https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Open Access Earth System Science Data Discussions

Open-access energy demand data for South and Southeast Asia

Oliver G Coombes[1,2], Kieran M R Hunt[2,3], and Hannah C Bloomfield[4]

[1]Department of Earth Science and Engineering, Imperial College London, UK
[2]Department of Meteorology, University of Reading, Reading, UK
[3]National Centre for Atmospheric Science, University of Reading, Reading, UK
[4]Department of Civil and Geospatial Engineering, School of Engineering, Newcastle University, Newcastle upon Tyne, UK

Correspondence: Kieran M R Hunt (k.m.r.hunt@reading.ac.uk)

Abstract. Open-access electricity demand data are essential for meteorology–energy and climate–energy research, forecasting, and resilience planning. Yet in South and Southeast Asia (SASEA), such records are fragmented across sources, reported in inconsistent formats, and often difficult to find or access. This is a serious limitation in a region where electricity use and system stress are sensitive to monsoon variability, humid and dry heat, and other natural hazards.

In this paper, we present and describe a harmonised electricity demand dataset for twelve SASEA countries (Bangladesh, Bhutan, India, Malaysia, Myanmar, Nepal, Oman, Philippines, Singapore, Sri Lanka, Taiwan, and Thailand) at daily national resolution, spanning 2013–2025 with country-dependent coverage. We compiled raw data from national utilities, regulators, and international providers using reproducible retrieval workflows (e.g., APIs and automated scraping of PDFs/XLS/web portals). All records were standardised to megawatt-hours (MWh), and aligned to local-calendar daily totals (i.e., midnight-to-midnight in local standard time).

To support and encourage transparent downstream use, we also provide the raw extracted series, alongside harmonised daily aggregates, metadata, and our processing and scraping scripts. We also publish diagnostics quantifying completeness, gaps, and outliers flagged using a range of statistical methods. Independent validation against Ember monthly electricity demand shows strong agreement in temporal variability for most countries.

Our open dataset will enable regional and cross-country analysis of demand seasonality, growth and variability; evaluation of weather–demand sensitivity using reanalysis or forecasts; and event-based studies of disruption and recovery during extremes. We finish with a short case study application of our dataset and discussion on how it should and should not be used.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

1 Introduction

Reliable and open-access electricity demand data are important for understanding electricity consumption, monitoring long-term trends, and supporting energy system planning. Monitoring long-term trends is useful for planners who want to separate secular growth from weather-driven variability and seasonality, identify the impact of policy (e.g., during Covid), and quantify the effect of emerging drivers (e.g., transport electrification). Beyond national planning, harmonised datasets allow for the appraisal of prospective cross-border interconnections. Across South and Southeast Asia (SASEA), many studies have shown strong links between temperature and electricity demand across different spatial and temporal scales (e.g. Gupta, 2012; Hung and Huang, 2015; Ang et al., 2017; Joshi et al., 2022a; Hunt and Bloomfield, 2025a; Islam et al., 2025). In SASEA, however, consistent and standardised reporting is often lacking. Where data are available, they may be fragmented across formats, aggregated inconsistently, or recorded in different time conventions, making integration and intercomparison of datasets across countries challenging. These gaps, which can range in length from hours to years, are particularly important in regions influenced by the Asian monsoon, where strong seasonality in temperature and humidity lead to highly variable electricity use. High temperatures and extreme weather have already been linked to grid strain and outages in the region (Zhong and Charturvedi, 2014; Hunt and Bloomfield, 2025b).

Electricity demand data have been used for studies across SASEA, including Bangladesh (Uddin et al., 2019; Miraz et al., 2021; Hussain et al., 2025; Islam et al., 2025), Bhutan (Chophel et al., 2025), Cambodia (Chreng et al., 2022), India (Debnath et al., 2022; Chandra et al., 2024; Gulati et al., 2021; Hunt and Bloomfield, 2025a), Indonesia (Akil et al., 2023; Atma et al., 2025), Malaysia (Tai et al., 2021; Zaim et al., 2023), Nepal (Chapagain et al., 2021; Rajbhandari et al., 2021), Oman (Al-Abri and Okedu, 2023), Pakistan (Ali et al., 2013; Aziz et al., 2024; Shah et al., 2020; Ullah et al., 2023), the Philippines (Albiento et al., 2023; Parreño, 2022; Santos, 2021; Torculas et al., 2023), Singapore (Dhar et al., 2025; Liang, 2021), Sri Lanka (De Silva and Samaliarachchi, 2013; Priyadarshana et al., 2021; Shiwakoti et al., 2025), Thailand (Chapagain et al., 2020; Parkpoom and Harrison, 2008), and Vietnam (Phu, 2021). These use national or regional demand data at a range of aggregations (yearly down to sub-hourly), with the majority using those data to train statistical or machine learning models to predict demand across various timescales. The remainder typically focus on identifying drivers of demand (e.g., temperature, Covid-19). Many of these studies are from countries that do not openly publish demand data, or make it difficult to access through APIs or reports.

To address this, we have compiled a harmonised, open-access dataset of daily electricity demand for twelve countries in SASEA: Bangladesh, Bhutan, India, Malaysia, Myanmar, Nepal, Oman, Philippines, Singapore, Sri Lanka, Taiwan, and Thailand. Countries were selected on the basis of data availability (to the best of our knowledge) and their shared exposure to strong intraseasonal variability. Together, they capture a wide range of energy systems, access to energy, and grid structures, making the dataset relevant for both regional comparison and cross-country analysis.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Open Access Earth System Science Data Discussions

The dataset is provided at daily, country-level resolution spanning 2013 to 2025, where availability varies per country. While some national sources publish data at sub-daily or sub-national levels, we focus on daily, national-level demand. All records are standardised into consistent units, dates, and ISO country codes, with accompanying metadata and reproducible processing scripts. To our knowledge, this is the first openly accessible, daily-resolution dataset spanning multiple monsoonal countries in SASEA.

This work complements and extends existing resources. The International Energy Agency (IEA) holds records of national daily demand data[1] which can be accessed through an API, but with limited coverage that excludes several countries in this region, such as Oman, Bhutan, Nepal, and Myanmar. Ember[2], a non-governmental energy research organisation provides harmonised electricity data at monthly resolution, which is too coarse for most applications. Our dataset fills these gaps by offering daily, openly available, and consistently processed demand records with broader coverage. In several cases, the raw data we collected also provide higher temporal resolution than those available through IEA or Ember. Where higher-than-daily resolution is available, we include it in our published dataset.

This resource can support a range of users. Climate scientists can link daily demand to weather or reanalysis data to study variability and extremes (Staffell and Pfenninger, 2018; Bloomfield et al., 2022). Energy planners can use it to track demand growth, benchmark infrastructure needs, and assess resilience (Steinbuks et al., 2017; Joshi et al., 2022b). Economists and policy analysts can evaluate market trends and the impacts of policy interventions using electricity demand as a proxy for economic activity (Fezzi and Fanghella, 2020; Wang et al., 2021). Development practitioners and international agencies can use demand and reliability data to monitor energy access and service quality, including disruptions and outages, and to support policy design (IEA et al., 2024; Odarno and Kinuthia, 2018; Dunn et al., 2019). By improving access to consistent daily records, this dataset will support both scientific and practical applications in energy planning and management for a region where such data are typically scarce.

This paper presents an open-access, high-resolution and daily energy demand datasets for twelve countries across South and Southeast Asia. The datasets are derived from national utility reports, scraped from official websites or APIs where possible, and cleaned into a consistent, user-friendly format. Regional calendar systems have been standardised to Gregorian dates, and the dataset is align with gridded weather reanalysis products such as ERA5 from Copernicus Climate Change Service (2023) for integration into climate-energy models.

2 Data sources

Table 1 summarises the key characteristics of the twelve SASEA countries in their original resolutions and reporting conventions.

[1] https://www.iea.org/data-and-statistics/data-tools/real-time-electricity-tracker
[2] https://ember-energy.org/data/electricity-data-explorer/

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

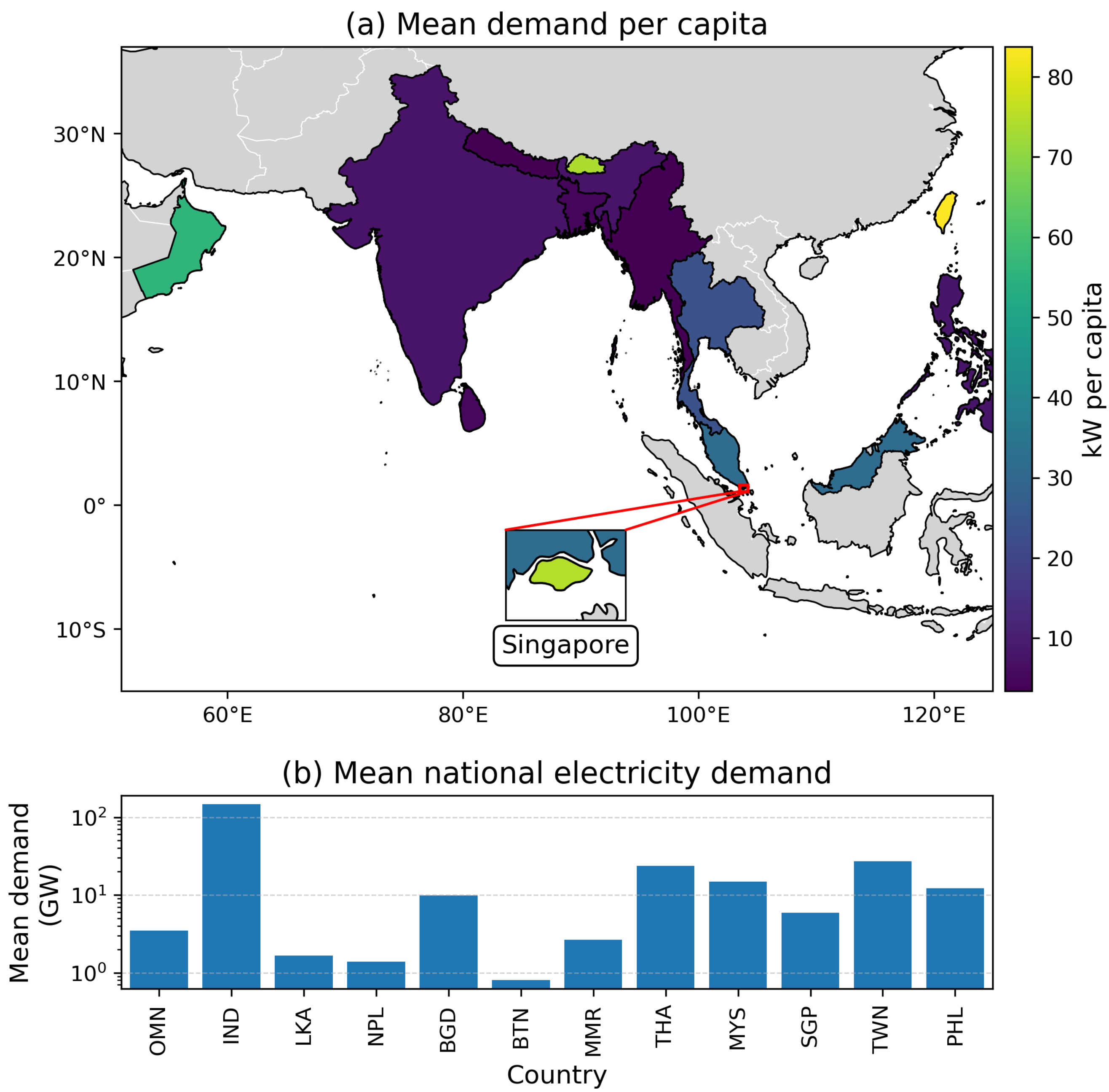

Figure 1. Average country-level electricity demand. (a) Map of mean electricity demand per capita (kW per capita) for the twelve included SASEA countries, computed from the harmonised daily aggregated dataset (`daily_agg_overall_demand.csv`) and median population estimates over the corresponding period. Daily energy totals are converted to mean power by dividing by 24 h prior to population normalisation. (b) Mean national electricity demand (GW; logarithmic scale), shown for the same set of countries and ordered approximately west-to-east.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Table 1. Summary of raw electricity-demand data sources and reporting conventions by country. MW = megawatts, MWh = megawatt-hours, MU = million units (GWh).

Country	Source	Link	Access	Units	Native frequency (TZ)
Bangladesh	Bangladesh Power Development Board (BPDB)	https://www.bpdb.gov.bd	Web scrape	MW	Daily (UTC+06:00)
Bhutan	Bhutan Power System Operator (BPSO)	https://www.bpso.bt	API	MWh	Daily (UTC+06:00)
India	Grid India	https://grid-india.in/	PDF scrape	MU	Daily (UTC+05:30)
Malaysia	Grid System Operator (via IEA)	https://www.gso.org.my/	IEA API	MW	Daily (UTC+08:00)
Myanmar	Ministry of Electric Power (MOEP)	https://moep.gov.mm	Web scrape	MWh	Daily (UTC+06:30)
Nepal	Nepal Electricity Authority (NEA)	https://www.nea.org.np	PDF scrape	MWh	Daily (UTC+05:45)
Oman	Oman Power and Water Procurement	https://www.omanpwp.om	XLS scrape	MWh	Hourly (UTC+04:00)
Philippines	IEMOP (via IEA)	https://www.iemop.ph/	IEA API	MW	Daily (UTC+08:00)
Singapore	Energy Market Authority (EMA)	https://www.ema.gov.sg	XLS scrape	MW	30-min (UTC+08:00)
Sri Lanka	Public Utilities Commission of Sri Lanka (PUCSL)	https://www.pucsl.gov.lk	API	MW	15-min (UTC)[a]
Taiwan	International Energy Agency (IEA)	https://www.iea.org/data-and-statistics	IEA API	MW	Daily (UTC+08:00)
Thailand	EGAT (via IEA)	https://www.egat.co.th	IEA API	MW	Daily (UTC+07:00)

[a]Sri Lanka is reported in UTC in the raw source; values are converted to local standard time before aggregation in the harmonised daily dataset.

Although the core characteristics are summarised in Table 1, several country-specific considerations are important. Reporting formats vary widely: some utilities provide structured, machine-readable data (e.g., Bhutan via a public archive, Sri Lanka via station-level records, i.e., demand reported for each power station by ID and timestamp), while others publish daily reports in PDF format (India, Nepal). Nepal reports inconsistently between the Bikram Sambat and Gregorian calendars. Myanmar's dataset contains significant gaps, which are documented in the accompanying metadata. Finally, a subset of countries (Oman, Singapore, Sri Lanka, India, Bangladesh) publish sub-daily and/or sub-national data. These raw records were preserved in their original resolution, but in the harmonised dataset they

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

were aggregated to daily, country-level values to ensure consistency and comparability. Sub-daily data are retained for completeness, but daily resolution was chosen as the best way to provide analysis across countries.

3 Processing & Harmonisation

3.1 Unit standardisation

All data were converted to megawatt-hours (MWh), as it is an appropriate unit for national grids. For datasets originally in megawatts (MW), values were multiplied by the duration in hours represented by each time step (e.g., 24 for daily, 0.5 for 30-minute intervals). Datasets reported in million-units (MU), which are equivalent to gigawatt-hours (GWh), were converted by multiplying by 1000 to obtain MWh. Generation was taken as a proxy for demand because grid-scale electricity storage remains minimal across SASEA (Chernyakhovskiy et al., 2021). Where sub-daily or sub-national records were available, values were aggregated to daily, country-level totals. For example, Sri Lanka reported demand at the power station level, which required station-level summation before daily aggregation.

3.2 Time alignment and special cases

All timestamps were first converted to local standard time prior to aggregation, ensuring that daily totals – midnight to midnight for the local timezone – reflect local energy-use cycles. For Sri Lanka, originally reported in UTC, this required a shift to UTC+5:30; without this adjustment, daily totals would have been split across local calendar days.

Nepal publishes demand data in both Gregorian and Bikram Sambat calendars. Because Bikram Sambat is the national standard, Gregorian months were occasionally recorded incorrectly (e.g. 31/08 followed by 01/08 instead of 01/09). These discrepancies were corrected, with all values reported in Gregorian dates in the harmonised dataset and the mapping documented in the metadata. Myanmar's dataset contains significant gaps, which are fully documented. Note that users should not mix local-day and UTC-day series when comparing across countries or with reanalyses.

3.3 Output format

The harmonised dataset contains three fields (Table 2): an ISO 8601 date, an ISO 3166-1 alpha-3 country code, and a floating-point daily electricity demand value in MWh. Data coverage varies by country, full ranges are documented in the metadata.

4 Missing data and quality control

No interpolation or imputation was applied. Missing entries are assumed to reflect genuine reporting gaps (e.g. outages, missing reports) and are documented in the metadata.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Table 2. Column names and formatting for the harmonised daily dataset.

Column Header	**Format**
`date`	ISO 8601 format: YYYY-MM-DD
`country`	ISO 3166-1 alpha-3 code (e.g., BGD for Bangladesh)
`demand_MWh`	Floating-point daily total electricity demand in MWh

To ensure data integrity, several quality control steps were implemented, including dataset completeness assessment, outlier detection, and consistency checks across countries.

4.1 Outlier detection

Outliers were identified on a gap-filled dataset, where gaps of three days or fewer were linearly interpolated. This temporary gap-filling was applied only to stabilise rolling calculations; the main harmonised dataset remains unfilled. Gap-filled days are recorded in the accompanying outlier CSV, but the main dataset remains unfilled. Outliers were flagged using three complementary methods commonly applied in time-series anomaly detection (Jamshidi et al., 2022; Leys et al., 2013; Shao et al., 2020):

1. **30-day rolling mean z-score**: Daily values deviating more than $\pm$ 0 standard deviations from the 30-day rolling mean were flagged. The rolling window accounts for long-term growth in demand.
2. **14-day rolling median magnitude ratio**: Values with ratios greater than 5.0 or less than 0.2 relative to the 14-day rolling median were flagged. Using the median provides robustness against skewness or extreme values.
3. **Rolling difference check:** Daily changes exceeding five times the average absolute difference within a 7-day window were flagged.

All flagged outliers are retained in the dataset, as these points may represent meaningful events such as outages, national holidays, or grid failures, which are important for forecasting and understanding demand variability. An accompanying CSV file contains all calculated scores and flags for transparency, so the end user can decide whether to filter or highlight them. See Table 3 for more information. Overlaps between methods occur and are recorded in the outlier dataset. Note that all of these thresholds are heuristic and intended for flagging, not automatic exclusion

Outlier counts vary considerably across countries. India (45) and Bangladesh (32) dominate, consistent with their longer records and relatively high outlier frequency (11.5 and 9.8 per 1000 days respectively). Bhutan has a relatively high count (13), due to a high outlier frequency (24.2 per 1000 days) driven mainly by extreme magnitude ratios. Countries like Sri Lanka, Singapore, and Taiwan had no flagged outliers, reflecting either more consistent reporting,

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Table 3. Counts of different outlier types per country. Outliers per 1000 days are normalised by the number of non-zero (reported) days in each country record.

Country	z-Score	Diff.	Mag. ratio	Total	Outliers / 1000 days	Start	End
Bangladesh	31	8	3	32	9.8	2015-12-17	2025-04-08
Bhutan	2	2	11	13	24.2	2023-04-01	2025-05-14
India	37	38	0	45	11.5	2013-01-01	2024-04-28
Malaysia	0	0	0	0	0.0	2017-07-24	2025-08-17
Myanmar	0	1	0	1	1.7	2018-03-30	2024-07-02
Nepal	3	0	0	3	3.7	2022-10-10	2025-01-11
Oman	2	1	0	3	0.8	2013-01-01	2022-09-30
Philippines	5	1	1	5	3.3	2020-10-09	2025-08-16
Singapore	0	0	0	0	0.0	2014-01-06	2025-04-21
Sri Lanka	0	0	0	0	0.0	2023-01-01	2025-04-08
Taiwan	0	0	0	0	0.0	2021-01-01	2025-07-31
Thailand	1	0	0	1	0.8	2021-07-01	2025-07-26

smaller variability, or pre-existing quality control. Low counts in Myanmar and Oman partly reflect shorter data coverage rather than inherently stable demand reporting, though both also have relatively low outlier frequency (1.7 and 0.8 per 1000 days respectively).

Figure 2 presents an UpSet plot showing overlaps between the three detection methods. Vertical bars indicate the number of days flagged by each combination of methods, providing insight into which methods uniquely capture anomalies versus those consistently identified by multiple approaches.

The rolling z-score method produces the largest number of outlier events (81 total, 41 unique) and has a substantial overlap with the rolling-difference check (51 total, with 36 shared). This is expected because the two methods target related but not identical behaviours, i.e., the z-score identifies days that are extreme relative to a local 30-day baseline (capturing unusually high/low demand for the season or long-term trend), while the difference check is finds more abrupt day-to-day jumps over a shorter window. Their large overlap therefore finds events that are both sharp and large, and thus consistent with genuine disruptions (e.g. outages, recovery periods, major holidays) or abrupt reporting artefact. Conversely, z-score–only events are more likely to identify sustained departures from the baseline (e.g., multi-day anomalies or step changes), whereas difference-only events tend to reflect isolated spikes or dips that may be too brief to strongly affect the rolling mean.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Median-magnitude ratio identified fewer anomalies (15 total, 9 of which are unique). Its lower total may reflect the method's robustness to extreme short-term anomalies (the median being more resistant to outliers than the mean) and therefore likely shows shifts in trend rather than isolated peaks/dips.

In addition to the statistical outlier flags discussed so far, users may wish to identify operational disruption events. These might include multi-day periods of zero demand, or abrupt drop-and-rebound ('V-shaped') events. For example, if we consider V-shaped events where day drops by 0% compared to , and then rebounds by 0% on (returning to within % of the pre-drop level), we count 19 for Bangladesh, 5 for Bhutan, 2 for India, 1 for the Philippines, and 0 elsewhere. This is consistent with Bangladesh's highly vulnerable infrastructure in the face of landfalling tropical cyclones (e.g., Hunt and Bloomfield, 2025b).

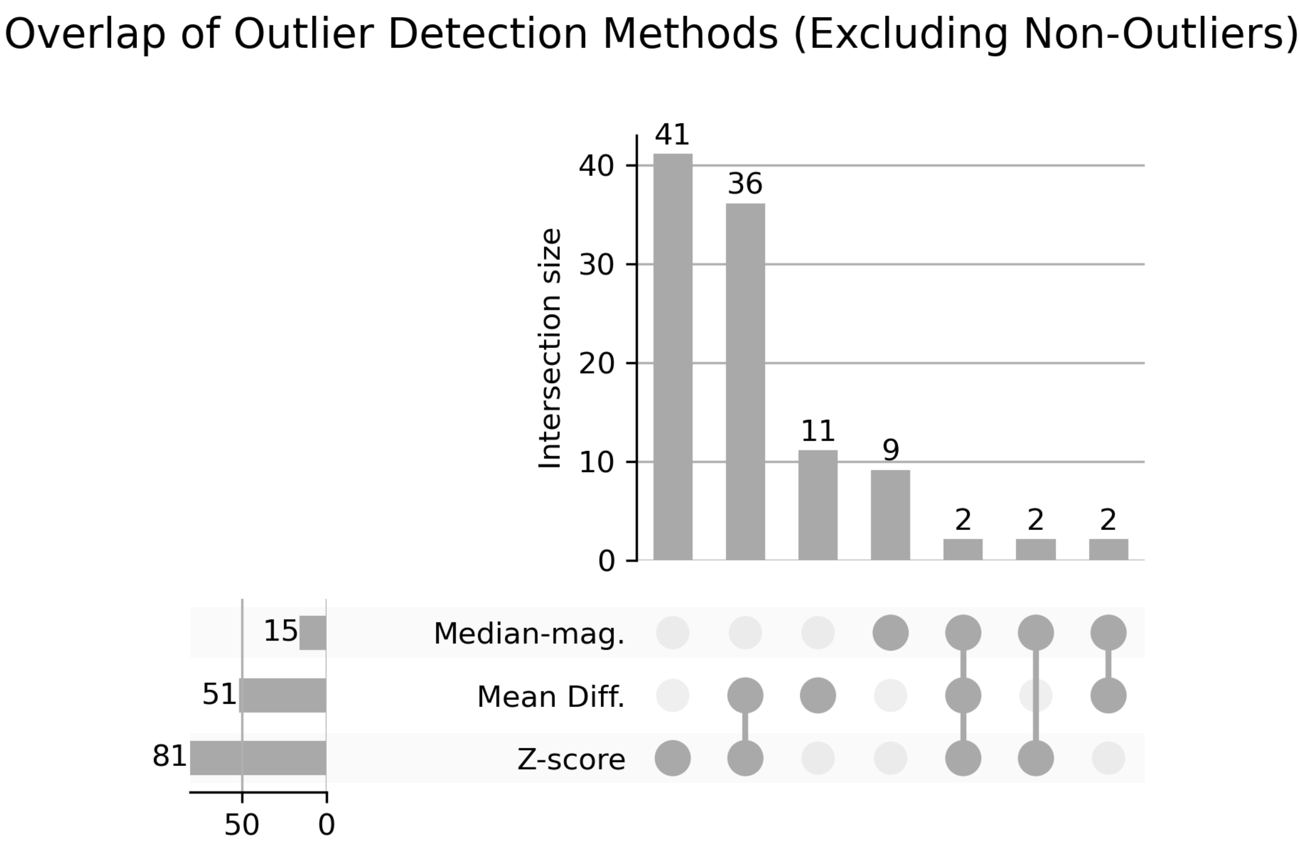

Figure 2. UpSet plot showing the overlap of three outlier detection methods (z-score, mean difference, and median magnitude) across all daily aggregated demand series, excluding days where no method flagged an outlier. The vertical bars indicate the number of days detected as outliers for each combination of methods, highlighting both unique and overlapping detections.

Figure 3 summarises the temporal distribution of flagged outliers with a monthly heatmap.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

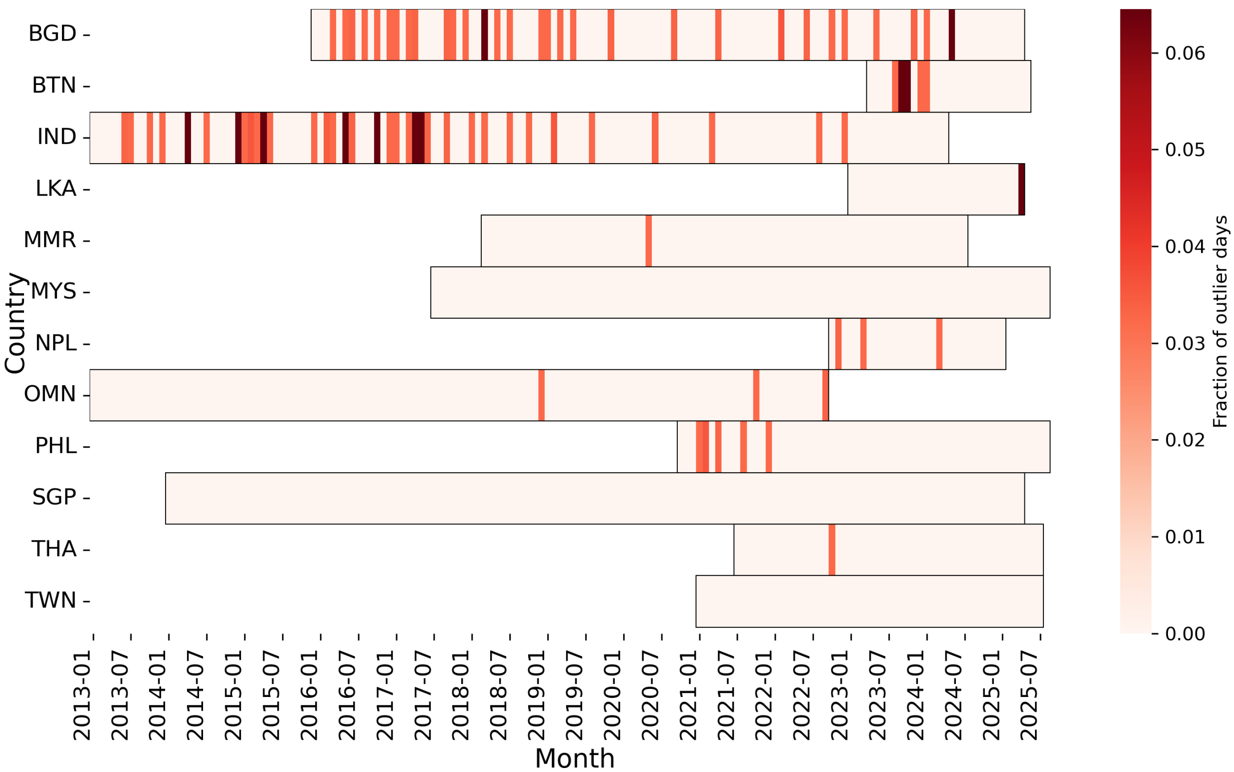

Figure 3. Heatmap showing the fraction of monthly aggregated demand values flagged as outliers for each country, with fractions clipped at 30% to emphasize moderate to high outlier prevalence. Each row represents a country, and each column represents a year-month period. Darker red indicates a higher proportion of outlier days within that month, allowing comparison of outlier patterns across countries and over time.

4.2 Dataset completeness

Completeness was evaluated for each country over the available reporting period, from dataset start date through the full series which varies per country. A day was considered 'complete' if it reported a positive electricity demand value. Days with missing records (gaps in reporting or scraping) or zero demand were considered 'incomplete'. Of course, zero-demand days may represent true outages and are scientifically meaningful; we mark them as incomplete only in the sense that they are not usable for typical demand-modelling workflows without special handling.

While zero-demand entries may reflect real system events such as outages, they were treated as non-informative for demand pattern analysis and therefore reduced dataset completeness in the same way as missing values. Completeness is thus defined as the proportion of complete days out of the total reporting period.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Table 4 summarises dataset start dates, durations, zero-demand days, gap days, and completeness scores. Most countries show high completeness above 86%, with several (Oman, Sri Lanka, Taiwan, Thailand) achieving nearly full coverage. Myanmar is the notable outlier, with only 25.8% completeness, limiting its usefulness for long-term trend analysis. However, it still contributes some insight to demand behaviour and could serve as a source to train foundational models .

Table 4. Dataset completeness per country, including start date, duration, zero-demand days, gap days, and completeness percentages.

Country	Duration (days)	Zero Generation Days	% Zero Gen.	Gap Days	% Completeness
Bangladesh	3303	0	0.00	37	98.88
Bhutan	641	3	0.47	100	83.93
India	4136	0	0.00	236	94.29
Malaysia	2718	0	0.00	5	99.82
Myanmar	2287	0	0.00	1696	25.84
Nepal	814	4	0.49	9	98.40
Oman	3560	0	0.00	0	100.00
Philippines	1545	0	0.00	8	99.48
Singapore	4013	0	3.84	0	96.16
Sri Lanka	731	0	0.00	0	100.00
Taiwan	1461	0	0.00	0	100.00
Thailand	1280	0	0.00	0	100.00

Together, these quality control steps ensure that the datasets are sufficiently reliable for subsequent analysis and forecasting, while explicitly accounting for gaps and anomalies inherent in electricity demand data.

Figure 4 highlights temporal and regional variation in dataset completeness. Most countries show high completeness with occasional one day gaps, typically missing data than zero-demand days. There are a few countries with evident coverage problems: Myanmar shows extended periods of missing data, Singapore displays a large amount of zero-demand days in early 2015, and Bhutan and India have regular missing data throughout their datasets. These patterns complement the overall completeness scores in Table 4.

Ratios of zero-demand and missing data days that lie on national holidays, recognised by the `holidays` package (dr prodigy and contributors, 2025), are identified in Figure 5. It was found that in most countries, these events mostly occur on holidays, other than in Myanmar which experiences outages on non-holiday days.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

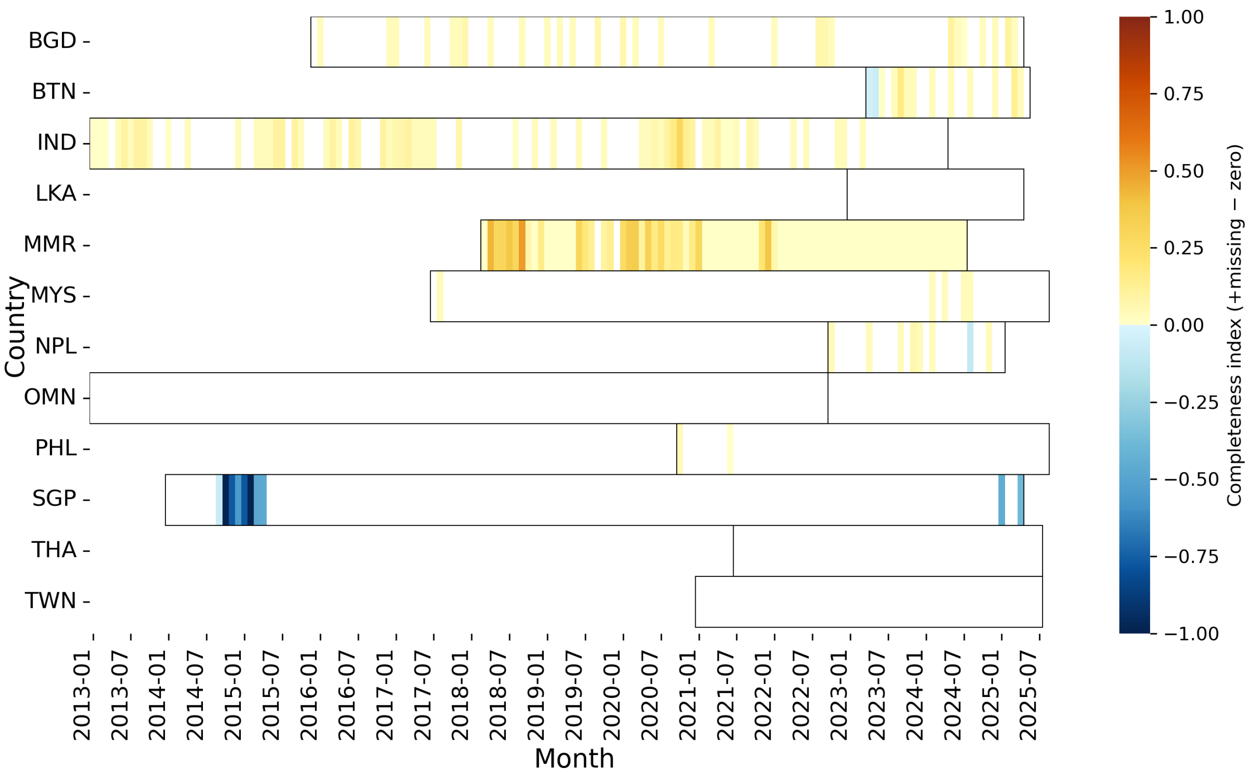

Figure 4. Monthly completeness heatmap for daily aggregated demand across countries, showing the balance between missing and zero-value days. Each row represents a country and each column represents a year-month period. Positive values (red) indicate a higher fraction of missing days, negative values (blue) indicate a higher fraction of zero-value days, and values near zero indicate relatively complete data. Extremes are clipped at ±50% to highlight months with substantial data issues while maintaining visibility of moderate deviations. Boxes indicate the bounds of the dataset period.

4.3 Correlation and statistical significance

Cross-country demand relationships are presented in Figure 6. Bhutan shows little correlation with any other country, which is likely due to its short record, while Myanmar also performs poorly, consistent with a low completeness percentage. Sri Lanka shows lower correlation than expected despite its high coverage. This could be due to the unusual seasonality of its weather (it is affected by a second monsoon, the 'northeast monsoon') in October to December. Equally, it could be due to different tourism seasons, or weekly working patterns.

In contrast, there is high correlation between Southeast Asian countries, particularly Malaysia, Taiwan, the Philippines, and Thailand (correlation of 0.72 between Thailand and the Philippines). This could be due to their proximity (and

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

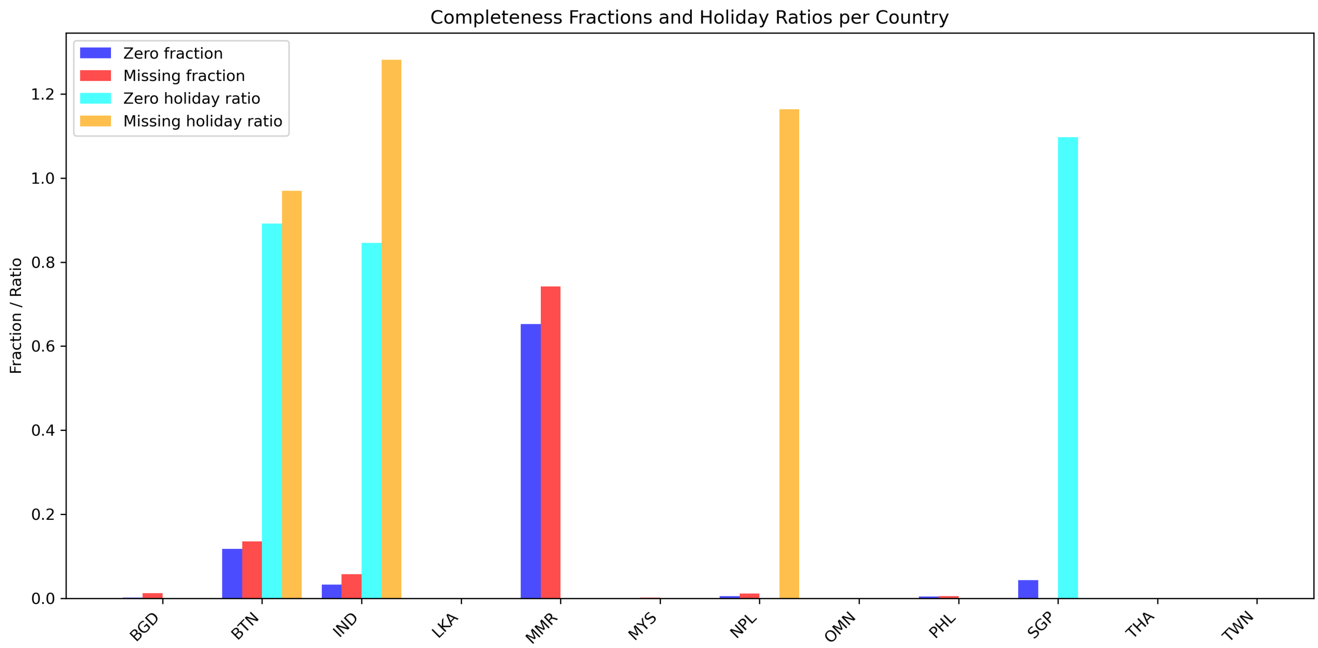

Figure 5. Fraction of zero-demand days and missing-demand days across each country, with ratios of these events occurring on holidays compared to non-holidays. Ratios greater than one indicate that the event is more likely to occur on a holiday.

thus shared large-scale meteorological forcings), but could also be a result of similar calendar structures or long-term trends.

To complement this, Figure 7 shows the average annual seasonal cycle correlation between each country. These cycles show slightly stronger cross-country alignment than daily demand correlation, reflecting reduced dependence on temporal overlap between countries. Southeast Asian countries again form a coherent cluster, with correlations of 0.65–0.7 between the Philippines, Taiwan, and Singapore. Oman also correlates unexpectedly well with Singapore, Taiwan, the Philippines, and Bangladesh. Bhutan has weak correlation with other countries, as do Myanmar and Sri Lanka. Of course, while seasonal-cycle correlations are more robust than daily correlations because they reduce sensitivity to temporal overlap, they are still sensitive to record length (as sampling noise decreases with record length), and so estimates from short series (e.g., Bhutan, Myanmar) should be interpreted cautiously.

4.4 Seasonal and weekly demand structure

Figure 8 summarises two characteristic patterns in the harmonised daily demand series: the mean annual cycle and the mean day-of-week cycle, both expressed as ratios relative to each country's mean demand. The monthly normalisation is performed year-by-year before averaging, reducing sensitivity to long-term demand growth and

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

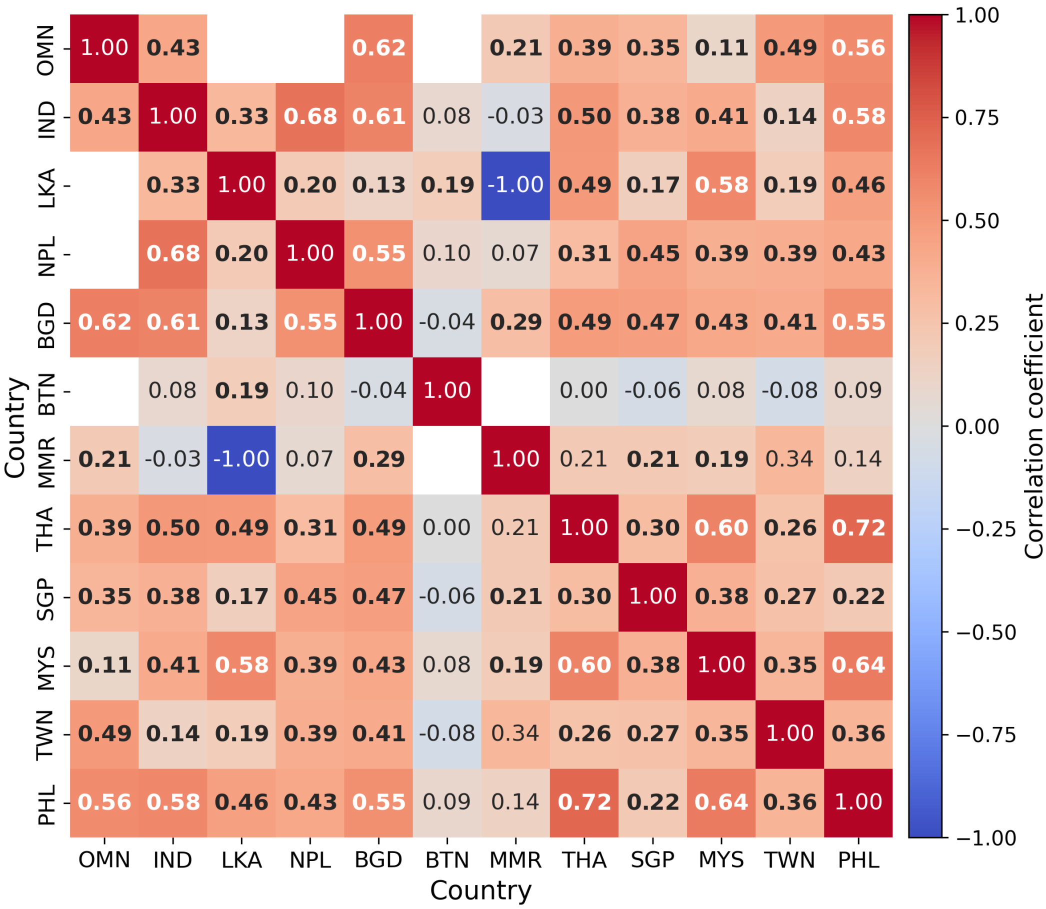

Figure 6. Cross-country electricity demand relationships. Pearson correlation matrix of daily electricity demand, highlighting regional co-movement patterns. Bolded correlations in panel indicate statistical significance at the 1% level.

avoiding distortions from partial years. These diagnostics provide a compact overview of demand seasonality and intra-week structure that can inform model design (e.g., seasonal baselines, calendar regressors) and help identify unusual country-specific behaviour.

Seasonal structure varies strongly across the region. Oman exhibits the most pronounced annual cycle, with summer demand substantially higher than winter demand, consistent with cooling-driven load in the hot season. Several South and Southeast Asian countries show more moderate but still coherent seasonality, with demand typically elevated during warmer and/or more humid parts of the year and suppressed during cooler months. In contrast, Singapore and Malaysia show relatively weak seasonal modulation, consistent with smaller intra-annual temperature ranges and more stable demand profiles.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Open Access
Earth System Science Data
Discussions

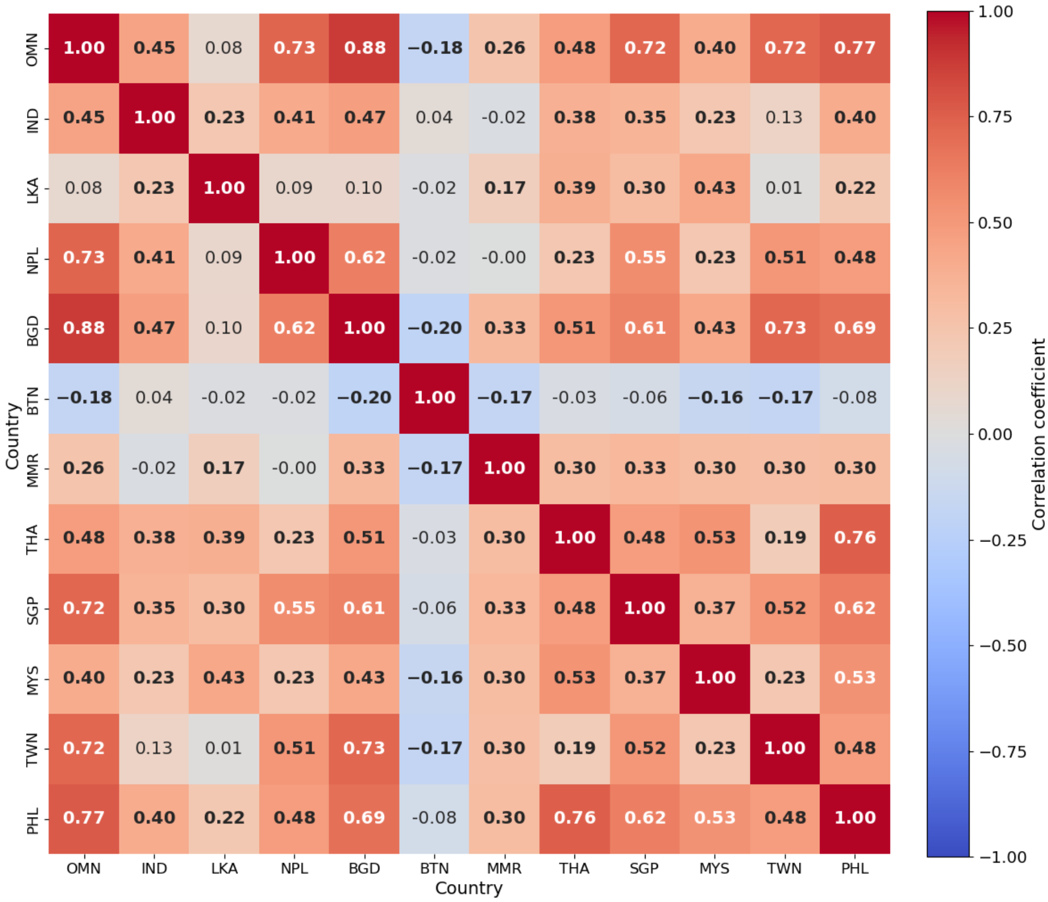

Figure 7. Correlations of the average annual seasonal cycle of electricity demand between SASEA countries. Each cell shows the Pearson correlation between mean seasonal profiles, aggregated over all available years. Bolded correlations indicate statistical significance at the 1% level.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

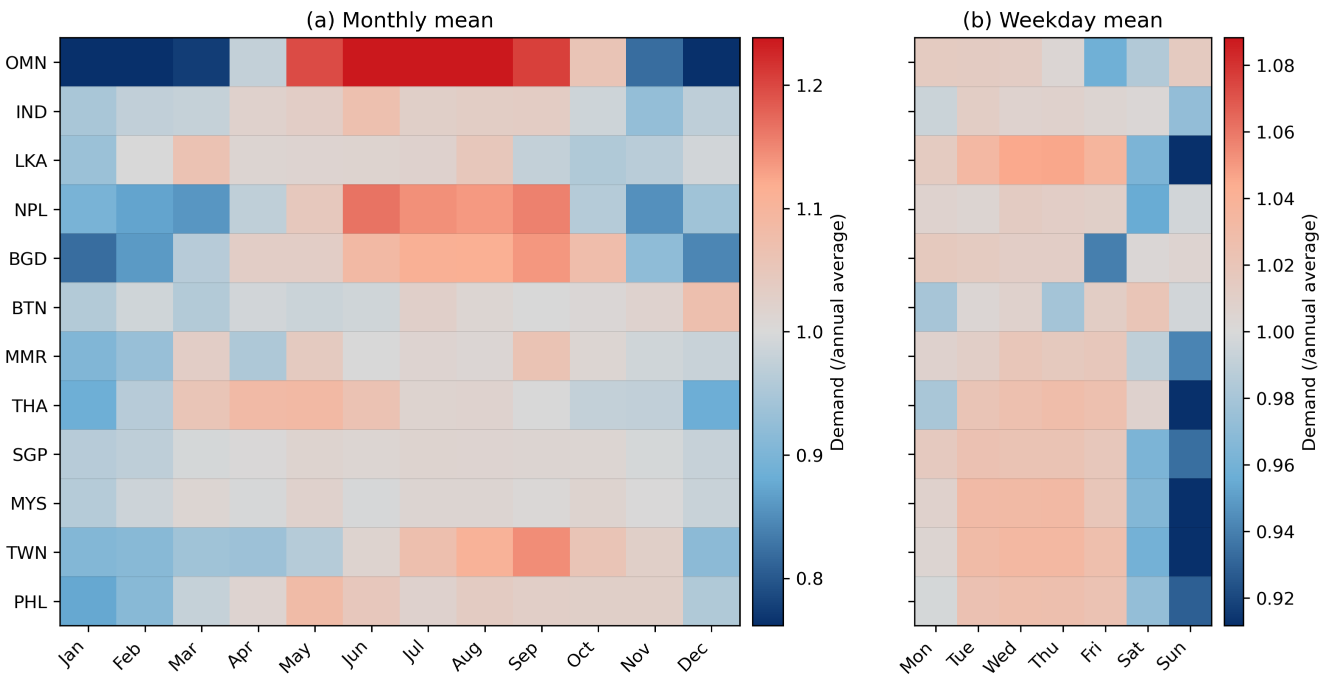

Figure 8. Normalised seasonal and weekly structure of daily electricity demand across the twelve SASEA countries. In panel (a), monthly mean demand is computed for each year, normalised by that year's mean demand, and then averaged across years with at least 8 months of data to reduce sensitivity to long-term growth and incomplete years. Panel (b) shows mean demand by day-of-week, normalised in the same way. Countries are ordered approximately west-to-east. Colours are centred on 1, so that values above 1 indicate above-average demand and values below 1 indicate below-average demand.

The weekday diagnostics show a broadly consistent reduction in demand on weekends across many countries, with lowest values often occurring on Sunday (or Friday in Oman, consistent with differing weekend conventions). The magnitude of the weekly cycle differs across countries: some systems exhibit only weak weekday modulation, while others show clearer reductions on non-working days, indicating a larger contribution from industrial/commercial activity relative to residential load. Together, these patterns demonstrate that the dataset captures physically and socio-economically plausible variability at both seasonal and weekly timescales, supporting its use in weather-sensitive demand studies and operational/strategic energy analyses.

4.5 Validation against independent sources

Mean and median monthly demands were validated against Ember electricity demand data for 10 of 12 countries in the dataset (Nepal and Bhutan are not covered by Ember) (Ember, 2025)., as shown in Figure 9. Note that Ember definitions may vary from ours (e.g., including net imports, or behind-the-meter solar) and so offsets do not strictly imply errors.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Overall, correlation between national government / energy provider sources is strong. Most countries show low mean absolute percentage errors (MAPE), typically less than 10%, indicating that the sourced data broadly aligns with the values reported by an independent non-governmental organisation. This provides confidence in the integrity of the sourced electricity demand data.

Two major exceptions stand out. Malaysia (MAPE = 29.6%) and Taiwan (MAPE = 20.6%) exhibit relatively large errors. In both cases, however, the correlation between the two series remains high with of 0.97 and 0.87 respectively. This suggests that the discrepancies are primarily due to scale difference rather than incorrect demand trends. These deviations could reflect differences in reporting standards or definitions of demand units between government / provider and Ember's methodology. Importantly, all overall trends remain aligned, implying that timing and structure of the demand data is reliably captured.

5 Dataset descriptions

5.1 Raw datasets

The release includes twelve country-level raw demand datasets, each provided as a daily CSV file alongside an accompanying YAML metadata file. CSVs contain electricity demand time series converted from their original reporting formats (e.g. PDF, XLS, API) into a uniform, accessible structure. File names follow the convention `raw_[ISO3]_demand.csv` (e.g. `raw_IND_demand.csv`). Metadata files document variable definitions, units, temporal coverage, access methods, and known issues, with links to the original sources and scripts for automated retrieval.

Table 5 provides a high-level overview of the raw demand files included in the archive. Each dataset is also reported in its original units and temporal resolution before harmonisation, ensuring fidelity to the source material.

Each dataset retains its original reporting conventions:

- **Daily totals**: Bangladesh, Bhutan, India, Malaysia, Myanmar, Nepal, Philippines, Taiwan, Thailand.
- **Sub-daily demand**: Oman (hourly), Singapore (30-min), Sri Lanka (15-min).
- **Units**: as reported by the original source (MW, MWh, or MU).

5.2 Processed datasets

To facilitate cross-country comparison and time-series analysis, we provide a set of harmonised, processed datasets derived from the original files (Coombes et al., 2025). These datasets standardise units, temporal resolution, column names, and quality control indicators.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Open Access Earth System Science Data Discussions

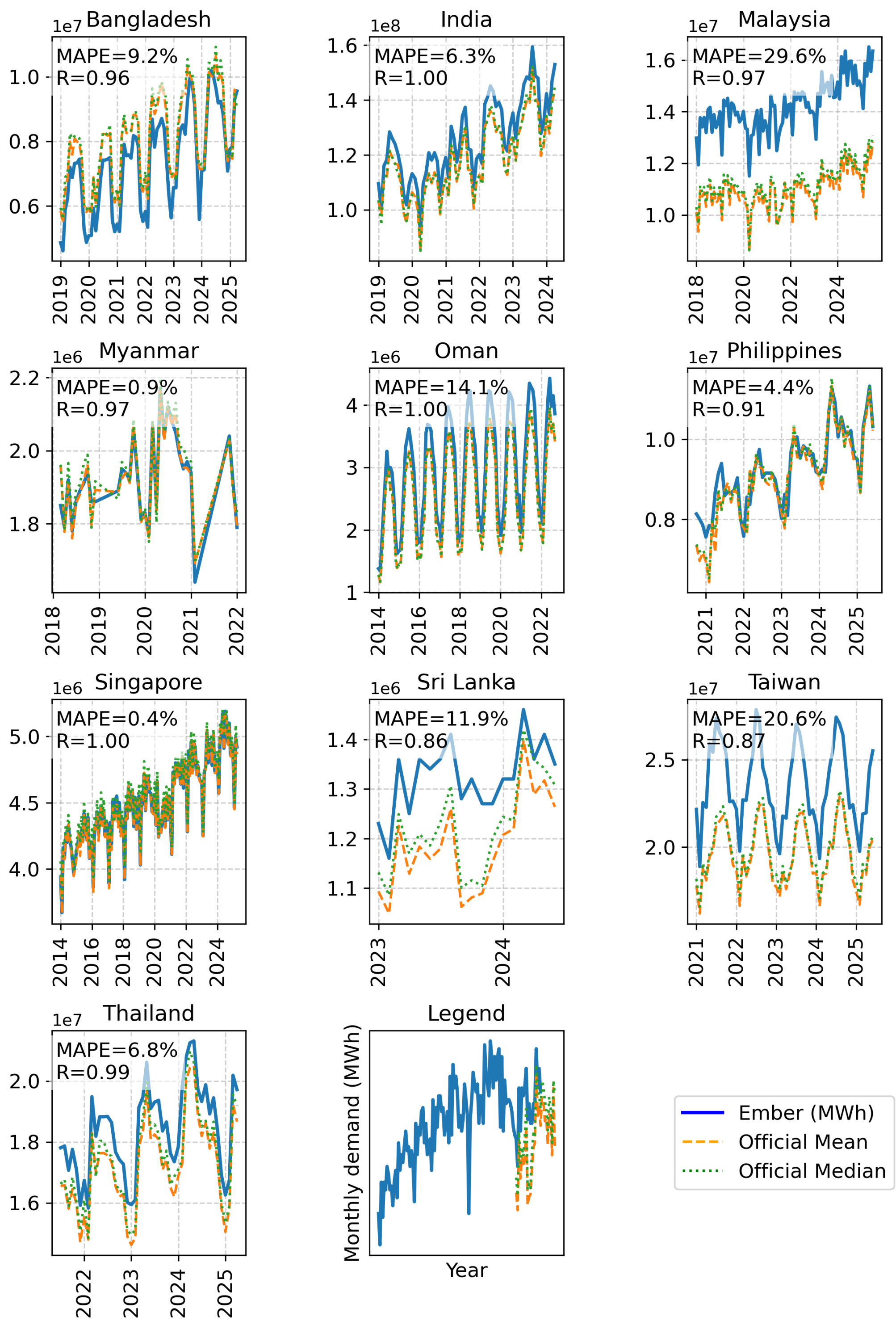

Figure 9. Monthly means (orange) and medians (green) for each country in our dataset plotted alongside mean monthly electricity demand values from Ember (2025) (blue).

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Table 5. Summary of raw demand datasets included in the release. All files are provided as CSVs with associated YAML metadata.

Country	File name	Resolution	Units	Metadata
Bangladesh	raw_BGD_demand.csv	Daily	MW	Yes
Bhutan	raw_BTN_demand.csv	Daily	MWh	Yes
India	raw_IND_demand.csv	Daily	MU	Yes
Malaysia	raw_MYS_demand.csv	Daily	MW	Yes
Myanmar	raw_MMR_demand.csv	Daily	MWh	Yes
Nepal	raw_NPL_demand.csv	Daily	MWh	Yes
Oman	raw_OMN_demand.csv	Hourly	MWh	Yes
Philippines	raw_PHL_demand.csv	Daily	MW	Yes
Singapore	raw_SGP_demand.csv	30-min	MW	Yes
Sri Lanka	raw_LKA_demand.csv	15-min	MW	Yes
Taiwan	raw_TWN_demand.csv	Daily	MW	Yes
Thailand	raw_THA_demand.csv	Daily	MW	Yes

5.2.1 Daily aggregated dataset

The `daily_agg_overall_demand.csv` file provides the main harmonised dataset for cross-country analysis. It aggregates electricity demand to daily totals in megawatt-hours (MWh) for each country, standardising column names and date formats (ISO 8601). Each row represents a single day for a given country, with three core fields: `date`, `country` (ISO 3166-1 alpha-3 code), and `generation_MWh`.

Both the harmonised overall dataset and the per-country aggregated datasets (`daily_agg_[ISO3]_demand.csv`, e.g., `daily_agg_BGD_demand.csv`) extend from the first available record in each source through to the most recent access date. This ensures that users can work with consistent harmonised series, while also accessing the most up-to-date demand records for individual countries. Each file is accompanied by a country-specific YAML metadata file, documenting the original source, units, coverage, and known issues.

5.2.2 Daily aggregated dataset with outlier detection

The file `daily_agg_overall_demand_outliers.csv` extends the harmonised daily aggregated dataset by including diagnostic columns for anomaly detection, missing-value tracking, and gap-filling. These additions enable users to construct quality-controlled time series for modelling and forecasting without needing to re-implement anomaly detection methods (see Section 4.1).

Both the baseline daily dataset and the outlier-extended version are UTF-8 encoded CSV files, with dates in ISO8601 format (`YYYY-MM-DD`) and demand expressed in megawatt-hours (MWh). Table 7 summarises the dataset structure.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Table 6. Structure and coverage of harmonised daily aggregated datasets.

Dataset	Description	Coverage
`daily_agg_overall_demand.csv`	Combined daily demand totals for all countries, harmonised to MWh.	2013–2025
`daily_agg_[ISO3]_demand.csv`	Country-level daily demand totals, one file per country.	Start date varies by source; extends to most recent available date.

Table 7. Column descriptions for daily aggregated datasets. Core columns appear in both datasets; diagnostic columns are specific to the outlier-extended dataset.

Column	Description	Unit / Format
Core columns (in both datasets)		
`date`	Calendar date of observation	YYYY-MM-DD
`country`	ISO 3166-1 alpha-3 country code	Text
`demand`	Daily aggregated electricity demand	MWh
Additional diagnostic columns (only in outlier dataset)		
`is_zero`	Indicator for zero-demand day	Boolean (True/False)
`was_missing`	Indicator for originally missing value	Boolean
`gap_filled`	Indicator if value was gap-filled (imputed)	Boolean
`rolling_zscore`	Z-score of demand relative to rolling window mean (30 days)	Unitless
`diff`	First difference in demand (current minus previous day)	MWh
`rolling_diff_mean`	Rolling mean of daily differences (30 days)	MWh
`rolling_median`	Rolling median demand (30 days)	MWh
`mag_ratio`	Ratio of current demand to rolling median	Unitless
`outlier_z`	Outlier flag based on rolling z-score threshold	Boolean
`outlier_diff`	Outlier flag based on deviation in daily difference	Boolean
`outlier_mag_ratio`	Outlier flag based on magnitude ratio threshold	Boolean
`is_outlier`	Combined indicator (true if flagged by any method)	Boolean

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Open Access Earth System Science Data Discussions

This dataset is provided only in combined form across all twelve countries, unlike the baseline daily datasets which are available both combined and per-country.

In the outlier-extended daily aggregated dataset (`daily_agg_overall_demand_outliers.csv`), intermediate columns such as rolling means, rolling variances, and magnitude ratios are provided primarily for diagnostic purposes, but they can also be used directly in analysis. If the `was_missing` column is `True` and `gap_filled` is `False`, the gap was longer than three days, and all outlier intermediate and flag columns will contain `NaN` values because rolling statistics cannot be computed on missing data.

5.2.3 Statistical datasets

For each country, aggregated electricity demand statistics are provided at monthly, quarterly, and yearly resolutions. File names follow the pattern `stats_[resolution]_[ISO3].csv` (e.g., `stats_monthly_BGD.csv`), with an overall combined dataset as `stats_overall_all.csv`. All files are UTF-8 encoded CSVs with ISO 8601 date conventions (`YYYY-MM-DD`); demand is expressed in megawatt-hours (MWh). Statistics are computed directly from reported daily demand values without interpolation.

Missing or zero-demand days reduce the sample size (`count`) but are otherwise excluded from summary statistics. As a result, some statistics that are left blank, particularly where a given month, quarter, or year has a single, or no data points. For example, Myanmar in 2024 has no calculated standard deviation as there is a single data point.

Table 8 describes the column structure for these datasets.

Table 8. Column descriptions for aggregated statistics datasets.

Column	Description	Unit / Format
`year_month` / `year_quarter` / `year`	Aggregation period identifier	YYYY-MM or YYYY-Q or YYYY
`min`, `max`	Minimum / maximum daily demand in period	MWh
`mean`, `median`	Mean and median daily demand in period	MWh
`std_dev`	Standard deviation of daily demand	MWh
`total`	Total demand summed across days in period	MWh
`load_factor`	Ratio of mean to maximum demand	Unitless
`p5`, `p25`, `p75`, `p95`	5th, 25th, 75th, 95th percentiles of daily demand	MWh
`count`	Number of valid daily observations in period	Count
`start_date`, `end_date`	Temporal coverage of the dataset	YYYY-MM-DD

The dataset covers electricity demand for twelve countries, aggregated at multiple temporal resolutions (monthly, quarterly, and yearly). This multi-scale structure enables analysis of both long-term trajectories and short-term fluctuations in demand.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Open Access Earth System Science Data Discussions

Table 9. Central tendency and dispersion of daily electricity demand in the harmonised dataset (GWh).

Country	Min	Max	Mean	Median	SD	Start	End
Bangladesh		0		0		2015-12-17	2025-04-08
Bhutan	0 0			0		2023-07-01	2025-05-14
India		0		0		2013-01-01	2024-04-28
Malaysia	0					2017-07-24	2025-08-17
Myanmar		0		0	0	2018-03-30	2024-07-02
Nepal		0				2022-10-10	2025-01-11
Oman						2013-01-01	2022-09-30
Philippines					0	2020-10-09	2025-08-16
Singapore	0	0	0		0	2014-01-06	2025-04-21
Sri Lanka	0			0		2023-01-01	2025-04-08
Taiwan						2021-01-01	2025-07-31
Thailand	0					2021-07-01	2025-07-26

Country	5th	25th	75th	95th
Bangladesh				
Bhutan		0	0	
India		0 0		
Malaysia	0	0		0
Myanmar		0		
Nepal	0			0
Oman			0	
Philippines		0		0
Singapore		0		
Sri Lanka				
Taiwan		0		
Thailand			0	

Table 10. Electricity-demand percentiles of the daily distribution in the harmonised dataset (GWh).

295 A key diagnostic metric applied is the load factor, defined as the ratio of average demand to peak demand within a given time window. Higher load factors typically indicate more efficient and stable utilisation of generation capacity, whereas lower values suggest higher peak stress relative to baseline consumption.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Tables 9 and 10 summarise electricity demand statistics for each country. Demand levels span several orders of magnitude, from small systems such as Bhutan averaging 19,500 MWh, to large systems such as India which averages 3.5 million MWh.

High-demand systems include Taiwan and Thailand with average demands of around 646,000 MWh and 565,000 MWh respectively. Malaysia and the Philippines are slightly lower at 356,000 MWh and 294,000 MWh. These countries show relatively high stability, with standard deviation below 70,000 MWh in Taiwan and Thailand (less than 12% of their means), and under 40,000 MWh in Malaysia and the Philippines (less than 15% of their means).

Mid-sized systems such as Oman and Singapore average around 83,500 MWh and 148,000 MWh respectively. Their standard deviations of 25,000 MWh and 11,000 MWh indicate a third of Oman's mean demand and less than 10% of Singapore's.

Smaller systems such a Sri Lanka, Nepal, and Myanmar show smaller mean demands at 39,800 MWh, 33,200 MWh, and 63,400 MWh, around half that of Oman and Singapore. Their standard deviations of around 4500 MWh (<12%), 5300 MWh(<17%), and 4900 MWh(<8%) respectively point to relative variances of less than 12% for Sri Lanka and Myanmar and <18% for Nepal.

Bhutan stands out as the smallest system, with a mean of just under 20,000 MWh and standard deviation of 3500 MWh, less than 18% of the mean.

Overall, these values suggest that absolute variance scales with grid size, as expected; however, relative variance does not. Larger grids generally show lower proportional variability than smaller grids. Variance within mid- to small-sized grids is more heterogeneous, countries such as Myanmar and Singapore indicate much lower variability than Oman, Nepal, and Bhutan. This highlights that grid size alone is not indicative of grid stability.

Figure 10 highlights overall statistics per country (Table 9). It emphasises overall trends in demand, with most countries showing an increase.

5.2.4 Trend dataset

A separate summary of long-term demand statistics is provided in `dataset_trend_summary.csv`. This dataset is derived from daily demand values (excluding zero-generation and missing days) and retrains outlier observations, which may still reflect meaningful demand events. Rolling windows of 90 days were used to construct mean and variance trends, and augmented Dickey-Fuller (ADF) tests of the aggregated dataset. Linear regression slopes are estimated to characterise gradual changes over time. The 90-day window was chosen to capture subseasonal variability, roughly corresponding to a three month season.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

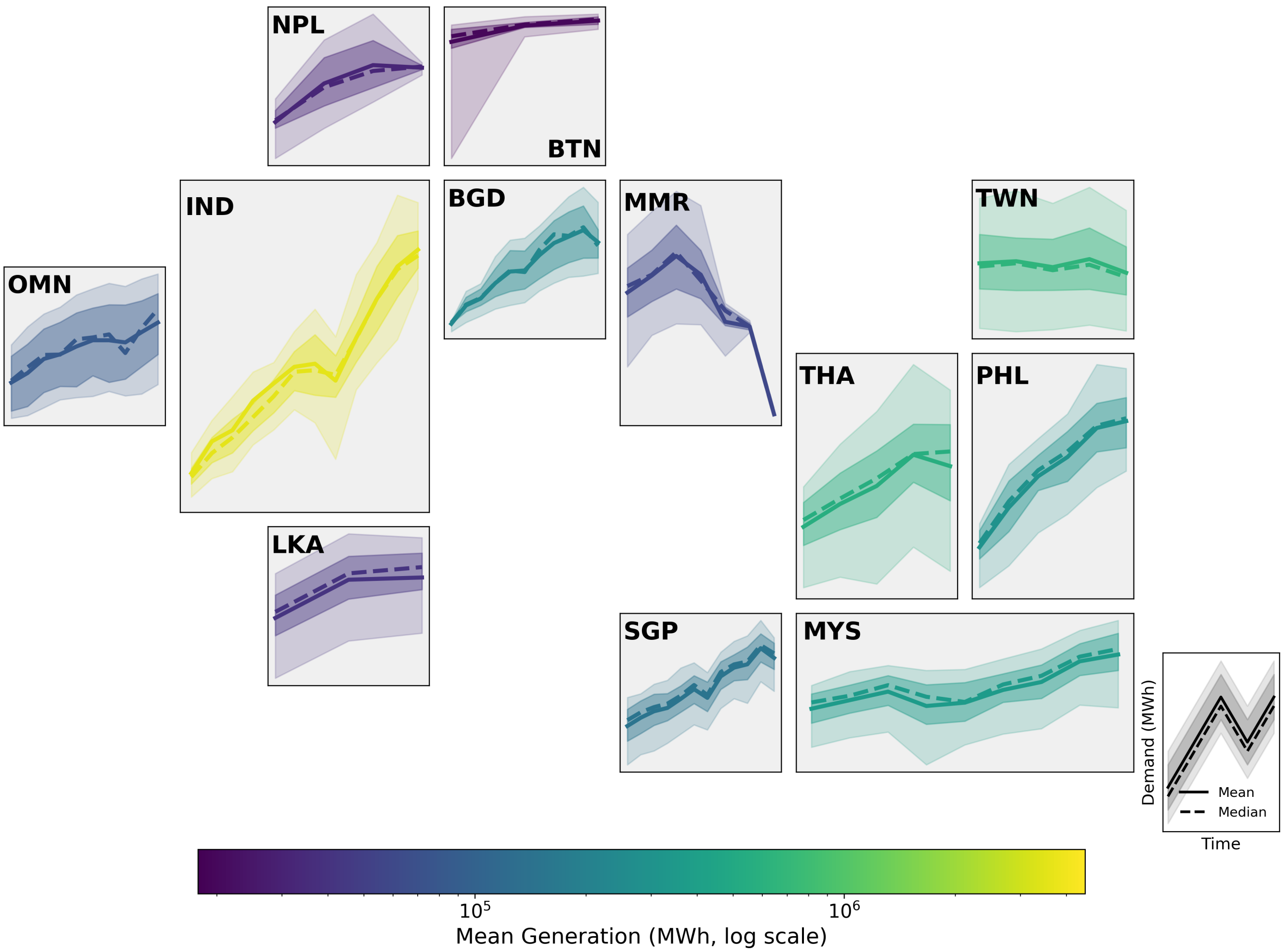

Figure 10. Annual electricity demand distributions for all twelve countries. Each panel shows yearly mean, median, interquartile range, and 5th–95th percentile range of daily electricity demand. Panels are arranged geographically to approximate east-to-west orientation, and colours indicate mean annual demand.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Variance is computed as total variability within each window, without decomposing seasonal cycles or removing trends. Here, a positive slope in the variance trend indicates that variability in demand increases with time. No interpolation of missing days is applied.

Table 11 describes the column structure.

Table 11. Column descriptions for trend summary dataset.

Column	**Description**	**Unit / Format**
`country`	ISO 3166-1 alpha-3 country code	Text
`start_date, end_date`	Coverage of demand series	YYYY-MM-DD
`num_nonzero_days`	Number of days with non-zero demand	Count
`rolling_mean_slope`	Estimated slope of 90-day rolling mean series	MWh/day
`rolling_mean_pvalue`	Significance level of mean trend regression	-value
`rolling_var_slope`	Estimated slope of 90-day rolling variance series	$(\text{MWh})^2$/day
`rolling_var_pvalue`	Significance level of variance trend regression	-value
`adf_statistic`	Augmented Dickey–Fuller test statistic (stationarity test)	Unitless
`adf_pvalue`	Significance level of ADF test	-value

Temporal dynamics were assessed using 90-day rolling diagnostics and Augmented Dickey–Fuller (ADF) tests (Table 12). Rolling-mean trends indicate sustained long-term growth in electricity demand across all SASEA countries, with the strongest absolute increases in the largest systems (e.g., India and Thailand). If we express the trend as a relative change, there is a wide range: mature systems such as Taiwan and Malaysia show fairly small fractional increases (% % year^{-1}), whereas rapidly expanding systems such as Nepal and Bhutan have much faster relative growth (0% % year^{-1}). So while there is region-wide demand expansion, there is also markedly different growth rates depending on national context and baseline system size.

Rolling-variance slopes provide a measure of how shorter-timescale variability evolves. Several countries have increasing variance (e.g., Bangladesh, Malaysia, Nepal, Oman, Taiwan, and Thailand), which is consistent with growing variability alongside rising mean demand. In contrast, declining variance in countries such as India, Bhutan, Myanmar, and Singapore suggests a relative stabilisation of demand variability over the sample period, although interpretation should be treated cautiously because variance trends can also be affected by structural changes in reporting and by non-stationary growth in the mean.

ADF tests were applied to the non-zero daily series to assess whether demand is better characterised as stationary around a stable mean (after accounting for deterministic components) as opposed to long-term drift. Several countries have statistically significant ADF statistics (e.g., Oman and Malaysia), indicating a stationary baseline, implying that models allowing for stochastic trends (or integrating differencing/explicit growth terms) may still be appropriate.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Table 12. Trend statistics of electricity demand per country, computed from non-zero daily demand values. Rolling-mean slopes (MWh day) and rolling-variance slopes were estimated by linear regression applied to 90-day rolling windows through each series; variance slopes are shown scaled by 0 . Rolling-mean trends in % yr are computed as 00 × × slope ‾, where ‾ is the mean daily demand for that country. ADF statistics refer to Augmented Dickey–Fuller tests applied to each non-zero daily demand series. Asterisks denote statistical significance (* 0 0 , ** 0 0 , *** 0 00).

Country	**Non-zero days**	**Rolling mean slope** (MWh day)	**Rolling mean trend** (% yr)	**Rolling variance slope** (0 MWh^2 day)	**ADF stat.**
Bangladesh	3360	43.31***	6.63***	2.87***	-2.43
Bhutan	667	7.13***	13.18***	-0.46***	-3.65**
India	3900	396.39***	4.12***	-527.18***	-3.22*
Malaysia	2942	16.18***	1.65***	0.41***	-3.45**
Myanmar	591	4.81***	2.77***	-0.11***	-3.53**
Nepal	812	13.73***	15.08***	0.13***	-2.36
Oman	3560	11.17***	4.88***	0.26***	-4.01**
Philippines	1765	57.90***	7.08***	-0.22	-2.68
Singapore	3948	7.14***	1.76***	-0.03***	-2.73
Sri Lanka	829	9.15***	8.34***	-0.01	-2.98*
Taiwan	1673	9.62***	0.55***	1.66***	-2.57
Thailand	1487	51.26***	3.30***	10.88***	-2.77

Others show weaker evidence for stationarity (e.g., Bangladesh, the Philippines, Taiwan, and Thailand), suggesting that detrending or regression-based approaches may be needed before further analysis.

Trend (`dataset_trend_summary.csv`) and summary statistics datasets (`stats_monthly/quarterly/yearly_[ISO3].csv`) are derived from the harmonised daily series. Zero-demand days are excluded from trend calculations to prevent distortion of means and variances, and no interpolation is applied. Outlier values are retained because they may be associated with genuine demand events, disruptions, or reporting artefacts (and we do not know which). Users should therefore treat gaps explicitly and apply filtering choices that match their application.

5.2.5 Gap datasets

Gap datasets identify continuous periods of missing or zero-reported demand, as detected in the daily aggregated datasets. File names follow the pattern `gap_[ISO3].csv` (e.g., `gap_BGD.csv`). Each row records the start and end date of a gap, along with its duration in days (inclusive; e.g., a gap from 1970-01-01 to 1970-01-02 has duration 1, and

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

to 1970-01-03 has duration 2). Datasets are available for Bangladesh (BGD), Bhutan (BTN), India (IND), Malaysia (MYS), Myanmar (MMR), Nepal (NPL), and the Philippines (PHL). Countries without gaps (Oman, Singapore, Sri Lanka, Taiwan, and Thailand) do not have associated files.

The column structure is as follows:

– `start_date`: First day of the gap (YYYY-MM-DD).

– `end_date`: Last day of the gap (YYYY-MM-DD).

– `duration`: Length of the gap in days (inclusive).

Currently, no distinction is made between gaps arising from missing days and those due to zero-reported demand, but such metadata could be added in future updates.

These datasets are complementary to the completeness measures (Section 5.2.6), offering finer-grained information on when gaps occur and how long they persist. They are useful for preprocessing (e.g., deciding when to apply imputation), for robustness checks in statistical analysis, and for understanding the temporal clustering of data issues.

5.2.6 Completeness dataset

The completeness dataset provides a high-level summary of temporal coverage and reliability for each country, stored in the file `dataset_completeness.csv`. It combines information on zero-demand days, consecutive missing days, and the overall dataset range to compute the percentage of usable days. Zero-demand days are days explicitly reported with 0 MWh. These are considered non-informative for trend analysis but may still reflect operational outages or reporting peculiarities. True gap days are consecutive days with no recorded data.

The `dataset_completeness_%` reflects the proportion of days with valid demand data over the full dataset range, including both zero-demand and missing days. All values are computed relative to the daily aggregated overall demand dataset. Table 13 details the column schema.

This dataset provides a rapid assessment of data quality across countries and highlights where imputation, caution, or further preprocessing may be necessary before statistical analyses or modelling.

The completeness dataset (`dataset_completeness.csv`) summarises coverage per country as the percentage of usable days over the dataset range, where both zero-demand and missing days are considered unusable. These datasets thus provide an overview of data quality and can be used to guide preprocessing and imputation decisions.

6 Example meteorological application

We now briefly discuss one of the possible applications of our dataset – building predictive models of daily demand using meteorological drivers. Here we take an extremely simplistic approach, fitting a third-degree (i.e., nonlinear)

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Table 13. Column descriptions for the completeness dataset.

Column	Description	Unit / Format
`country`	ISO 3166-1 alpha-3 country code	Text
`dataset_completeness_%`	Percentage of non-missing days over dataset range	Percentage (0–100)
`total_missing_days`	Total number of days with no data available	Count (days)
`zero_generation_days`	Number of valid days with zero reported demand	Count (days)
`true_gap_days`	Number of consecutive missing days (true gaps)	Count (days)
`dataset_range_days`	Total number of days in dataset range	Count (days)
`largest_gap_days`	Length of the largest continuous missing-data gap	Count (days)

polynomial of daily maximum 2-m dewpoint temperature to daily demand for each of the twelve countries in our dataset (Fig. 11), using only the last 365 days of available demand data, so as to avoid the need for detrending mean or variance.

Despite the simple method, some interesting features emerge. All countries feature a positive correlation between the two variables, and for all except Sri Lanka, the correlation coefficient is significantly different from zero with at least 95% confidence. Those countries nearest to the Equator (Singapore, Malaysia, Sri Lanka) have a fairly weak relationship between dewpoint and demand, either because the variability of the former is relatively small (several degrees, versus several tens of degrees in other countries) and so unable to drive large swings in demand; or because demand is saturated as air conditioning is always on. Nevertheless, even in Sri Lanka (where the relationship is the weakest), we see that the days of lowest demand tend to have the lowest dewpoint, and the days of highest demand tend to occur during days with the highest dewpoint.

Moving further north, the remaining nine countries exhibit much stronger relationships between dewpoint and demand. In countries where temperature falls sufficiently low (e.g. Taiwan and Nepal), we see the classic ‘u-shaped’ response that is common in European countries (Bessec and Fouquau, 2008). Also of interest is the relatively poor correlation between Indian national demand and dewpoint. This is likely due to the population being distributed across different climates (from the tropical south to the subtropical and hilly north) and different socioeconomic groups. More advanced data driven models that take these factors into account report very high skill (e.g., Hunt and Bloomfield, 2025a)

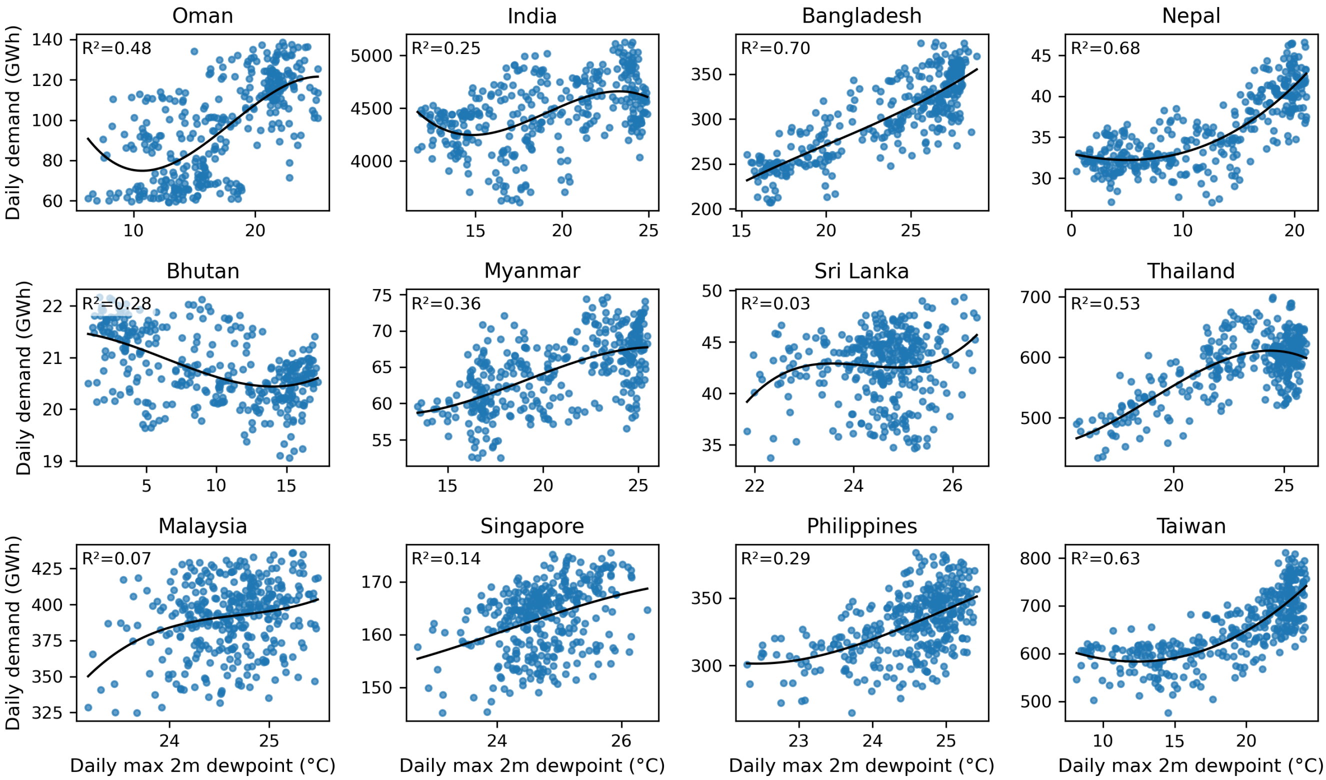

Figure 11. Relationship between daily electricity demand and dewpoint temperature for twelve SASEA countries. Each panel shows the national daily demand (GWh) plotted against area-averaged daily maximum 2-m dewpoint temperature (°C; from ERA5) for the most recent 365 non-zero days available in each record. The black curve shows a degree-3 polynomial regression fitted after removing outliers.

7 Discussion

7.1 Relationship to existing open electricity datasets

Several global or regional initiatives provide electricity data, but often at resolutions or with coverage that limit their use for weather-sensitive analysis in SASEA. The IEA real-time electricity tracker provides demand and generation series for many countries via a centralised interface, but coverage remains incomplete in this region and source definitions vary across countries. Ember provides high-quality harmonised electricity indicators globally, but at monthly resolution, which is too coarse for most weather-demand applications and for many operational studies of extremes, outages, or intraseasonal variability. In contrast, our dataset focuses on daily (and where available, sub-daily) national demand series for monsoonal and monsoon-adjacent countries, including those commonly missing

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

from global trackers. By providing a single consistent daily dataset across twelve countries, we enable cross-country comparisons at a timescale that is directly compatible with meteorological reanalysis products such as ERA5.

7.2 How to use this dataset

The dataset is designed to support applications where consistent daily demand records are required across multiple countries. Examples of this may include (i) weather–demand relationships and forecasting, e.g., using reanalysis data to train models that can predict demand as a function of weather (e.g., Hunt and Bloomfield, 2025a); (ii) resilience analysis, for example linking unusual demand behaviour to natural hazards (e.g., Hunt and Bloomfield, 2025b) and comparing between countries; (iii) planning and benchmarking; and (iv) economic or other studies, e.g., where demand can be used as a proxy for economic or political activity – such as the high level of disruption to the Myanmar dataset. The diagnostics provided with our dataset (e.g., completeness) are intended to help users quickly assess suitability for their chosen application.

7.3 How not to use this dataset

While this dataset is suitable for many research and planning applications, users should be cautious in some cases including, but not limited to, the following:

- Treating the data as fully homogeneous "ground truth" across countries. Some of our series measure operational demand, others use generation as a proxy for demand, and some may include trade or import definitions that differ between countries.
- Inferring outages or true "zero demand" events directly from zeros. In some cases, these may be genuine outages, but in others they could be reporting artefacts. Users interested in outage analysis should combine these series with independent operational records where possible.
- Computing cross-country correlations without accounting for temporal overlap (correlations are sensitive to record length) and non-stationarity (similar trends will inflate correlations).
- Assuming daily totals are directly comparable across time zones without checking time zone alignment.

7.4 Interpreting gaps and outliers

We do not apply interpolation in the released harmonised daily dataset. Instead, we provide completeness metrics and gap datasets identifying missing periods. Users may then choose to exclude long-gap periods e.g., from trend estimates, or to apply conservative imputation for short gaps, e.g., when training data-driven models that require continuous sequences.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

We also recommend that users do not automatically exclude flagged outliers without first inspecting their timing relative to known events (e.g., tropical cyclones). Our outlier flags are best used as diagnostic indicators for sensitivity tests (e.g., modelling with and without flagged days).

7.5 Future work

This dataset is intended as a foundation that can and should be extended. Next steps could include (i) expanding coverage to countries outside SASEA, (ii) adding sub-national demand series where available, and (iii) automated updates and near-real-time extensions.

8 Conclusions

We have developed and released a harmonised, open-access electricity demand dataset for twelve South and Southeast Asian countries at daily resolution (with sub-daily raw series preserved where available). The dataset addresses a major barrier to climate– and weather–energy research and regional energy planning in SASEA, i.e., the lack of standardised, intercomparable demand records across countries in this region.

We collected raw demand data were from national utilities, regulators, and international sources using automated reproducible retrieval and scraping scripts. We then converted all series into a consistent daily format with common units (MWh), ISO-standard dates, handling time zones and national calendar conventions. In addition, we computed quality control diagnostics (completeness metrics, gap summaries, and multi-method outlier flags) in order to support transparent preprocessing decisions. Validation against independent monthly demand estimates demonstrates strong agreement in timing and variability for most countries.

By making these harmonised time series and processing scripts openly available, we intend that our work will enable more robust analysis of demand variability, demand growth, and weather sensitivity across SASEA. Our dataset supports applications in data-driven demand forecasting, resilience assessment, and climate-impact research, and provides a flexible platform for future expansion to additional countries and regions.

9 Code availability

All code used for the analyses is available on GitHub at https://github.com/C4ntasaur/sasea-demand. The repository is primarily written in Python and includes a README with instructions for use and structure. Dependencies are documented within each script via import statements. The code is distributed under the MIT License.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

10 Data availability

All datasets generated and analysed in this paper are available on Zenodo (Coombes et al., 2025) (https://doi.org/10.5281/zenodo.17175212). The data are provided as comma-separated value (CSV) files accompanied by YAML metadata files describing structure, variables, and locations. The dataset is versioned and citable through the assigned DOI.

Author contributions. KH and HB conceived the project. KH and OC scraped the raw data. OC processed and published the data. OC and KH wrote the paper and created the figures. KH and HB edited the paper.

Competing interests. The authors declare that they have no competing interests

Acknowledgements. This work was supported by the CROCUS programme (NERC Climate system and biodiversity science for Challenges, Risks, and Opportunities – Collaborating in Understanding and Solutions), led by the University of Reading, which funded the author's research. KMRH is supported by a NERC Independent Research Fellowship (MITRE; NE/W007924/1). This study used data from the Copernicus Climate Change Service (C3S). The results contain modified Copernicus Climate Change Service information (2020). Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. Specifically, ERA5 hourly data on single levels from 1940 to present were obtained from the Copernicus Climate Data Store (Hersbach et al., 2018).
ChatGPT was used to write parts of this manuscript.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

References

Akil, Y. S., Fortuna, D. M., and Hasniaty, A.: ELD analysis under tropical climate using wavelet and regression approaches for power management: Indonesia case study, International Journal of Sustainable Engineering, 16, 197–210, 2023.

Al-Abri, A. and Okedu, K. E.: Overview of the Global Electricity System in Oman Considering Energy Demand Model Forecast, Energy Engineering, 120, 2023.

Albiento, A. Q., Nopia, J. Z. L., Umpacan, M. C., Mabborang, R. C., and Molina, M. G.: Forecasting Peak Electricity Consumption Demand in Luzon by utilizing ARIMA Model [J], European Journal of Computer Science and Information Technology, 2, 2023.

Ali, M., Iqbal, M. J., and Sharif, M.: Relationship between extreme temperature and electricity demand in Pakistan, International Journal of Energy and Environmental Engineering, 4, 36, 2013.

Ang, B. W., Wang, H., and Ma, X.: Climatic influence on electricity consumption: The case of Singapore and Hong Kong, Energy, 127, 534–543, 2017.

Atma, H., Ruzzenenti, F., and van den Broek, M. A.: Exploring the evolution of long-term electricity demand and load curves in emerging economies: A case study of Indonesia's energy transition, Energy Strategy Reviews, 60, 101 805, 2025.

Aziz, A., Mahmood, D., Qureshi, M. S., Qureshi, M. B., and Kim, K.: AI-based peak power demand forecasting model focusing on economic and climate features, Frontiers in Energy Research, 12, 1328 891, 2024.

Bessec, M. and Fouquau, J.: The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach, Energy Economics, 30, 2705–2721, 2008.

Bloomfield, H. C., Brayshaw, D. J., Deakin, M., and Greenwood, D. M.: Hourly historical and near-future weather and climate variables for energy system modelling, Earth System Science Data, 14, 2749–2766, https://doi.org/10.5194/essd-14-2749-2022, 2022.

Chandra, P. K., Bajaj, D., Sharma, H., Bareth, R., and Yadav, A.: Electrical load demand forecast for Gujarat state of India using machine learning models, in: 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1–6, IEEE, 2024.

Chapagain, K., Kittipiyakul, S., and Kulthanavit, P.: Short-term electricity demand forecasting: Impact analysis of temperature for Thailand, Energies, 13, 2498, 2020.

Chapagain, K., Acharya, S., Bhusal, H., Katuwal, S., Lakhey, O., Neupane, P., Sah, R. K., Tamang, B., and Rajbhandari, Y.: Short-term electricity demand forecasting for Kathmandu Valley, Nepal, Kathmandu University Journal of Science, Engineering and Technology, 15, 2021.

Chernyakhovskiy, I., Joshi, M., Palchak, D., and Rose, A.: Energy storage in south asia: understanding the role of grid-connected energy storage in south asia's power sector transformation, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2021.

Chophel, U., Chhoedron, D., and Jamtsho, K.: Dynamics of Electricity Demand: Analysis and Forecasting for Bhutan, Zorig Melong| A Technical Journal of Science, Engineering and Technology, 8, 169–174, 2025.

Chreng, K., Lee, H. S., and Tuy, S.: Electricity demand prediction for sustainable development in Cambodia using recurrent neural networks with ERA5 reanalysis climate variables, Energy Reports, 8, 76–81, 2022.

Coombes, O. G., Hunt, K. M. R., and Bloomfield, H. C.: Open access electricity demand data for South and Southeast Asia (daily, 2013–2025), https://doi.org/10.5281/zenodo.17175212, 2025.

Copernicus Climate Change Service: ERA5 hourly data on single levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.adbb2d47, accessed on 2023-03-07, 2023.

De Silva, G. B. and Samaliarachchi, L. A.: Peak electricity demand prediction model for Sri Lanka power system, Engineer: Journal of the Institution of Engineers, Sri Lanka, 46, 2013.

Debnath, R., Bardhan, R., Misra, A., Hong, T., Rozite, V., and Ramage, M. H.: Lockdown impacts on residential electricity demand in India: A data-driven and non-intrusive load monitoring study using Gaussian mixture models, Energy Policy, 164, 112 886, 2022.

Dhar, S., Parmar, A., Qiu, H., Senga, J. R. L., and Viswanathan, S.: A Stability-Driven Framework for Long-Term Hourly Electricity Demand Forecasting, arXiv preprint arXiv:2507.15001, 2025.

dr prodigy and contributors: Python Holidays: Generate and work with holidays in Python, https://pypi.org/project/holidays/, version used: 0.58 (Accessed: 20 September 2025), 2025.

Dunn, L. N., Geddes, J. A., Wilber, M., Mullen, E., Ruggles, T. H., Sánchez, J. J., McNerney, J., Rogers, M. A., Yang, H., Jones, C., and Kammen, D. M.: High-resolution power interruption data: revealing spatiotemporal patterns and environmental drivers, Energy Policy, 2019.

Ember: Monthly Electricity Data, Dataset, https://ember-energy.org/data/monthly-electricity-data/, monthly electricity generation, emissions and demand data for 85 geographies, 2025.

Fezzi, C. and Fanghella, V.: Real-time estimation of the short-run impact of COVID-19 on economic activity using electricity market data, Environmental and Resource Economics, https://doi.org/10.1007/s10640-020-00467-4, 2020.

Gulati, P., Kumar, A., and Bhardwaj, R.: Impact of Covid19 on electricity load in Haryana (India), International Journal of Energy Research, 45, 3397–3409, 2021.

Gupta, E.: Global warming and electricity demand in the rapidly growing city of Delhi: A semi-parametric variable coefficient approach, Energy economics, 34, 1407–1421, 2012.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., et al.: ERA5 hourly data on pressure levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS)[data set], 2018.

Hung, M.-F. and Huang, T.-H.: Dynamic demand for residential electricity in Taiwan under seasonality and increasing-block pricing, Energy Economics, 48, 168–177, 2015.

Hunt, K. M. and Bloomfield, H. C.: Building and explaining data-driven energy demand models for Indian states, Environmental Research: Energy, 2, 025 003, 2025a.

Hunt, K. M. R. and Bloomfield, H.: Landfalling tropical cyclones significantly reduce Bangladesh's energy security, Authorea Preprints, 2025b.

Hussain, F., Hasanuzzaman, M., and Abd Rahim, N.: Multivariate machine learning algorithms for energy demand forecasting and load behavior analysis, Energy Conversion and Management: X, 26, 100 903, 2025.

IEA, IRENA, UNSD, World Bank, and WHO: Tracking SDG7: The Energy Progress Report 2024, https://trackingsdg7.esmap.org/, 2024.

Islam, M. T., Turja, S. A., and Habib, A.: Enhanced power demand forecasting for Bangladesh: using feature engineering associated with environmental and economic impact, Journal of Data, Information and Management, pp. 1–19, 2025.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Jamshidi, E. J., Yusup, Y., Kayode, J. S., and Kamaruddin, M. A.: Detecting outliers in a univariate time series dataset using unsupervised combined statistical methods: A case study on surface water temperature, Ecological Informatics, 69, 101 672, 2022.

Joshi, J., Magal, A., Limaye, V. S., Madan, P., Jaiswal, A., Mavalankar, D., and Knowlton, K.: Climate change and 2030 cooling demand in Ahmedabad, India: opportunities for expansion of renewable energy and cool roofs, Mitigation and Adaptation Strategies for Global Change, 27, 44, 2022a.

Joshi, M., Palchak, D., de Silva, T., and Stephen, G.: Reliability and Resiliency in South Asia's Power Sector: A Framework for Improving Planning, Tech. Rep. NREL/TP-6A40-81826, National Renewable Energy Laboratory (NREL), https://www.nrel.gov/docs/fy22osti/81826.pdf, 2022b.

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L.: Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median, Journal of experimental social psychology, 49, 764–766, 2013.

Liang, E. B. J.: Short-term load forecasting in Singapore's energy market, https://dr.ntu.edu.sg/entities/publication/a5078642-0074-430e-bc56-93a89a2824a1, 2021.

Miraz, K. F. A., Biswas, K., and Rudra, M.: Forecasting Electricity Demand of Bangladesh and its relation with physical and natural variables using group method of data handling model, Ph.D. thesis, Brac University, 2021.

Odarno, L. and Kinuthia, B.: How Open Data Can Help Solve Kenya's Energy Access Problems, https://www.wri.org/insights/how-open-data-can-help-solve-kenyas-energy-access-problems, 2018.

Parkpoom, S. and Harrison, G. P.: Analyzing the impact of climate change on future electricity demand in Thailand, IEEE Transactions on Power Systems, 23, 1441–1448, 2008.

Parreño, S. J.: Forecasting electricity consumption in the Philippines using ARIMA models, International Journal of Machine Learning and Computing, 12, 279–285, 2022.

Phu, L. V.: Nonlinear temperature response of electricity loads and implications for power development policies in Vietnam, Energy and Buildings, 251, 111 339, 2021.

Priyadarshana, A., Lokupitiya, R., Kuruppuarachchi, D., and Lokupitiya, E.: Using weather patterns to forecast electricity consumption in Sri Lanka: An ARDL approach, International Energy Journal, 21, 257–268, 2021.

Rajbhandari, Y., Marahatta, A., Ghimire, B., Shrestha, A., Gachhadar, A., Thapa, A., Chapagain, K., and Korba, P.: Impact study of temperature on the time series electricity demand of urban Nepal for short-term load forecasting, Applied System Innovation, 4, 43, 2021.

Santos, A. G.: Forecasting residential electricity demand in the Philippines using an error correction model, Philippine Review of Economics (Online ISSN 2984-8156), 57, 121–151, 2021.

Shah, I., Iftikhar, H., and Ali, S.: Modeling and forecasting medium-term electricity consumption using component estimation technique, Forecasting, 2, 163–179, 2020.

Shao, J., Xu, T., and Zhou, K.: CGMTSA: An R package for continuous glucose monitoring time series data analysis, bioRxiv, 2020.

Shiwakoti, R. K., Limcharoen, P., and Uduwage, D.: SHORT-TERM ELECTRICITY DEMAND FORECASTING IN Sri Lanka USING STATISTICAL AND DEEP LEARNING MODELS, in: Proceedings The 13th World Construction Symposium| August, p. 1261, 2025.

https://doi.org/10.5194/essd-2026-76
Preprint. Discussion started: 26 February 2026

Staffell, I. and Pfenninger, S.: The increasing impact of weather on electricity supply and demand, Energy, 145, 65–78, https://doi.org/10.1016/j.energy.2017.12.051, 2018.

Steinbuks, J., de Wit, J., Kochnakyan, A., and Foster, V.: Forecasting Electricity Demand: An Aid for Practitioners, Tech. Rep. 2017/73, World Bank, https://doi.org/10.1596/26189, 2017.

Tai, V., Tan, Y., Faiza, N., Che, H., Chia, C., Saw, L., and Ali, M.: Long-term electricity demand forecasting for Malaysia using Artificial Neural Networks in the presence of input and model uncertainties, Energy Engineering: Journal of the Association of Energy Engineers, 118, 715, 2021.

Torculas, E., Rentillo, E. J., and Ambita, A. A.: Forecasting of Energy Consumption in the Philippines Using Machine Learning Algorithms, in: Asian Conference on Intelligent Information and Database Systems, pp. 424–435, Springer, 2023.

Uddin, S. M., Malaker, M., and Alam, M. S.: Forecasting Electricity Demand of Bangladesh by Trend Analysis Method, International Conference on Mechanical, Industrial and Materials Engineering, https://icmime-ruet.ac.bd/2019/DIR/Contents/Technical%20Papers/Industrial%20Engineering/IE-21.pdf, 2019.

Ullah, I., Hasanat, S. M., Aurangzeb, K., Alhussein, M., Rizwan, M., and Anwar, M. S.: Multi-horizon short-term load forecasting using hybrid of LSTM and modified split convolution, PeerJ Computer Science, 9, e1487, 2023.

Wang, X., Si, C., Gu, J., Liu, G., Liu, W., Qiu, J., and Zhao, J.: Electricity-consumption data reveals the economic impact and industry recovery during the pandemic, Scientific Reports, 11, 19 960, https://doi.org/10.1038/s41598-021-98259-3, 2021.

Zaim, S., Yusoff, W. N. S. W., Mohamad, N. N., Radi, N. F. A., and Yaziz, S. R.: Forecasting of electricity demand in Malaysia with seasonal highly volatile characteristics using SARIMA–GARCH model, Matematika, pp. 293–313, 2023.

Zhong, R. and Charturvedi, S.: Scorching Heat Exposes India's Power Woes, https://www.wsj.com/articles/scorching-heat-exposes-indias-power-woes-1402491922, 2014.