Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A provenance-tracked State-day panel of India's power system built from official public reports, 2018–2026

Rohit Kumar

therohitkumar.build@gmail.com

Dr. B. R. Ambedkar National Institute of Technology Jalandhar https://orcid.org/0009-0004-7936-9683

Research Article

Keywords: power system data, India, electricity demand, renewable generation, electricity prices, data provenance, publication delay, open data, State control areas, energy forecasting

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11546744/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

India publishes daily information on electricity demand, scheduling, generation and prices, but as thousands of report files whose layouts change, spread across the grid operator, the Central Electricity Authority and the meteorological department. We describe a State-day panel for 34 Indian State control areas from 2018-04-01 to 2026-08-31 (104,550 State-days, 107 variables). It joins, from official sources only, energy met and shortages, scheduled and actual drawal with over- and under-drawal, conventional generation by fuel, renewable generation, bid-area exchange prices and deviation charge rates, system frequency stress, carbon intensity and weather. Each of the 9,614 source files it uses is recorded with its URL, retrieval time and SHA-256 hash, and the panel can be rebuilt deterministically from the raw files. Each variable carries an availability delay, based on measured publication times where available, so forecasting studies can respect what was known each day. Technical validation reproduces the sources' own totals exactly and agrees with an independent copy of the generation data within 1% on 100.0% of 75,485 State-days.

Background & Summary

India operates one of the world's largest synchronous grids, and its States are both the units of demand and the units that are scheduled, settled and regulated. Each day the national load despatch centre of Grid Controller of India Limited (Grid-India) publishes a Power Supply Position report with State demand, scheduled and actual drawal and the resulting over- or under-drawal [1]. The Central Electricity Authority (CEA) publishes station-wise generation [2] and State renewable generation [3] and maintains the national CO_2 baseline database [4]. Grid-India publishes the bid-area exchange prices and deviation charge rates used for settlement [5] under successive deviation settlement regulations [6,7]. The India Meteorological Department (IMD) distributes gridded daily rainfall and temperature [8–10].

These sources are public, but they are not a dataset. They are PDF and spreadsheet reports whose sheet order, column positions, date conventions and even file names change over the years. Several failure modes affect naive use:

- report pages that carry the previous day's date;
- two differently defined renewable tables in the same workbook, in varying order;
- a web archive that points many report dates to one shared file;
- typing errors in station records;
- sources that move to a new address or stop publishing without notice.

Third-party aggregations of some of these reports exist, such as the copy from the India Data Portal used here for a cross-check (Technical Validation), but their processing is not documented in a way that allows them to be rebuilt from the original reports, which limits reproducibility and comparison.

This data descriptor documents a panel built for a research programme on forecasting, context-aware assessment of schedule deviations and risk-aware scheduling. It follows three principles.

1. **Official sources only.** Every value comes from a report published by Grid-India, CEA or IMD. Aggregator copies are used solely as an external cross-check.
2. **Provenance and determinism.** Each raw file is stored unchanged with URL, retrieval time, size and SHA-256 hash; parsers and build steps are deterministic, and the panel carries a content checksum (8099e9fdc72c7ea7…).
3. **Information-set awareness.** Each variable is annotated with the delay after which it is publicly available, set from publication times measured in the publishers' own file metadata where these exist, so that forecasting and decision studies do not use data that were not yet published.

The design follows the documentation practice of datasheets for datasets [11], the FAIR principles [12] and the processing transparency advocated for European power system data [13]. In Europe, a central transparency platform and processed open datasets of this kind exist [13,14], together with reanalysis-based series of renewable output [15,16], and open data and software are increasingly seen as a condition for reproducible energy research [17]. Data descriptors document open power-system data elsewhere: a transmission, generation and demand dataset for a highly renewable European system [18], a harmonised open dataset for the Brazilian power system [19], open maps of the global power grid [20] and of wind and solar farms [21], daily consumption-based carbon intensity of European power systems with a one-day latency [22], and a multi-scale time-series benchmark for machine learning on power grids [23]. For India, published datasets project future electricity demand [24] and estimate urban energy use by district [25]; they do not cover the daily operational reports of the State control areas. This descriptor documents such a panel, built from the official reports alone. Two companion manuscripts, on day-ahead forecasting and on day-ahead scheduling, use version 1.0 of the panel; neither had been published when this descriptor was written. This descriptor documents version 1.1, which adds the renewable generation reports that the Central Electricity Authority has published on a new portal since November 2025.

Figure 1 summarises the workflow. The panel (104,550 State-days × 107 columns, 10.25 million values) covers 34 State control areas and 3,075 days.

Methods

Sources and acquisition

Table 1 lists the sources. All files were downloaded without login from public pages by scripted HTTP requests on 14–15 September 2026, and the renewable reports published on CEA's new portal on 25 October 2026. Grid-India reports were enumerated through the file listing interface of its website, which

returns one record per published attachment (4,834 records for the Power Supply Position reports). CEA National Power Portal reports follow a date-based URL pattern. CEA renewable reports up to 18 November 2025 were enumerated from the listing behind the report page of cea.nic.in. From November 2025 CEA publishes the same report on its RE generation portal (https://gen-re.cea.gov.in/reports), one file per day under a date-based name; the old listing has no later report. Version 1.1 of the panel takes the portal files from 1 November 2025, and the old listing is kept on every date it covers. IMD gridded files are served per year and variable.

Every request, including those that failed, is written to a per-source manifest with URL, local path, HTTP status, size, SHA-256 hash and UTC retrieval time. A response smaller than a source-specific minimum size is recorded as unavailable rather than stored. During acquisition a local network login page intercepted some requests; such pages are detected by content and are never stored, and the affected requests were retried. The raw layer holds 18,233 official files (5.03 GB). The 9,614 files of Table 1 (4.53 GB) are the inputs of the panel; the other 8,619 (National Power Portal daily sub-reports 6 and 10, its daily coal report and two regulation documents of the Central Electricity Regulatory Commission) were downloaded for checks, are kept with their manifests and are not used in this version. An earlier pilot collection of aggregator copies (162 files) is kept unchanged with its own hash manifest and is used only for cross-checks.

Table 1. Official sources used in the panel. Files are those retrieved successfully; “not available” counts requests that returned an error or no report.

Publisher and product	Content used	Files	Not avail.	Size (MB)
Grid-India, daily Power Supply Position report	State demand, shortages, drawal schedule, over/under-drawal, max over/under-drawal; regional generation and inter-regional flows; frequency profile	3,448	5	2667.2
Grid-India, deviation settlement rate files	bid-area day-ahead and real-time clearing prices, normal rate of charges for deviation	404	0	87.6
CEA National Power Portal, daily generation report (sub-report 2)	station programme and actual generation, monitored capacity, capacity under outage, coal-stock days	3,022	74	1053.1
CEA, daily renewable generation report (cea.nic.in listing, to 18 Nov 2025)	State wind, solar and other renewable generation	2,361	8	382.3
CEA, daily renewable generation report (RE generation portal, from 1 Nov 2025)	as above	304	1	13.4
CEA, CO_2 baseline database (v22.0 and earlier)	station net generation and absolute emissions by financial year	48	0	68.0
IMD Pune, gridded daily data	rainfall ($0.25°$), maximum and minimum temperature ($1°$)	27	0	254.2

Entities

The unit of analysis is the State control area as reported by Grid-India. There are 34 entities: all 28 States; five Union-Territory control areas covering six of India's eight Union Territories, namely Delhi, Chandigarh, Puducherry, Dadra and Nagar Haveli and Daman and Diu (reported separately before their merger and summed into one entity), and Jammu and Kashmir together with Ladakh, which the reports give as one area; and the control area of the Damodar Valley Corporation, a central utility. The Andaman and Nicobar Islands and Lakshadweep do not appear in the reports. Industrial and railway entities that appear in some reports are excluded. A registry maps every spelling used by any source to an entity, and assigns each entity to its region and exchange bid area; no label in any parsed report remained unmapped.

Parsing and harmonisation by source

Power Supply Position reports.

Spreadsheet reports are read by locating the State table and mapping columns from their header text; positional reading fails on files that contain empty spacer columns. PDF reports are read from word

coordinates: lines are clustered by vertical position, and numbers are assigned to columns by their horizontal centres. The data date is the printed “Date of Reporting” minus one day.

In 2 cases, two reports with different figures carried the same printed date because the table page had not been updated. The report whose cover letter implied a different, otherwise uncovered date was moved to that date. Within the study window from 2018-04-01 to 2026-08-31, 129 data dates (4.2%) have no usable report. Of these, 90 are caused by the website itself: its listing points 118 report titles to only 23 files, one of which is shared by 72 report dates.

Deviation settlement rate files.

Three layouts correspond to regulatory regimes, which the panel records per day:

- **R1** (2018-12-19 to 2022-12-04): frequency-linked charges with daily average exchange prices.
- **R2** (2022-12-01 to 2024-09-15): the 2022 Regulations [6], in force from 5 December 2022. Files in the new layout are published from 2022-12-01, so both layouts exist for the first days of December 2022, and the panel keeps the later listing.
- **R3** (from 2024-09-16): the 2024 Regulations [7]. The normal rate of charges for deviation is the highest of the day-ahead clearing price, the real-time clearing price and an ancillary-service-based rate.

Rates published in paise per kWh are converted to rupees per MWh. Block values are summarised per day and bid area as mean, maximum, standard deviation and the share of blocks at the regulatory cap. Where the same date appears in several files, the later listing is kept and the choice logged.

Daily generation report.

Each spreadsheet is a hierarchy of region, State, sector, fuel type, station and unit rows. Column positions differ between years and are located from header text. In some files, sector and type rows appear without labels or carry subtotal values; these are recognised and skipped. The sum of station rows equals the report’s own State total on every accepted date (largest absolute difference 4.5×10^{-13} GWh). 3 files with a shifted layout and no total rows cannot be validated and are excluded.

Station coal-stock days outside 0–180 are treated as entry errors. State-days containing an isolated station value above three times its capacity-based maximum, while the station’s recent capacity factor is normal, are flagged. Stations are aggregated by entity and fuel (coal including lignite, gas, diesel, hydro, nuclear).

Daily renewable generation report.

A workbook contains up to three tables in varying sheet order:

- renewable generation located in the State, including inter-State generating stations;

- renewable generation in the State's own control area (from 2020-06-01);
- a station list for inter-State stations.

Tables are classified by the title printed inside each sheet and checked by magnitude, because the first total can never be below the second; both definitions are released.

The date printed in the report is used. The website listing date differed from it for 10 files. Rows are accepted when wind, solar and other renewable generation add up to the total. Files that omit the "other" column (58 State-days) are recognised because wind plus solar then equals the total. Rows matching neither layout (47 and 23 State-days for the two definitions) are set to missing. One file containing the same table twice is not used. The portal files have the same layout and are read by the same parser.

Carbon intensity.

Station names in the CO_2 database and in the daily generation report do not match reliably, so emission factors are built per entity, fuel and financial year. A factor is CEA absolute emissions divided by CEA net generation. Because the daily report gives gross generation, factors are multiplied by the net-to-gross ratio of each entity and fuel over the financial years in which the daily report is complete (national ratios: 0.927 for coal, 0.966 for gas).

Daily emissions are gross generation by fuel times the factor. A database version is published months after its financial year ends (version 22.0 in August 2026 [4]). The panel therefore provides factors lagged by two financial years for use as forecast inputs, and by one financial year for ex-post accounting. A consumption-based intensity adds net drawal at the day's regional generation intensity.

Weather.

IMD binary grids are read as day $\times$ latitude $\times$ longitude arrays; the orientation was verified on land and sea cells and on known climatology. Entity weather is the mean over one to four load centres (capital and largest demand cities). It uses the nearest valid temperature cell and the mean of rainfall cells within $\pm 0.25^{\circ}$. Cooling and heating degree days use base temperatures of 24 °C and 18 °C.

Quality control

Source tables are never altered. Values that fail a plausibility rule are set to missing in the panel and flagged:

- energy met of zero or more than five times its 29-day median (6);
- maximum demand inconsistent with energy met (16);
- energy shortage exceeding energy met (18);
- over/under-drawal exceeding energy met plus scheduled drawal (2);

- generation above capacity × 24 h with a 15% overload margin (13 State-days);
- isolated station typing errors (145 State-days).

Missing values are not imputed. The only fill values are structural zeros: a fuel with no monitored capacity in an entity on a reporting date is recorded as zero generation.

Publication timing and information sets

For each variable the field registry records an availability delay L in days: a forecast issued on day τ may use the variable up to date $\tau - L$. The registry uses one day for Power Supply Position and weather variables, two days for generation, renewable and carbon variables, and 14 days for prices. These delays are assigned per source. For the two Grid-India products they were set from upload times measured in the publisher's file metadata, which also show how often the assigned delay holds:

- Power Supply Position reports appear a median 10.6 h after the end of the data day (90th percentile 36.7 h; 1,064 uploads for report dates from 1 April 2025). For 76.8% of 496 report dates the first upload was online by noon of the next day, the latest time at which a one-day delay holds for a forecast issued at noon.
- Daily deviation rate files are uploaded in batches a median 8.8 days after the data date (range 4.7–71.7 days, 267 files from January 2025); 92.1% arrived within the 14 days of the registry.

No upload times were measured for the Central Electricity Authority and IMD sources, so their delays are assumptions based on how the reports are issued; IMD's gridded files in particular are final products released after the year ends (Usage Notes). The registry's delays are therefore defaults that hold for most but not all days, and studies that need a strict real-time information set should lengthen them.

Use of AI tools

During the preparation of this manuscript, the author used Claude (Anthropic) to assist with English language editing, drafting of revised text, consistency checking across tables and narrative text and formatting of the revised manuscript. All experimental design, data collection, model development, the original code implementation, and interpretation of results were performed entirely by the author. The author carefully reviewed and verified all AI-assisted output and takes full responsibility for the content of this publication.

Data Records

The files described in this section are version 1.1 of the RE-FUSED record on Zenodo [26]. The main record is refused_state_day.parquet, one row per entity and day from 2018-04-01 to 2026-08-31 (104,550 rows, 107 columns). It is accompanied by the following files:

- refused_entities.csv: entity registry;

- FIELD_REGISTRY.csv: source, unit, availability delay and missing share of every column;
- QC_REPORT.md: quality-control counts and cross-source checks;
- DATASHEET.md: datasheet;
- SOURCE_REGISTRY.md: source registry;
- CHECKSUM.txt: content hash;
- the interim tables at native resolution, including station-day generation (1,547,930 rows), bid-area prices, entity–fuel–year emission factors (453 cells) and entity-day weather.

Columns are grouped by prefix. dem_ is demand, dev_ deviation, gen_ conventional generation and re_ renewables. prc_ holds prices and co2_ carbon. sys_, reg_ and nat_ are system, regional and national context, and wx_ is weather. cal_, avail_ and qc_ are calendar fields, source availability flags and quality flags.

Energies are in GWh (1 MU), power in MW, prices in rupees per MWh and emissions in tonnes of CO_2.

Prices begin on 2018-12-19, the first published deviation rate file. State renewable data begin on 2019-07-01 and end on 2026-08-31. IMD gridded weather ends on 31 December 2025. Conventional generation is missing for 74 days between 13 March and 31 May 2020, when no daily generation reports were published. Maximum under-drawal is printed in the source reports only from 2023, which explains most missing deviation cells before 2025.

Data Overview

Figure 2 shows, for each source, the share of entity-days with data in every month, and Table 2 the mean share of missing cells by variable block before and after 1 April 2025. Gaps follow the sources: no value is imputed (Methods), and the dates at which sources begin and end are listed under Data Records.

Table 2. Mean share of missing cells by variable block, before 1 April 2025 and after.

Block	2018-04 to 2025-03 (%)	2025-04 to 2026-08 (%)
Demand	5.1	0.0
Deviation	20.9	10.3
Conventional generation	3.1	0.0
Renewables	27.4	3.2
Prices	18.5	0.0
Carbon	12.6	5.5
System stress	19.1	14.7
Regional and national	7.8	2.7
Weather	0.0	46.9

Technical Validation

Internal consistency with the sources' own totals.

For every accepted daily generation report, the sum of parsed station rows reproduces the State totals printed in the same report (largest absolute difference 4.5×10^{-13} GWh over 3,019 dates). This check detected three parser failure modes during development, which were corrected:

- subtotal rows without labels;
- sector and type rows carrying values;
- a report date stored only as a date cell.

For renewable generation, the in-report arithmetic (wind + solar + other = total) holds on all accepted rows by construction, and the classification of the two table definitions is checked by magnitude on every file. The 18 reports published on both CEA listings (1 to 18 November 2025) give identical values in all 5,040 parsed cells (0 differences), so the change of portal introduces no break in the series.

Agreement with an independent copy.

A previously collected aggregator copy of the same CEA reports was compared with the panel up to 31 March 2025 (Figure 3):

- **Conventional generation:** 100.0% of 75,485 State-days agree within 1%, and 99.1% within 0.01 GWh.
- **Renewable generation** (State including inter-State stations): 99.9% of 60,619 State-days agree within 1%.

The remaining renewable disagreements trace to the copy: it assigns some reports to the website's listing date, which for 10 files differs from the date printed in the report (for example, a report for 30 September listed on 30 October 2021).

Dates.

Grid-India's listing dates, the dates printed on report tables and the dates on cover letters were compared for all Power Supply Position reports. Collisions between different reports claiming the same date were resolved by the cover letter in 2 cases, and no report is used for two dates.

Physical plausibility.

- Quality-control exclusions are rare and listed individually (Methods).
- The national net-to-gross ratio of coal generation (0.927) is consistent with auxiliary consumption at coal stations. Entity-level coal emission factors lie between 0.84 and 1.04 t CO_2 per MWh of gross generation (median 0.91).
- Weather at the Delhi load centre has a mean January maximum temperature of 19.4 °C and a May mean of 39.7 °C. The grid cell containing Mumbai received 1,532 mm of rain in July 2023, confirming the grid orientation.

Publication timing.

Figure 4 shows the measured delays used for the information sets. Most Power Supply Position reports appear about half a day after the end of the data day, with a tail beyond a day at weekends. Deviation rate files arrive in weekly batches.

Determinism.

Rebuilding the panel from the raw files reproduces the content hash exactly; this was checked by a full rebuild of version 1.0 and by rebuilding version 1.1 from its renewable raw files in an empty folder. The hash is computed over row values independent of file encoding.

Usage Notes

Respecting information sets.

A model issuing a forecast on day τ may use a column only up to date τ minus its availability delay in the field registry. Using the latest row for every column assumes instant publication, which is not what grid operators, utilities or researchers observe. Carbon columns ending in _acc and IMD final gridded weather are ex-post products.

Known gaps and breaks.

- Scheduled drawal in the Power Supply Position report is the final schedule, including revisions made during the day. It is not a day-ahead schedule, and deviations measured against it are smaller than deviations from day-ahead positions would be.
- Deviations are daily net values. Over- and under-drawal within a day offset each other; maximum over- and under-drawal are the only intra-day severity indicators.
- Conventional generation is location-based and includes central stations whose output is scheduled to other States. It is not the energy supplied to the State.
- Price series change definition with the regulatory regimes R1–R3 (column prc_regime).
- Gridded weather ends on 31 December 2025, the last year that IMD had published. State renewable data from 19 November 2025 come from CEA's new portal (version 1.1); version 1.0 has none after 18 November 2025.

Figures in this descriptor.

Companion studies use the period from 1 April 2025 as a reserved evaluation period. For that reason, the figures and value comparisons here use dates up to 31 March 2025; the coverage figure shows availability only.

Reuse terms.

All source files are public reports of Government of India bodies. Their websites do not state an open licence. The release therefore consists of:

- the code and manifests needed to re-download and verify every raw file;
- the derived panel and interim tables, with attribution to Grid-India, CEA and IMD.

Raw report files are not redistributed. Users should cite the original publishers alongside this descriptor. After our downloads of September 2026 the Central Electricity Authority restructured its website: its daily renewable reports from September 2024 onwards are on the RE generation portal, but older renewable

reports and the CO_2 database files were no longer at the addresses in the manifests when checked on 25 October 2026. The manifests keep each file's address and SHA-256 hash, so a copy obtained from the authority can still be checked.

Declarations

Data Availability

The State-day panel is built only from reports published by Grid Controller of India, the Central Electricity Authority and the India Meteorological Department. The manifests in the record give the address, retrieval time and SHA-256 hash of every report used, so a copy obtained from the agencies can be checked; some older Central Electricity Authority files are no longer at the addresses recorded in

September 2026, because the authority has since moved its website. The panel described here is version 1.1.0 of the RE-FUSED record on Zenodo [26], at https://doi.org/10.5281/zenodo.23685972, with the panel and documentation under the CC BY 4.0 licence and the code under the MIT licence. It holds the State-day panel (refused_state_day.parquet), the entity and field registries, the quality-control report, the datasheet, the source registry, the content checksum, the interim tables at native resolution, the manifests of every downloaded file and the code that downloads, parses, builds and validates the panel. Version 1.0.0 (https://doi.org/10.5281/zenodo.23301898) is the earlier, frozen panel used by the companion studies; it differs only in the renewable columns from 19 November 2025 onwards.

Funding

This research did not receive a specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Code Availability

The acquisition, parsing, build and validation code is part of the release (Python 3.10). The steps are:

- acquire/ downloads the files and writes the manifests;
- b01–b06 parse each source into interim tables;
- b07 builds the panel, field registry, quality-control report and checksum;
- b08 recomputes every validation statistic reported here.

The numbers in this descriptor are generated from the validation output, and its figures from the panel, by two scripts included in the release. A leakage test checks that features built with the registry's availability delays do not change when unpublished values are perturbed.

The code, the panel and its documentation are released together on Zenodo, and the DOI of the version described here is https://doi.org/10.5281/zenodo.23685972. The code and build logs are also openly available at https://github.com/Ohi-Me/RE-FUSED.

References

1. Grid Controller of India Limited Daily Power Supply Position Reports, National Load Despatch Centre. Daily PSP Report page of grid-india.in, files listed through the site's public file interface; reports 2017–2026, accessed 14 September 2026, 2026. URL https://grid-india.in/en/reports/daily-psp-report
2. Central Electricity Authority (2026) Daily Generation Report, National Power Portal. Published reports of the National Power Portal, daily sub-report 2 (public-reports/cea/daily/dgr/DD-MM-YYYY/dgr2-YYYY-MM-DD.xls); 2018–2026, accessed 14 September 2026, URL https://npp.gov.in/publishedReports

3. Central Electricity Authority. Daily Renewable Generation Report, Renewable Project Monitoring Division. Reports to 18 November 2025 from the report listing of cea.nic.in, accessed 14 September 2026; reports from November 2025 from CEA's RE generation portal, accessed 25 October 2026 (2026) URL https://gen-re.cea.gov.in/reports
4. Central Electricity Authority. CO
$_2$
baseline database for the Indian power sector, user guide version 22.0. Ministry of Power, Government of India; database files downloaded on 14 September 2026 from the CDM CO
$_2$
baseline database page of cea.nic.in, which has since moved; every file URL and hash is in the release manifest (2026) August URL https://cea.nic.in
5. Grid Controller of India Limited. Deviation Settlement Mechanism Rate Files, National Load Despatch Centre DSM Rates page of grid-india.in; files 2018–2026, accessed 14 September 2026, 2026. URL https://grid-india.in/en/markets/dsmrates
6. Central Electricity Regulatory Commission Central Electricity Regulatory Commission (Deviation Settlement Mechanism and Related Matters) Regulations, 2022. Notification No. L-1/260/2021/CERC, dated 14 March 2022, 2022. URL https://www.cercind.gov.in/Regulations/168_reg.pdf
7. Central Electricity Regulatory Commission Central Electricity Regulatory Commission (Deviation Settlement Mechanism and Related Matters) Regulations, 2024. Notification No. L-1/260/2021/CERC, dated 5 August 2024, 2024. URL https://www.cercind.gov.in/regulations/192-Noti.pdf
8. Pai DS, Latha Sridhar M, Rajeevan OP, Sreejith NS, Satbhai, Mukhopadhyay B (2014) Development of a new high spatial resolution ($0.25^\circ \times 0.25^\circ$) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. MAUSAM 65(1):1–18. https://doi.org/10.54302/mausam.v65i1.851
9. Srivastava AK, Rajeevan M, Kshirsagar SR (2009) Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region. Atmospheric Sci Lett 10(4):249–254. https://doi.org/10.1002/asl.232
10. India Meteorological Department Gridded daily rainfall (0.25°) and maximum/minimum temperature (1°) data, Climate Monitoring and Prediction Group, Pune. Yearly binary files 2017–2025, accessed 14 September 2026, 2026. URL https://imdpune.gov.in/cmpg/Griddata/Rainfall_25_Bin.html
11. Timnit Gebru J, Morgenstern B, Vecchione JW, Vaughan H, Wallach (2021) Hal Daumé III, and Kate Crawford. Datasheets for datasets. Commun ACM 64(12):86–92. https://doi.org/10.1145/3458723

12. Mark D, Wilkinson M, Dumontier J, Aalbersberg et al (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3: 160018. https://doi.org/10.1038/sdata.2016.18
13. Frauke Wiese I, Schlecht W-D, Bunke C, Gerbaulet L, Hirth M, Jahn et al (2019) Open power system data – frictionless data for electricity system modelling. Appl Energy 236:401–409. https://doi.org/10.1016/j.apenergy.2018.11.097
14. Lion Hirth J, Mühlenpfordt, Bulkeley M (2018) The ENTSO-E transparency platform – a review of Europe's most ambitious electricity data platform. Appl Energy 225:1054–1067. https://doi.org/10.1016/j.apenergy.2018.04.048
15. Stefan Pfenninger and Iain Staffell (2016) Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy 114:1251–1265. https://doi.org/10.1016/j.energy.2016.08.060
16. Iain Staffell and Stefan Pfenninger (2016) Using bias-corrected reanalysis to simulate current and future wind power output. Energy 114:1224–1239. https://doi.org/10.1016/j.energy.2016.08.068
17. Pfenninger S, DeCarolis J, Hirth L, Quoilin S, Iain Staffell (2017) The importance of open data and software: Is energy research lagging behind? Energy Policy 101:211–215. https://doi.org/10.1016/j.enpol.2016.11.046
18. Tue V, Jensen, Pinson P (2017) RE-Europe, a large-scale dataset for modeling a highly renewable European electricity system. Sci Data 4:170175. https://doi.org/10.1038/sdata.2017.175
19. Ying Deng, Karl-Kiên Cao W, Stegen HR (2023) Kai von Krbek, Rafael Soria, Pedro Rua Rodriguez Rochedo, and Patrick Jochem. Harmonized and open energy dataset for modeling a highly renewable Brazilian power system. Sci Data 10:103. https://doi.org/10.1038/s41597-023-01992-9
20. Arderne C, Zorn C, Nicolas C, Koks EE (2020) Predictive mapping of the global power system using open data. Sci Data 7:19. https://doi.org/10.1038/s41597-019-0347-4
21. Dunnett S, Sorichetta A, Taylor G, Eigenbrod F (2020) Harmonised global datasets of wind and solar farm locations and power. Sci Data 7:130. https://doi.org/10.1038/s41597-020-0469-8
22. Zhang S, Zhao W, Zhu B, Song CYX, Jiang H, Fang J, Ciais P, Xuan N, Gentine P, Davis SJ (2025) Zhu Liu, and Guo Yu Qiu. A near-real time daily European power consumption and carbon intensity dataset (ECON-PowerCI). Sci Data 12:1693. https://doi.org/10.1038/s41597-025-05978-7
23. Zheng X, Xu N, Trinh L, Wu D, Huang T, Sivaranjani S, Liu Y, Xie L (2022) A multi-scale time-series dataset with benchmark for machine learning in decarbonized energy grids. Sci Data 9:359. https://doi.org/10.1038/s41597-022-01455-7
24. Barbar M, Mallapragada DS, Alsup M, Stoner R (2021) Scenarios of future Indian electricity demand accounting for space cooling and electric vehicle adoption. Sci Data 8:178. https://doi.org/10.1038/s41597-021-00951-6
25. Tong K, Nagpure AS, Ramaswami A (2021) All urban areas' energy use data across 640 districts in India for the year 2011. Sci Data 8:104. https://doi.org/10.1038/s41597-021-00853-7

26. Rohit Kumar RE-FUSED: State-day panel of India's power system from official data, with forecasting, forecast correction and scheduling code, results and logs. Version 1.1.0, Dataset, Zenodo, 2026. https://doi.org/10.5281/zenodo.23685972

Figures

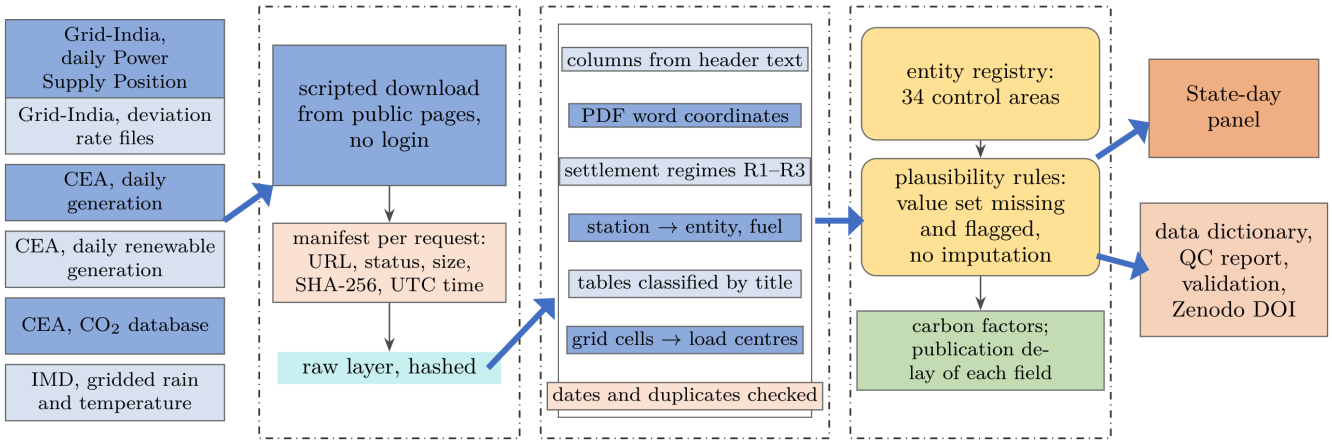


Figure 1

The workflow for constructing the State-day panel from official public reports. Source files are downloaded with request-level manifests and SHA-256 hashes, parsed using source-specific procedures, harmonised across 34 State control areas, and checked using plausibility rules. The final release includes the panel, data dictionary, quality-control report, validation results, and Zenodo DOI.

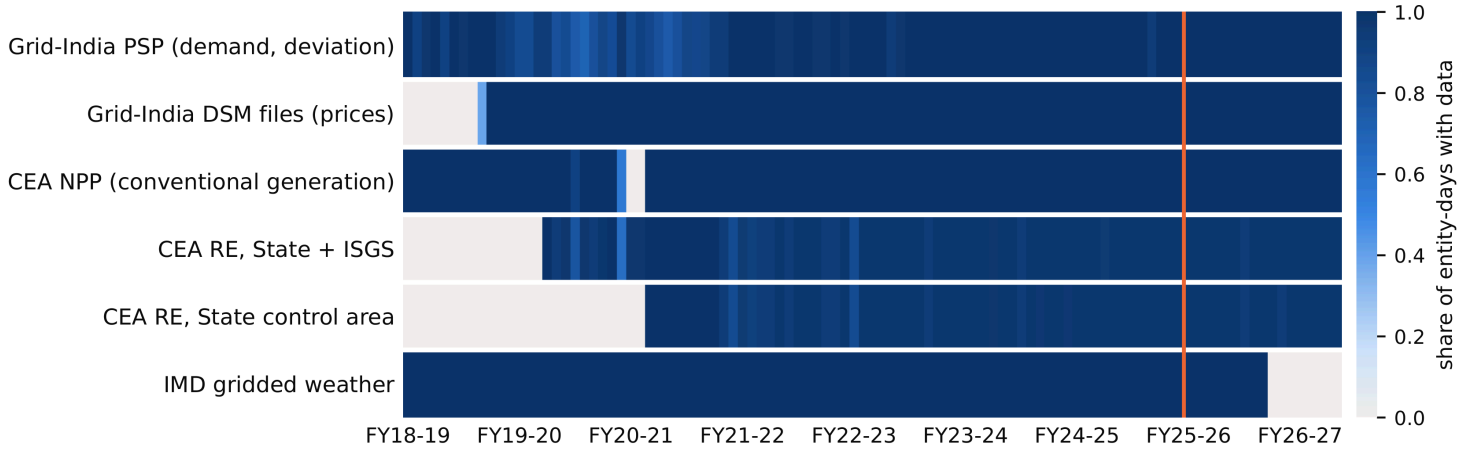


Figure 2

Share of entity-days with data, by source and month. The orange vertical line marks 1 April 2025, after which the period is reserved for evaluation in companion studies. The figure illustrates source coverage and data availability over time.

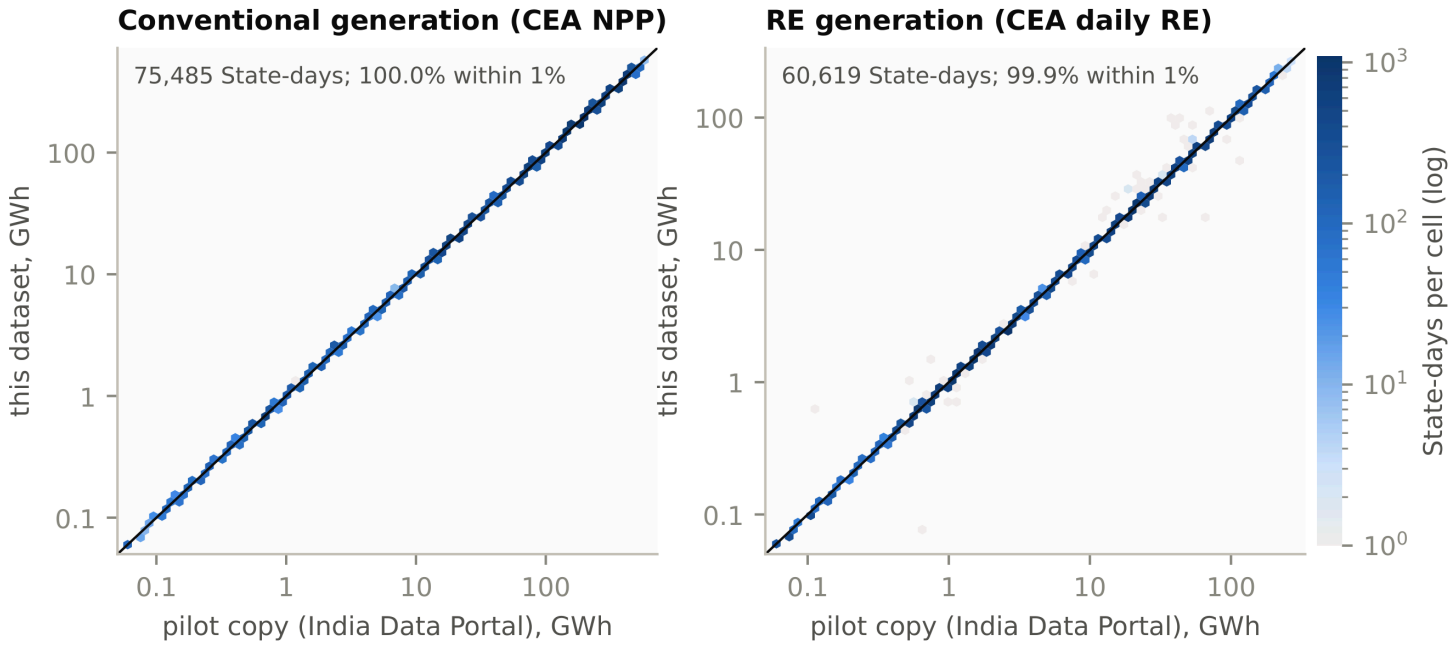


Figure 3

State-day conventional and renewable generation in the panel compared with an independent copy of the corresponding Central Electricity Authority reports from the India Data Portal, through 31 March 2025. Both axes use logarithmic scales, and the diagonal line indicates equality. Conventional generation agrees within 1% for 100.0% of 75,485 State-days, while renewable generation agrees within 1% for 99.9% of 60,619 State-days.

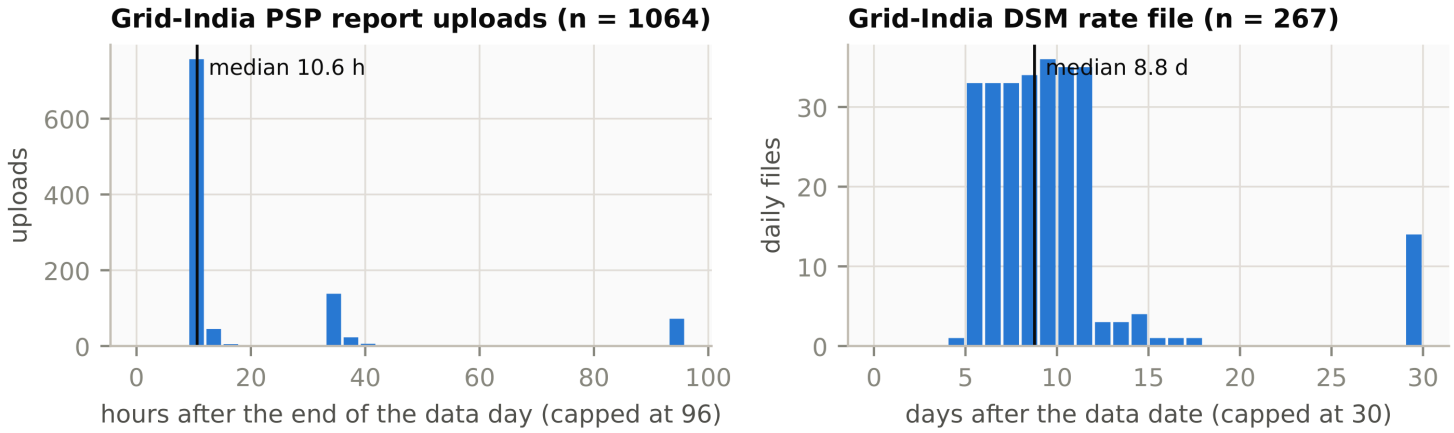


Figure 4

Publication delays measured from the publishers' file metadata. The left panel shows the upload delays for Grid-India Power Supply Position reports (1,064 uploads; median 10.6 hours after the end of the data day). The right panel shows delays for daily deviation settlement rate files (267 files; median 8.8 days after the data date). The distributions illustrate why forecasting studies must account for data availability delays.