

Supplementary Information

Daily operating states of residential electricity demand in India: Within-household variation, thermal conditions and peak demand

Krithika Panicker, Prashant Anand, Abraham George

Department of Architecture and Regional Planning, Indian Institute of Technology Kharagpur, West Bengal 721302, India

Supplementary Table S1. Characteristics of survey-linked households in the final sample

Monitoring location	Survey households	Household size, median (IQR)	Rooms, median (IQR)	Floor area, median (IQR)	AC ownership, n (%)	Water-heater ownership, n (%)
Aurangabad district	31	5.0 (4.0–6.0)	5.0 (3.0–6.0)	750 (475–900) ft ²	6 (19.4)	7 (22.6)
Gonda district	11	6.0 (5.5–9.5)	4.0 (3.0–5.0)	800 (450–800) ft ²	0 (0.0)	1 (9.1)
Kanpur rural	13	5.0 (5.0–6.0)	3.0 (2.0–4.0)	1,425 (1,200–1,500) ft ²	1 (7.7)	1 (7.7)
Pune district	28	5.0 (3.8–6.0)	3.0 (3.0–4.0)	700 (465–850) ft ²	5 (17.9)	6 (21.4)
Pune city	43	4.0 (3.0–4.0)	4.0 (3.0–4.0)	750 (600–947.5) ft ²	16 (37.2)	25 (58.1)
Overall	126	5.0 (4.0–6.0)	4.0 (3.0–5.0)	750 (557.8–900) ft ²	28 (22.2)	40 (31.7)

Note. Survey data were available for 126 of 133 households; floor area was available for 106. Values are median (IQR) or n (%). IQR, interquartile range; AC, air conditioner.

Data preparation and load construction

Supplementary Table S2. Household-day quality-control rules

Stage	Operational criterion	Treatment
Duplicate interval	Same household, date and 15-minute block recorded more than once	Duplicate records replaced by their arithmetic mean
Missing interval	No more than 4 missing blocks/day; longest run no more than 2; interior gap only; energy change no more than 5%	Linear interpolation between adjacent valid readings
Recorded zero	No more than 4 zero blocks/day; longest run no more than 2; no edge zeros; zero share no more than 10%; energy change no more than 5%	Interpolation under the stated limits; zero-quality category retained
Implausible day	Daily electricity below 0.05 or above 100 kWh; or maximum demand above 20 kW	Excluded from subsequent analysis
Primary operating-day sample	Active electricity at least 0.20 kWh; stable persistent-power estimate; acceptable zero quality	Retained for state classification

Supplementary Table S3. Daily demand measures and exact compact features used in the analysis

Indicator or feature	Operational definition	Analytical use
Persistent power	Hybrid fifth-percentile household-day estimate described below.	Persistent component used in active-load construction.
Active power	Observed 15-minute whole-house power minus persistent power, bounded at zero.	Power above the estimated persistent component.
Active electricity	Sum of 15-minute active power multiplied by 0.25 h.	Daily active-demand electricity.
Maximum demand	Largest 15-minute average-power observation.	Daily maximum 15-minute average power.
Peak excess	Maximum demand minus estimated persistent power.	Daily maximum demand above persistent power.
Peak time	Clock time of the 15-minute interval containing maximum demand.	Clock time of daily maximum demand.
Morning share	Sum of the smoothed normalised hourly active-load profile from 05:00 to 09:59.	Clustering feature: morning active-load share.
Daytime share	Sum of the smoothed normalised hourly active-load profile from 10:00 to 15:59.	Clustering feature: daytime active-load share.
Evening share	Sum of the smoothed normalised hourly active-load profile from 16:00 to 21:59.	Clustering feature: evening active-load share.
Peak-hour sine	$\sin(2\pi h/24)$, where h is the hour containing the maximum smoothed normalised hourly share.	Clustering feature: circular encoding of peak hour.
Peak-hour cosine	$\cos(2\pi h/24)$, where h is the hour containing the maximum smoothed normalised hourly share.	Clustering feature: circular encoding of peak hour.

Indicator or feature	Operational definition	Analytical use
Top-two-hour share	Sum of the two largest values in the smoothed normalised 24-hour active-load profile.	Clustering feature: concentration in the two highest hours.
Number of local peaks	Number of hourly shares greater than the preceding hour, at least as large as the following hour, and at least 0.05 of the daily active-load share.	Clustering feature: number of local maxima.

Supplementary Table S4. Construction of the analytical sample

Analytical stage	Household-days	Households	Share of candidate days (%)
Candidate household-days	71,038	139	100.00
Retained after missing-data rules	68,397	139	96.28
Retained after daily-value filters	62,496	139	87.98
Final operating-day sample	44,571	133	62.74
Household-balanced training subset	17,747	133	24.98
Final households matched to survey	—	126	—
Final households without survey match	—	7	—

Note. For centroid fitting, households with more than 180 eligible days were sampled to 180 days without replacement (seed 42); all eligible days were retained for shorter records. The fitted centroids were applied to all 44,571 operating days.

Supplementary Table S5. Sensitivity of the state solution to missing- and zero-reading treatment

Analytical variation	Days	Households	Silhouette	ARI vs primary	Minimum profile correlation
Primary repaired refit	44,571	133	0.29	1.00	1.00
Same days without short-zero repair	44,571	133	0.29	0.99	1.00
Alternative strict zero-screening refit	44,644	133	0.29	0.99	1.00
Complete days only	44,235	133	0.30	0.98	1.00
Lenient inclusion of zero-heavy days	52,489	136	0.25	0.63	0.19

Note. Alternative state profiles were matched to the primary S1–S4 ordering before correlation. The alternative strict zero-screening sample was reconstructed independently from the raw profiles; because screening and operating-day eligibility were reapplied within that alternative pipeline, its retained day count can differ slightly from the primary sample.

Persistent-power estimation and active-load construction

Persistent power was estimated using a hybrid fifth-percentile procedure. The fifth percentile of the 96 quarter-hour power observations was used as the primary estimate. When this percentile was zero and zero-valued observations constituted no more than 10% of the day, it was replaced by the fifth percentile of positive observations. The estimate was constrained not to exceed mean daily power. Active power was calculated as observed power above this persistent component, bounded at zero. Hybrid first- and tenth-percentile variants were evaluated as sensitivity analyses.

Supplementary Table S6. Sensitivity to the persistent-load definition

Hybrid estimator	Silhouette	ARI vs q05	Minimum profile correlation	Median persistent electricity (kWh/day)	Median active electricity (kWh/day)	Median peak excess (kW)
q01	0.295	0.951	0.996	0.989	3.398	0.869
q05	0.293	1.000	1.000	1.385	3.021	0.843
q10	0.292	0.950	0.994	1.691	2.733	0.825

Operating-state model selection and robustness

Supplementary Table S7. Profile representations evaluated in K-means model selection

Representation	Construction	Representation form
Raw 96-interval profile	L1-normalised 15-minute active-load profile (96 values).	96-dimensional quarter-hour profile.
Smoothed hourly profile	Four quarter-hour active-load shares summed within each clock hour, followed by a three-hour circular moving average and L1 normalisation (24 values).	24-dimensional smoothed hourly profile.
Cosine-normalised hourly profile	Smoothed hourly profile additionally L2-normalised before Euclidean K-means.	24-dimensional L2-normalised hourly profile.
Hourly principal components	PCA fitted only on the household-balanced training profiles; components retained to reach at least 90% cumulative explained variance, with a maximum of 10 components; scores standardised before clustering.	PCA scores retaining at least 90% cumulative explained variance, with a maximum of 10 components.
Compact temporal features	Seven features in Supplementary Table S3; non-finite values replaced by training-set medians and features z-standardised using the training subset.	Seven standardised temporal features.
Hybrid representation	Standardised hourly-PCA and compact-feature blocks were separately scaled by the square root of their dimensionality and concatenated.	Concatenated scaled PCA and seven-feature blocks.

Supplementary Table S8. Selected compact-feature candidate solutions (k = 2–4)

k	Silhouette	Negative silhouette (%)	Davies–Bouldin	Minimum share (%)	Maximum share (%)	Mean seed ARI	Status
2	0.32	2.17	1.36	43.28	56.72	0.99	Comparison model
3	0.32	1.92	1.31	21.96	52.09	1.00	Comparison model
4	0.29	1.60	1.28	14.45	47.77	0.98	Selected four-state solution

Supplementary Table S9. Stability and sensitivity of the selected four-state solution

Analysis	Runs or sample	Assignment agreement	State-specific or profile metric
Repeated initialisation	5 fits	Mean 0.98; minimum 0.96	—
Household bootstrap	50 resamples	Median ARI 0.91; 5th percentile 0.51	S2 recovery: median 0.97; 5th percentile 0.15
Region-balanced refits	50 refits	Median ARI 0.65; 5th percentile 0.62	S2 recovery: median 0.48; 5th percentile 0.40
Common monitoring period	22,241 days	ARI 0.96	Minimum centroid similarity 0.98
1-hour smoothing	44,571 days	ARI 0.62	Minimum profile correlation 0.94
3-hour smoothing	44,571 days	ARI 1.00	Minimum profile correlation 1.00
5-hour smoothing	44,571 days	ARI 0.58	Minimum profile correlation 0.35 (S2)

Note. ARI, adjusted Rand index. Household-bootstrap model stability used 50 resamples; confidence intervals in explanatory analyses used 1,000 resamples. S2 recovery denotes the state-specific agreement statistic for the refitted state matched to primary S2.

Weather, appliance-ownership, household-dynamics and peak analyses

Weather assignment and thermal indicators

Historical daily weather corresponding to the monitoring records was obtained from the Open-Meteo Historical Weather API [2]. Daily mean 2-m air temperature (temperature_2m_mean) was retrieved in the Asia/Kolkata time zone using fixed coordinates for Pune (18.5204° N, 73.8567° E), Aurangabad (19.8762° N, 75.3433° E), Gonda (27.1339° N, 81.9619° E) and Kanpur (26.4499° N, 80.3319° E). Pune city and Pune district used the Pune series; the remaining sites used their corresponding city series. Weather records were matched by monitoring location and local calendar date. The API default Best Match model selection was used.

Cooling degree days above 24 °C were calculated as $CDD_{24} = \max(0, T_{\text{mean}} - 24 \text{ }^{\circ}\text{C})$, consistent with prior work in tropical and emerging-economy settings [1]. Positive CDD_{24} values were divided into monitoring-location-specific tertiles of unique location-date observations; daily mean temperature was divided separately into monitoring-location-specific quartiles. Seasons were defined as winter (December–February), summer (March–May), monsoon (June–September) and post-monsoon (October–November).

Eligibility criteria and analytical measures

Supplementary Table S10. Eligibility criteria and principal measures for explanatory analyses

Analysis	Eligibility or design	Reported measure
Weather and season	Within-household contrasts required at least 14 retained days in each compared condition. Positive CDD ₂₄ days were divided into monitoring-location-specific tertiles; mean temperature was divided into monitoring-location-specific quartiles.	Change in household state share, with common-period and equal-region standardised sensitivities.
Air-conditioner ownership	At least 30 valid days, at least 10 distinct CDD ₂₄ values, CDD ₂₄ range at least 5 °C·day, and known ownership. Common-support sensitivity required at least 3 AC and 3 non-AC households in a region.	Difference in household-specific CDD ₂₄ response of active electricity associated with AC ownership; 60-, 90- and 120-day minimum-coverage sensitivities.
Electric water-heater ownership	At least 14 retained days per household-season and known ownership. Threshold sensitivities used 5, 14, 30 and 60 days.	Region- and season-adjusted ownership differences in morning active-use share and peak excess; household-equal, AC-adjusted and seasonal-interaction sensitivities.
Household state composition	At least 30 retained days spanning at least 2 months and 2 seasons.	State proportions, dominant state and normalised Shannon entropy.
State transitions	Consecutive calendar-day pairs; at least 20 total transitions per household and at least 5 transitions from the relevant origin state for the primary household-balanced probability.	Pooled transition frequencies and household-balanced persistence/switching probabilities with household-bootstrap confidence intervals.
Peak-day analysis	Pooled and household-relative upper 1%, 5% and 10% peak-excess days. Household-relative estimates required at least 20 retained days and at least 5 days in the selected upper-tail set.	State composition of pooled high-peak days and within-household change in state prevalence on each household's own high-peak days.

Operating-state prevalence variance with thermal conditions.

Within households, prevalence of the concentrated morning-peak state (S1) decreased by 13.06 percentage points between days with CDD₂₄ = 0 and days in the highest monitoring-location-specific tertile of positive CDD₂₄, by 11.78 percentage points from winter to summer, and by 13.72 percentage points between the coolest and hottest monitoring-location-specific temperature quartiles. Repeating the comparisons over the common monitoring period produced corresponding decreases of 14.33, 11.25 and 14.10 percentage points. The reduction in S1 was reproduced across all three thermal contrasts and both observation periods. Redistribution among the remaining states differed between the full-record and common-period analyses: the diffuse morning-led state (S2) increased across all three contrasts in the full-record analysis, whereas this pattern was not retained over the common period, where increases were observed for the concentrated daytime-peak state (S3) and extended evening-to-night state (S4). Detailed state-specific estimates and 95% confidence intervals are reported in Supplementary Table S11 and Supplementary Fig. S6.

Supplementary Table S11. Household-paired changes in state prevalence across full and common monitoring periods

Period	Thermal contrast	Households	S1, pp (95% CI)	S2, pp (95% CI)	S3, pp (95% CI)	S4, pp (95% CI)
Full	CDD ₂₄ : zero to highest positive tertile	89	-13.06 (-17.25 to -8.93)	+7.36 (3.29 to 11.26)	+1.19 (-2.05 to 4.57)	+4.50 (-0.85 to 9.93)
Full	Winter to summer	84	-11.78 (-15.67 to -8.20)	+4.74 (1.19 to 8.29)	+2.25 (-0.59 to 5.31)	+4.80 (-0.06 to 9.99)
Full	Coolest to hottest temperature quartile	91	-13.72 (-18.41 to -9.31)	+7.24 (3.51 to 11.21)	+1.83 (-1.31 to 5.02)	+4.65 (-0.57 to 9.65)
Common	CDD ₂₄ : zero to highest positive tertile	63	-14.33 (-20.22 to -9.36)	-0.50 (-4.52 to 3.71)	+5.45 (0.82 to 10.43)	+9.38 (2.16 to 16.43)
Common	Winter to summer	71	-11.25 (-15.85 to -7.13)	+0.30 (-2.86 to 3.70)	+4.60 (0.95 to 8.65)	+6.34 (1.02 to 11.92)
Common	Coolest to hottest temperature quartile	67	-14.10 (-19.46 to -8.70)	+1.96 (-2.07 to 6.05)	+5.18 (0.19 to 10.31)	+6.96 (0.42 to 13.06)

Note. Positive values indicate a larger state share in the warmer or summer condition. Common period: 14 June 2019–12 May 2020. pp, percentage points.

Supplementary Table S12. Air-conditioner-ownership differences in household-specific CDD₂₄ response

Measure	Estimate (95% CI), kWh/day per °C·day	Households	AC/non-AC households
Difference in median CDD ₂₄ response	0.56 (0.19 to 0.79)	105	28/77
Mean slope difference	0.59 (0.37 to 0.82)	105	28/77

Measure	Estimate (95% CI), kWh/day per °C·day	Households	AC/non-AC households
Region-adjusted AC coefficient	0.62 (0.39 to 0.85)	105	28/77
Common-support-region coefficient	0.63 (0.40 to 0.88)	94	27/67

Supplementary Table S13. Electric water-heater ownership associations with morning active-use share and peak excess

Model	Morning-share difference, pp (95% CI)	Peak-excess difference, kW (95% CI)	Households	Owner/non-owner
Primary region- and season-adjusted	+15.30 (9.42 to 20.28)	1.05 (0.76 to 1.35)	110	38/72
Household-equal weighted	+14.28 (9.21 to 19.25)	1.00 (0.70 to 1.32)	110	38/72
Additionally adjusted for AC ownership	+15.35 (9.81 to 20.33)	1.04 (0.78 to 1.31)	110	38/72

Note. Morning active-use share covers 05:00–09:59. pp, percentage points.

Supplementary Table S14. Water-heater ownership-by-season interaction contrasts

Seasonal contrast	Morning-share interaction, pp (95% CI)	Peak-excess interaction, kW (95% CI)
Monsoon – summer	+4.35 (1.20 to 7.72)	0.28 (-0.03 to 0.57)
Post-monsoon – summer	+5.62 (1.35 to 10.08)	0.23 (-0.10 to 0.54)
Winter – summer	+7.65 (3.50 to 11.79)	0.38 (0.07 to 0.66)

Note. Interaction contrasts are relative to summer; positive values indicate a larger ownership difference than in summer.

Supplementary Table S15. Household state composition and temporal diversity

Measure	Result
Households in final sample	133
Households eligible for primary entropy analysis	109
Dominant S1	15
Dominant S2	23
Dominant S3	11
Dominant S4	82
Mixed/tied dominant state	2
Median normalised entropy	0.71
Mean normalised entropy	0.64

Supplementary Table S16. State persistence and principal household-balanced transitions

Origin	Eligible households	Pooled persistence (%)	Household-balanced persistence, % (95% CI)	Principal switching probabilities, % (95% CI)
S1	50	66.62	47.55 (39.64–54.97)	S1→S2 24.88% (19.32–30.54); S1→S3 14.16% (10.25–18.87)
S2	99	42.35	33.93 (30.60–37.63)	S2→S4 39.84% (35.34–44.58); S2→S3 15.62% (12.98–18.10)
S3	91	39.42	30.49 (26.80–34.33)	S3→S4 40.21% (35.22–45.82); S3→S2 18.68% (15.70–21.96)
S4	102	73.40	59.13 (54.81–63.93)	S4→S2 19.10% (16.35–22.01); S4→S3 17.00% (14.04–20.33)

Note. Transition probabilities are conditional on the origin state. Household-balanced values are means across eligible households.

Supplementary Table S17. State composition of pooled and household-relative top-5% peak days

State	Pooled top-5% share (%)	Within-household change (pp)	95% CI (pp)	Eligible households
S1	46.25	+8.82	5.75 to 12.52	99
S2	12.61	-3.91	-6.55 to -0.69	99
S3	23.87	+5.71	2.12 to 9.49	99
S4	17.27	-10.63	-14.87 to -6.30	99

Note. The pooled threshold uses all retained household-days; household-relative changes use each household's own top-5% peak-excess days.

Supplementary Table S18. Appliance-ownership estimates under alternative persistent-load definitions

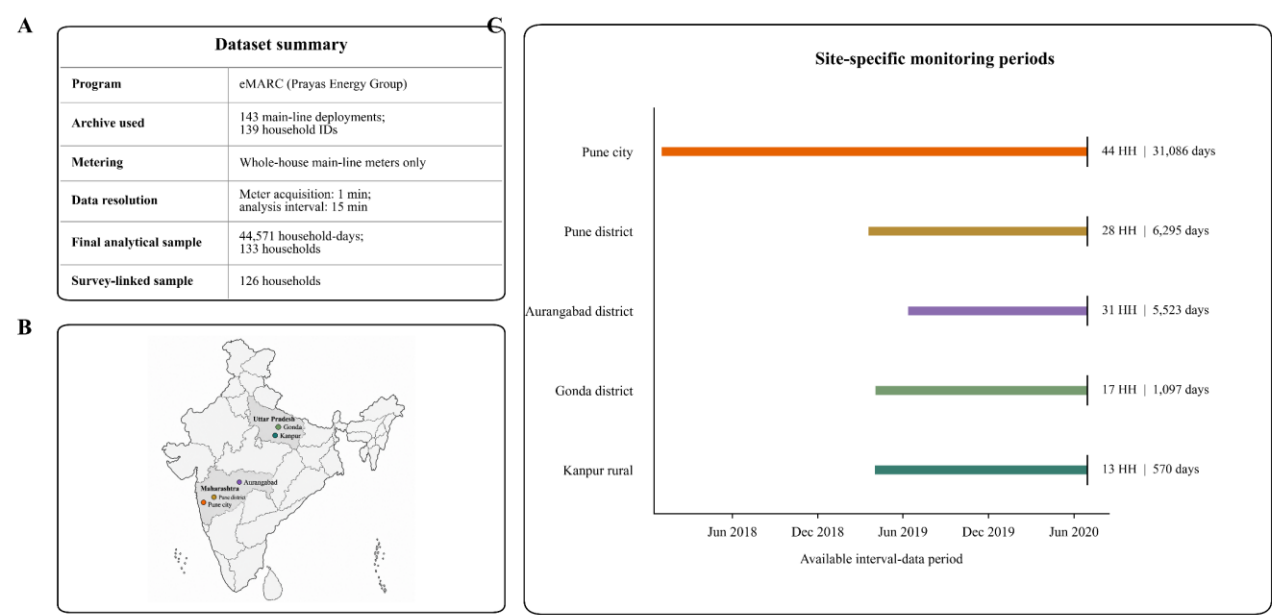
Hybrid estimator	AC median slope difference	AC region-adjusted coefficient	AC common-support coefficient	Water-heater morning-share difference (pp)	Water-heater peak-excess difference (kW)
q01	0.589	0.614	0.626	13.949	1.050
q05	0.558	0.616	0.630	15.297	1.049
q10	0.568	0.617	0.634	16.406	1.046

References

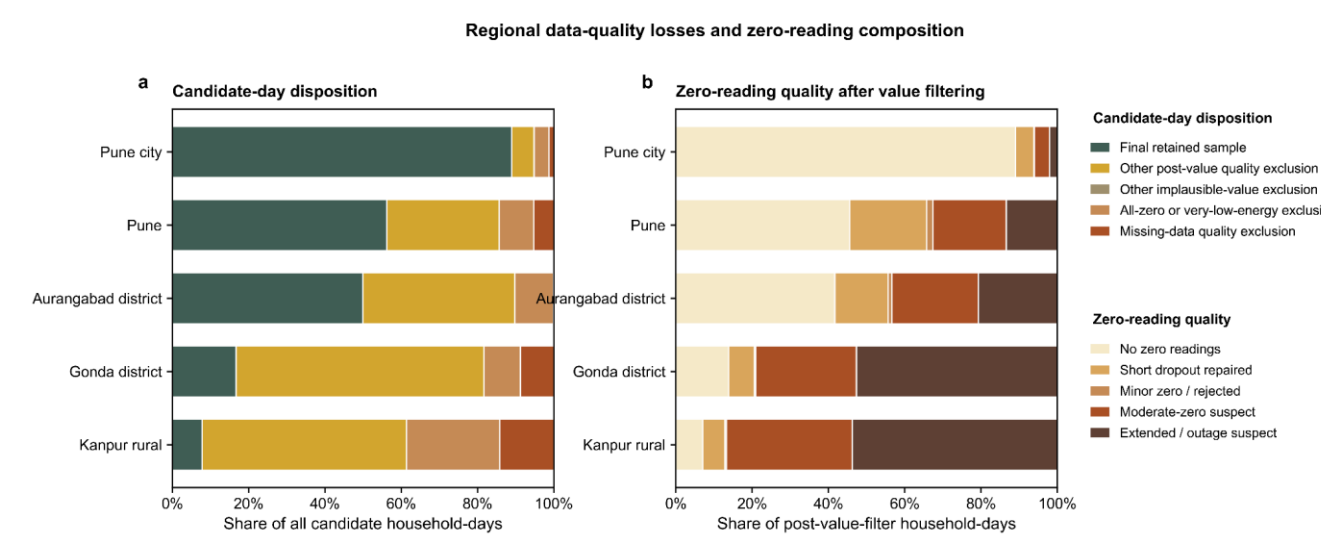
1. Pavanello, F. et al. Air-conditioning and the adaptation cooling deficit in emerging economies. Nat. Commun. 12, 6460 (2021). <https://doi.org/10.1038/s41467-021-26592-2>

2. Zippenfenig, P. Open-Meteo.com Weather API [Computer software]. Zenodo (2023). <https://doi.org/10.5281/zenodo.7970649>

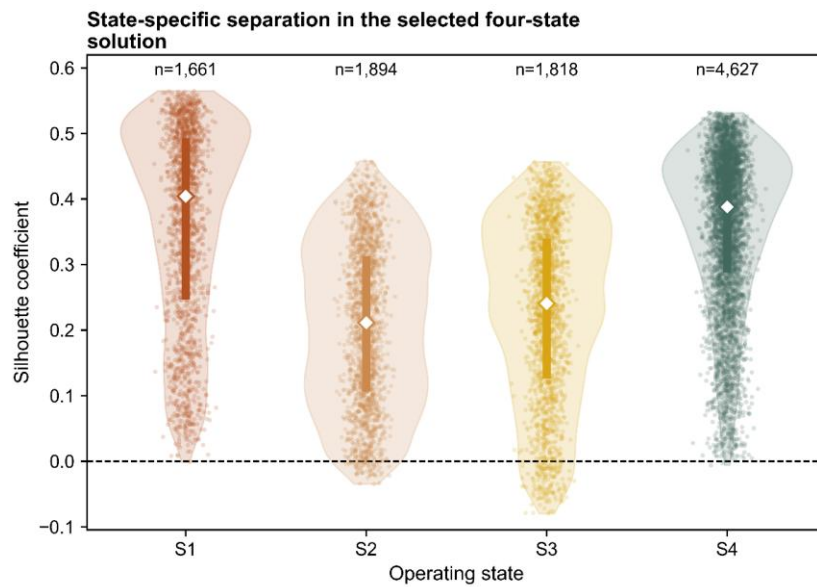
Supplementary figures



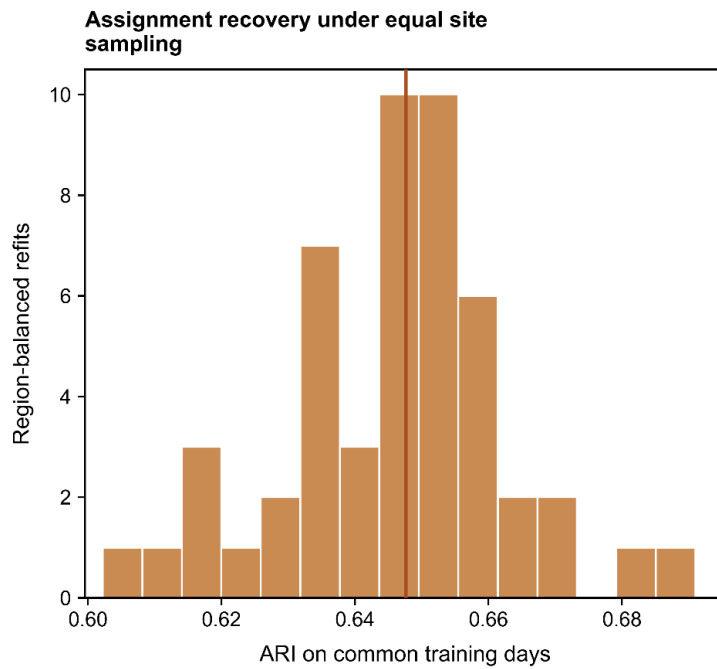
Supplementary Figure S1. eMARC dataset and monitoring context. (a) Dataset summary and analytical sample. (b) Monitoring locations represented in the dataset. (c) Site-specific monitoring periods and numbers of households and retained household-days.



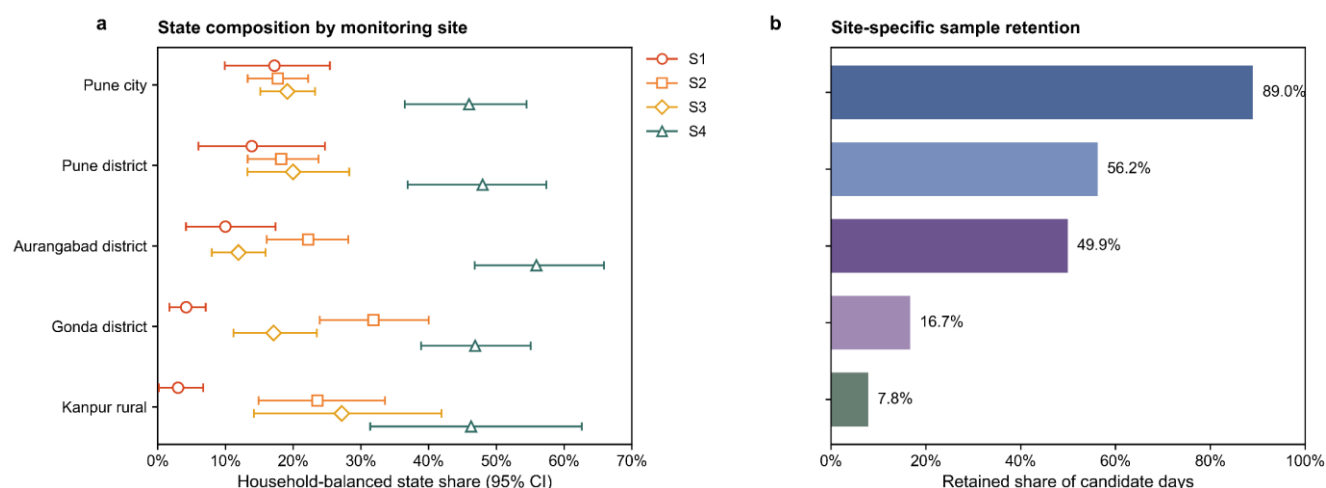
Supplementary Figure S2. Candidate-day disposition and zero-reading composition across the five monitoring locations. (a) Share of candidate household-days retained or excluded at successive quality-control stages. (b) Zero-reading quality composition after value filtering.



Supplementary Figure S3. State-specific silhouette distributions for the selected four-state solution.

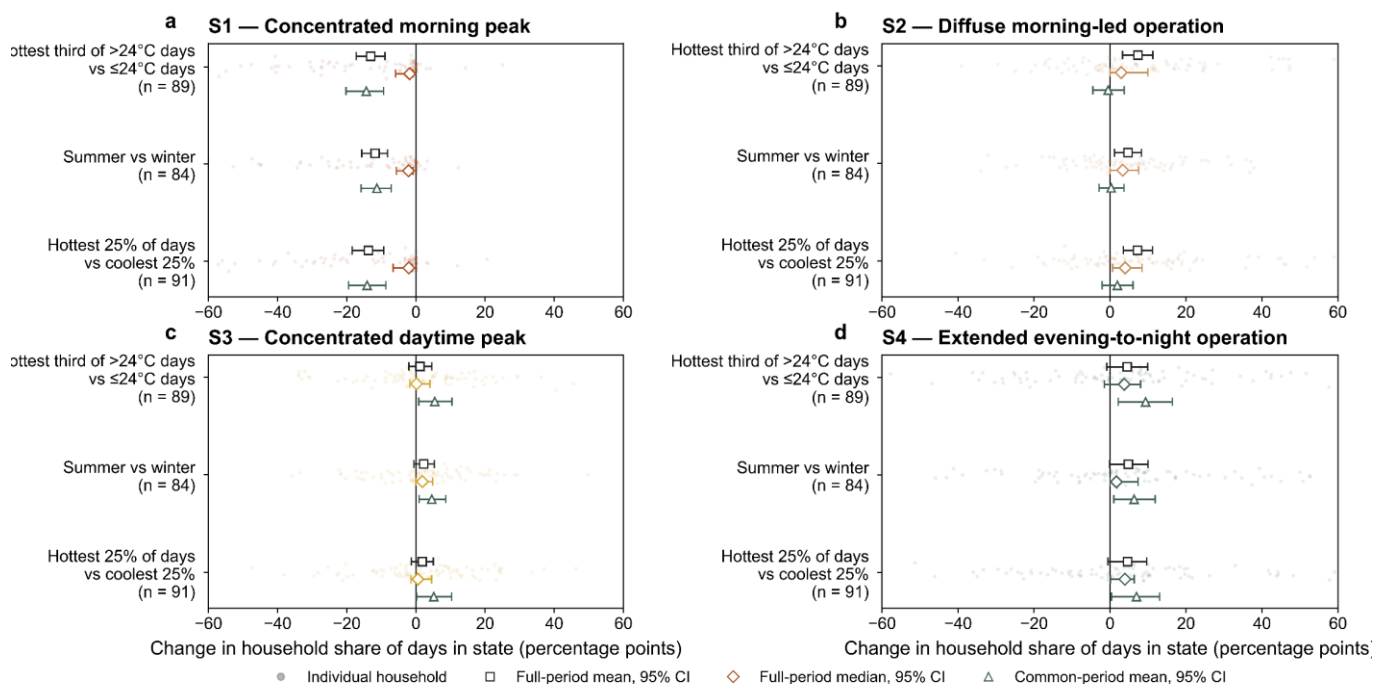


Supplementary Figure S4. Assignment agreement between the primary state solution and 50 region-balanced refits using adjusted Rand index (ARI) values on common training days across the refits



Supplementary Figure S5. Household-balanced regional state shares and site-specific sample retention. (a) Household-balanced state shares by monitoring location with 95% confidence intervals. (b) Retained share of candidate household-days at each monitoring location.

Operating-state prevalence variance with thermal conditions.



Supplementary Figure S6. Household-paired changes in operating-state prevalence under cooling-degree, seasonal and temperature-based definitions of thermal conditions. (a–d) Changes in within-household prevalence of S1–S4, respectively.