims are twofold: first, to quantify the descriptive statistics and event-level behaviour of voltage traces at rural distribution points and, second, to evaluate how much information a single measured variable (voltage) contains about imminent supply interruptions and system state. In parallel we will assemble the Maharashtra Rural Voltage Dataset (MRVD), a centralised, documented and filterable repository of cleaned voltage logs and site metadata intended to accelerate engineering and data-science investigations.

5.1 Pre-processing and data cleaning

All raw logger outputs will be converted to a standard CSV format and imported into a scripted processing pipeline. Timestamps will be parsed and converted to timezone-aware datetimes, and traces will be resampled to a uniform 60 s grid using mean aggregation; intervals that exceed an adjustable threshold will be marked as missing. Short, isolated gaps (on the order of a few consecutive samples) will be interpolated linearly to preserve continuity for event detection, while longer gaps will be retained as missing values so that downstream statistics account for actual uptime. Records exhibiting sustained constant values or other clear sensor artefacts will be identified and removed or annotated in the dataset. For each site we will compute basic provenance and completeness metrics (first/last timestamp, received samples, expected samples, uptime%) and store these alongside site metadata in the MRVD repository.

5.2 Event detection and aggregation

Voltage depressions ("notches") will be detected relative to a local rolling baseline to capture departures from the local operating level. The baseline will be defined as a centred rolling median (primary window: 1 hour) and events will be identified where the instantaneous voltage falls below a fraction of that baseline (primary threshold: $0.9\times$baseline). Contiguous samples satisfying the criterion will be grouped into discrete events and annotated with start and end times, duration (s), minimum voltage, depth (baseline minus minimum) and an integrated deficit ($V \cdot s$). Per-site aggregates will include descriptive statistics (mean, median, standard deviation), event frequency (events per day), median and interquartile event duration, median depth, and the fraction of samples below candidate thresholds. Sensitivity analyses will explore alternative thresholds (e.g. 0.85–0.95) and baseline window lengths (30–120 minutes) to characterise robustness.

5.3 Statistical comparisons and pattern discovery

We will look for systematic temporal patterns and cross-site differences using a combination of summary statistics and hypothesis tests. Diurnal structure will be examined by computing hour-of-day means and by visual inspection of day$\times$hour heatmaps; non-parametric tests (e.g. Mann–Whitney U) will be used to compare day versus night event rates and other non-normally distributed metrics. Where multiple validated sites are available, differences in event statistics will be assessed using Kruskal–Wallis tests. Time-series diagnostics (autocorrelation and partial autocorrelation functions) will be computed to assess serial dependence and inform feature construction for predictive models.

5.4 Exploratory prediction and evaluation

A central scientific question is how much predictive signal is present in voltage alone: specifically, can recent voltage behaviour be used to forecast a short-term supply interruption or a voltage notch with useful lead time and acceptable accuracy? To address this we will construct compact, interpretable predictive models that use sliding-window features and attempt to leverage machine learning to analyse the dataset to predict potential outage times before they happen.

5.5 Repository and reproducibility

All cleaned traces, per-file provenance metadata (timestamps, uptime, deviceID), site-level descriptors (district, site type, deployment notes) and the processing scripts will be assembled into the MRVD repository. The repository will be organised to allow filtering by district and by simple site attributes, and will include a README describing the data cleaning rules and event definitions. By publishing MRVD we aim to provide a common, reproducible basis for follow-up engineering studies (for example, storage/ride-through sizing, inverter control strategies or targeted field interventions) and to enable comparison of methods across research groups.

6 Discussion

The early signals of repeated short-duration depressions suggest transient network stress. If events are frequent and deep, ride-through or short-duration storage may be needed to maintain service quality. However, single-site evidence is limited: stationarity across seasons and a larger sample of sites are required for robust recommendations. Sensor limitations (sampling rate, ADC range, measurement chain) must be considered when interpreting event depth and duration.

7 Conclusions and next steps

7.1 Conclusions

This pilot validates the field-deployable approach to collecting voltage time-series at school distribution points. Preliminary results show notches that warrant detailed cross-site analysis.

7.2 Immediate next steps

- Recover corrupt files and retrieve missing SD card(s).
- Finalise QA and run event detection across all valid sites.
- Produce the three core figures and summary tables; prepare the cleaned dataset for MRVD release following ethics approval.
- If event rates justify, plan a larger deployment to sample multiple districts and seasonal variation.

8 Ethics, data management & sharing

Signed host consent forms are archived. Exact installation coordinates are stored separately and will be anonymised (district-level) for public release. The cleaned dataset (MRVD) and processing scripts will be published under an open data licence once QA and institutional ethics clearance are complete.

References

References

[1] Author(s) unknown – please replace. *e001372.full.pdf*. PDF. Local file: "e001372.full.pdf". 202X.

[2] Author(s) unknown – please replace. *Generalized congestion of power systems: insights from the massive blackouts in India.* PDF. Local file: "Generalized$_{c}ongestion_{o}f_{p}ower_{s}ystems_{i}nsights_{f}rom_{t}he_{m}as$ 2015.

[3] Author(s) unknown – please replace. *Rural electricity access in India in retrospect - A critical rumination.* PDF. Local file: "Rural electricity access in India in retrospect - A critical rumination.pdf". 2020.

[4] Author(s) unknown – please replace. *Solar Energy, DC Distribution and Microgrids: Ensuring Quality Power in Rural India.* PDF. Local file: "Solar$_{E}nergy_{d}c_{D}istribution_{a}nd_{M}icrogrids_{E}nsuring_{Q}uality_{P}$ 2019.

A Appendix A: Data cleaning and exploratory analysis (Excel workflow)

The initial data processing and exploratory analysis were performed in Microsoft Excel from the raw logger `.txt` files converted to `.csv`. The following steps document the reproducible procedure used for Unit A and can be followed to replicate the preliminary results reported in this paper.

1. **File conversion & import.** Convert the raw `.txt` logger output to UTF-8 `.csv`. Import into Excel via `Data` → `From Text/CSV`, ensuring timestamps are parsed correctly as text or ISO datetimes where required.

2. **Header handling & parsing.** Remove non-data header/footer rows inserted by the logger. Standardise the timestamp column to ISO 8601 where necessary and convert text timestamps to Excel datetime format.

3. **Sorting & duplicate removal.** Sort rows by timestamp and remove exact duplicate rows. Compute the inter-sample interval (e.g., `=A2-A1` formatted as seconds) to verify sampling cadence and identify clock jumps.

4. **Uptime & gap detection.** Estimate the expected sample count from the first and last timestamps using the nominal 60 s sampling period. Compute *uptime* as `received_samples / expected_samples`. Flag gaps where the inter-sample interval exceeds a threshold (e.g., $> 5 \times$ median interval) using conditional formatting.

5. **Basic descriptive statistics.** Use Excel functions to compute: `=AVERAGE()`, `=MEDIAN()`, `=STDEV.P()`, `=MIN()`, `=MAX()` on the voltage column. Compute percentage of samples below thresholds using `=COUNTIFS()`. where `k` corresponds to the number of samples on either side (for ∼1 hour at 85 s cadence use $k \approx 42$). `IFERROR()` can be used to handle edge cells. A moving average using `AVERAGE()` and `OFFSET()` was also used for simplicity in some checks.

6. **Event (notch) flagging.** Create a Boolean event column using a threshold relative to the baseline:

   ```
   =IF(voltage_cell < 0.9*baseline_cell, 1, 0)
   ```

   Collapse contiguous runs of ones into events by detecting event starts (event=1 and previous event=0) and counting consecutive samples per event to compute duration (multiply sample count by 60 s).

7. **Histograms & frequency tables.** Build a manual frequency table by defining bin edges (e.g., 2–5 V bins) and computing counts via `FREQUENCY()` or `COUNTIFS()`. Insert a Clustered Column chart from the frequency table and set *Gap Width* to ∼30–40% to increase bar thickness. Normalise counts by total samples to produce a density-style histogram.

8. **Documentation & provenance.** Save Excel workbooks with versioned filenames (e.g., `siteA_v1_clean.xlsx`) and include a change-log worksheet that records each manual edit, conversion or correction. Archive raw `.txt` files and SD-card images alongside cleaned outputs.

Note: Excel was selected for rapid, reproducible exploratory analysis. The full analysis pipeline (event detection replicated across multiple sites, sensitivity testing and modelling) will be implemented in a scripted environment (Python/R) once the complete multi-site dataset is available to ensure scalability and rigorous provenance.

B Appendix B: Bill of materials (BOM) and device specs

Component	Specification
Smoothing capacitor	470 μF, 35 V, electrolytic, low-ESR
Step-down transformer	230V to 12V AC, rated for 2-5 W (select per load)
Bridge rectifier	1 A or higher, 50 V rating
Logger	Pre-built ADC logger with RTC + SD support
Casing	IP65/66 rated enclosure (recommended)

C Nomenclature

Symbol	**Meaning**
V	Voltage (volts)
V_{baseline}	Rolling median baseline voltage (V)
Δt	Sample interval (s)
$MRVD$	Maharashtra Rural Voltage Dataset