Imperial College London
Department of Earth Science and Engineering
MSc in Environmental Data Science and Machine Learning

Independent Research Project
Final Report

Optimising Demand Response Strategies for Carbon Intelligent Electricity Usage

By
Daniel Kaupa

Email: daniel.kaupa24@imperial.ac.uk
GitHub username: esemsc-dbk24
Repository: https://github.com/ese-ada-lovelace-2024/irp-dbk24

Supervisors:
Dr. Mirabelle Muûls
Dr. Shefali Khanna
Dr. Gareth Collins

August 29th, 2025

ABSTRACT

This study builds a transparent, public-data pipeline that turns grid and weather feeds into actionable, half-hourly marginal emission factors (MEFs) for India and tests whether those signals can drive real-world load shifting. Using ERA5/ERA5-Land weather and national grid data (2020–2025), we fit a penalised GAM on net load; its derivative yields a raw MEF that we linearly calibrate to short-horizon ramps and map to ramp-based confidence. Correlations are modest at small ramps and stronger ≥1 GW—good enough to steer. Midday MEFs often exceed averages, consistent with PV saturation and fossil units on the margin. A conservative greedy scheduler (±2 h; ≤1/day; ≤3/week) over ~238k customers re-times ~3.7% of energy; per-home savings are small (~6.3 kg CO_2/yr) but MEFs change accounting (+313% vs AEFs). The pipeline is reproducible and best for ranking hours, while larger gains lie upstream.

1. INTRODUCTION

1.1 Background and motivation

Electrify everything and decarbonise fast. This dual imperative of the net-zero transition combined with rising living standards, digitalisation, data centre-growth, and expanding electricity access across the Global South, is pushing an increasing amount of demand onto the grid [36]. While meeting this rising demand can require more output from conventional fossil units, falling costs and policy support are simultaneously lifting the share of wind and solar into the supply mix [31]. On average, higher renewable penetration lowers emissions, but weather- and daylight-driven variability makes the grid's instantaneous emissions intensity more volatile. As a result, the emissions impact of electricity depends not only on how much is used, but also when it is used [4].

Recommendations for reducing grid emissions often first focus on reducing the electricity usage that causes them. The logic is sound: consume less, emit less. While this may be the most effective strategy, it encounters resistance with the comfort norms and service levels expected by customers. But the increased temporal variability of emissions intensity can offer a second, more palatable lever for customers: consume the same, but change the time of consumption to reduce emissions. In addition to the potential grid stability benefits, shifting time of usage has indeed shown promise for reducing emissions. But to evaluate the emissions reduction potential of a shift in usage, we of course need reliable emissions factors. The average emission factor (AEF) is perhaps the most familiar metric, measuring total emissions and total generation over a given geospatial and temporal range, and useful for describing systems. But when measuring the specific effects of shifting usage across time, the marginal emission factor (MEF) is the metric of choice in the literature [38].

The marginal emissions factor is the emissions rate associated with the next unit of electricity at the moment of use; the unit provided by whichever generator is setting the balance at that time. An accessible analogy is the "last-mile" of a delivery: most of the journey would happen anyway; the extra impact is in the final leg triggered by the order. Likewise, many plants run regardless of small changes in demand; the consequential impact sits with the unit that ramps up or down when consumption shifts. As weather-dependent renewables take a larger share, this rate can vary markedly across the day.

1.2 Operational Challenges

While the choice to use MEFs rather than AEFs is clear, there is far less agreement on how to determine these 'last-mile' emission rates in a way that provides reliable, high-resolution, and easily reproducible results. True MEF calculation requires extremely granular and detailed information to understand at any given point which units are online, the fuel and heat rates of those units, ramp and minimum up/down constraints, outages and maintenance, interconnector flows, transmission limits, and more. However, these data are usually proprietary to utilities and system operators, making them largely inaccessible to external stakeholders. Consequently, those seeking to use MEFs must instead rely on estimation techniques and attempt to reconstruct grid dispatch patterns from public data feeds and the occasional (usually scope-limited) proprietary drops. These reconstructions are often laden with assumptions and specific to a particular grid region and time; as a result, they are brittle as conditions change and difficult to reproduce across geographies.

1.3 Related Work

Hawkes (2010) regresses the first differences of system CO_2 rate on the first differences of total load using observed dispatch in Great Britain from 2002–2009 [19]. In this study, they report an average MEF of ~0.69 kg CO_2/kWh (±10%) from their model with a promising R^2 value of 0.95. While informative, this is an observational regression, and the high correlation alone should not be read as accurate causal identification. In modern grids, coincident drivers can move emissions and load together for reasons not caused by the load change itself (e.g. solar output rising with midday demand, wind regimes shifting around the evening peak, scheduled hydro releases, interconnector flows, or must-run/nuclear constraints). Even in first-differences, those co-movements can bias the slope (sometimes severely as VRE shares grow), yielding a tight fit that reflects diurnal/seasonal patterns rather than the true marginal response to an incremental load change.

WattTime (a leading provider of marginal operating emission rates) explicitly documents why pure differencing and simple binning can be biased, and notes they abandoned difference-based models in 2014 [37]. Although their full production methodology and validation datasets are not fully public, their publications offer a useful taxonomy of methods. These include the aforementioned simple differencing across adjacent intervals, binned regressions that condition on grid "states" (hour/season/load level), experimental and quasi-experimental designs, and price/heat-rate inference to identify the marginal fuel. Their current preference emphasises hybrid models that combine grid-state–conditioned regressions fuel-specific CEMS-based intensities, and curtailment detection [38].

1.4 Research Objectives

This paper (and its codebase) explores the middle ground between Hawkes' low-complexity, empirical estimation and WattTime's hybrid, causally validated signals, and adds a practical feasibility test.

Specifically, we attempt to:

1. Retain simplicity and access: rely only on widely available public grid and weather data so the method is deployable across regions.

2. Pursue robustness with modest tooling: apply robust, lightweight statistical methods to marginal emissions estimates that remain reasonably accurate even with limited data.
3. Test practical value: develop a lightweight optimisation to identify emissions-reduction opportunities by shifting routine household use within modest time windows using these marginal signals.

The accompanying codebase produces (i) a half-hourly marginal-emissions series for India from January 2021–May 2025, and (ii) recommended household load-shift schedules for Delhi and Mumbai covering late 2021 to mid-2023, under plausible behavioural constraints and grid-aware limits.

2. DATA

2.1 Data scope and sources

This study uses public grid and weather data from 2020–2025, though only 2021–2025 are used for estimating and analysing time-varying emissions. Residential consumption covers late-2021 to mid-2023 for customers in Delhi and Mumbai.

Weather was drawn from ERA5-Land [26] and ERA5 [20]. ERA5-Land is used as the primary source with ERA5 filling gaps and adding the cloud cover variables which are not present in the ERA5-Land dataset. All variables collected are as follows: 2-meter temperature, surface net solar radiation, total cloud cover, high cloud cover, medium cloud cover, low cloud cover, total precipitation, wind speed, and wind direction. These variables are at an hourly resolution for forty-five latitude/longitude points around Delhi and Mumbai.

Grid data was taken from CarbonTracker [6], which records grid data at a five-minute resolution for all of India. This data includes total demand met, generation by source, and associated emissions.

Anonymised residential smart-meter data at half-hour resolution was provided by Tata Power through the Hitachi–Imperial collaboration and used as the basis for the scheduling proof-of-concept.

2.2 Cleaning and processing

ERA5-Land and ERA5 variables were harmonised to common units and names. Though ERA5-Land was primarily used when available, ERA5 was used to impute gaps. As several ERA5-Land fields are recorded as cumulative totals over the forecast period (e.g. surface solar radiation), these fields were converted (de-accumulated) to per-interval values by differencing consecutive timesteps. To place ERA5-Land and ERA5 on a common grid, variable-specific spatial re-gridding was applied. Precipitation and solar-radiation fields were remapped with conservative methods so that area-integrated totals are preserved. Temperature and wind components used bilinear interpolation, which is appropriate for smoothly varying fields. Cloud-cover variables used nearest-neighbour assignment to avoid creating spurious fractional values.

The CarbonTracker dataset required minimal cleaning, but did require gap filling. Short gaps (≤ 80 minutes) were filled using linear interpolation. Longer gaps (including multi-day outages) were reconstructed using an averaged neighbor-day gradient approach: for each missing interval we find the nearest previous and following calendar days that have data for this time slot, compute the gradients for each time slot, average them, and then apply these gradients either forward or backward based on the nearest real value available as an anchor.

2.3 Temporal alignment

All sources were aligned to a half-hourly grid in India Standard Time (IST) to match the meter data.

ERA5/ERA5-Land was resampled to 30-minute resolution: temperature and wind speed via linear midpoint interpolation; cloud cover via linear midpoint with values clamped to [0,1]; wind direction via speed-weighted circular interpolation; precipitation by evenly splitting hourly totals; and solar radiation with an energy-conserving, ramp-aware split (sunrise/sunset shoulders shaped; half-hour values sum to the original hourly totals; daily integrals preserved). Static fields were carried forward from the current hour.

CarbonTracker reports five-minute data with start-of-interval timestamps (e.g., 09:55 represents 09:55–10:00), whereas meter data are traditionally end-of-interval [13, 29]. To align conventions, we first shifted the grid series to interval-end. We then aggregated to half-hour intervals: rate variables (generation by source, demand met, average intensity) were averaged within each half-hour, and totals (t CO_2) were summed. Final timestamps were snapped to :00 and :30, yielding half-hour series that align exactly with the consumption data and with the emissions estimates used later.

2.4 Feature construction for modelling

All plausible drivers were first engineered into modelling-ready predictors using consistent transformations across demand, generation by source, weather, and time. From this full set, variables were then selected empirically by screening for low collinearity (pairwise correlations and variance-inflation checks) and by keeping those that improved fit and stability in the level model while preserving a smooth, well-behaved derivative. The final specification reflects that testing and prioritises transparency and portability. The exact selection may vary by grid and data quality (see results section), and substitutes are straightforward where a given feature is weak or unavailable.

Net load (demand_minus_renewables → demand_minus_renewables_std).
Net load is total demand minus non-dispatchable renewable generation (solar and wind in our dataset) and approximates the residual demand that dispatchable (usually thermal) units must cover. In most hours, renewables do not set the marginal unit; fossil generators do, so modeling against net load is preferred to capture driver of operational emissions [7, 11]. We standardise it (zero mean, unit variance) so the smoother operates on a stable scale and the derivative (our marginal signal) is numerically well behaved.

Weather smooths (temperature_celsius_std, wind_speed_mps_std, surface_net_solar_radiation_kWh_per_m2_log1p_std).
Temperature and wind speed affect both demand and supply. We standardise them so penalties treat each on a comparable scale. Solar irradiance is heavy-tailed and spikes around midday;

applying a log-transform (log1p) compresses those tails and reduces leverage from a few very bright hours. These three variables enter as smooth functions so the model can learn non-linear responses (for example, rising emissions on very hot evenings when cooling demand surges).

Wind direction context (wind_dir_sin, wind_dir_cos).
Wind direction is encoded with sine and cosine to avoid artificial jumps at 360/0 and allow the model to "fully capture the circular characteristics" as stated by Bentsen et. all [3].

Generation context (hydro_share_std).
Instead of using raw generation by fuel, we model the generation mix with fuel shares. Raw outputs co-move strongly with total demand, inflating collinearity. Though shares increase variance inflation factor scores amongst themselves, they reduce the VIF score against demand. Additionally, shares naturally encode relative mix as the sum of shares equals 1. All shares were standardized, and hydro retained as it consistently improved the rank-order accuracy without meaningfully sacrificing R^2.

Sky condition indicator (is_sunny).
A simple flag was derived from irradiance greater than 0 to separate daytime from nighttime conditions. This provides a regime switch that helps the model avoid attributing solar-driven swings to demand alone.

Time structure (hour_sin, hour_cos, doy_sin, doy_cos, is_weekend).
Daily and seasonal patterns are smooth and periodic. We encode hour-of-day and day-of-year with sine/cosine pairs to respect their circular structure, so 23:30→00:00 and 31 Dec→1 Jan are treated as adjacent. This avoids edge discontinuities and the need for many dummy variables. A weekend indicator captures systematic behavioural shifts (e.g., later evening peaks).

2.5 Data availability and reproducibility

All weather inputs are publicly available. ERA5-Land and ERA5 can be accessed without charge, and the workflow is portable to other public sources that expose similar variables and metadata. For example, Open-Meteo [28] provides access to similar variables that could be used to suitable for reproducing the feature set where ERA5 access is constrained.

For electricity and emissions, CarbonTracker provides the India-wide signals used here. The approach is intentionally portable because analogous feeds exist in other regions in near-identical formats: Electricity Maps (multi-region) [14], the UK Energy Dashboard [16], the International Energy Agency (IEA) global datasets [15], and the US Energy Information Administration (EIA) series for the United States [12]. Substituting any of these sources requires only minor renaming and unit checks, as the analysis operates on standard quantities (demand met, generation by source, and emissions or emission factors) at hourly or sub-hourly resolution.

Household consumption data (Delhi and Mumbai) were provided under agreement by Tata Power through the Hitachi–Imperial collaboration. These data are not public; however, the optimisation component depends only on a generic half-hourly load profile and thus remains reproducible with any comparable smart-meter dataset.

3. METHODS: Estimating marginal emissions

3.1 Objective and design

Our aim is twofold. First, to fit a national-level model that explains half-hourly system CO_2 as a smooth function of net load and weather/time covariates. Second, to obtain a time-varying marginal signal by taking the partial derivative of the fitted mapping with respect to net load, evaluated at the observed covariates. The result yields a national marginal emissions series suitable for operational guidance.

3.2 Candidate approaches

Work began from a binned-regression template [1]: observations were stratified by local weather (solar irradiance, wind speed), then emissions (tCO_2) were regressed on demand (MW) and demand^2, with hour and month fixed. Despite a high level-model R^2, the rank correlation with short-horizon slopes was modest.

We then engineered a richer, low-collinearity feature set (§2.4), ran ablations to isolate key drivers (see Appendix C – Tables M1 -M4), and tested Polynomial ridge , Huber regression, Groupwise OLS, and a penalised generalised additive model, using bootstrap checks for stability. The results (see §5.1 and Appendix C -Tables M5 and M7) pointed toward a model that is flexible where needed, yet penalised enough to keep the derivative smooth.

3.3 Final Estimator

We adopt a penalised generalised additive model (pyGAM) to model half-hourly CO_2 emissions. The model uses smooth functions of standardised net load , temperature, wind speed, and log-compressed solar radiation (denoted Q, T, W, S), plus linear context terms of wind direction (sin/cos), hydro-generation share, Fourier time controls (hour and day-of-year sin/cos), a weekend flag, and a sunny/not sunny flag. This model can be described in the following equation:

$$\mathrm{Y} = \beta_0 + f_Q(Q_{std}) + f_T(T_{std}) + f_W(W_{std}) + f_S(log(1 + S_{raw})_{std}) + x_{lin}^{\top}\beta + \varepsilon.$$

Each $f(\cdot)$ is a spline learned jointly with the others by penalised least squares. We use an $\boldsymbol{L2}$ curvature penalty (a penalty on the integrated second derivative ($\int (f''(u))^2 du$), which balances the fit of the model to the data with smoothness of the curve. This regularisation helps prevent overfitting and stabilises the first derivative $\partial\widehat{\mathrm{Y}}/\partial\mathrm{Q}$, which we interpret as the marginal emissions signal. The final model set 20 basis functions (splines) per smooth and set the penalty (λ) to 50 for each smooth. These settings were chosen via a small grid search that prioritised maximising marginal-signal alignment (Pearson/Spearman) with only minor trade-offs in level R^2.

3.4 From levels to marginal effects

After fitting the level model, the marginal series is the partial derivative of the fitted value with respect to raw net load, holding all other covariates at their observed values:

$$\widehat{ME_t} = \frac{\partial \hat{Y}}{\partial Q_{raw}} \Big|_{Q=Q_t\ ,\ X = X_t}.$$

Because the model is trained on the standardised net load $Q_{std} = \frac{Q_{raw} - \mu_Q}{\sigma_Q}$, we must also undo the scaling via the chain rule:

$$\frac{\partial \hat{Y}}{\partial Q_{raw}} = \frac{1}{\sigma_Q}\frac{\partial \hat{Y}}{\partial Q_{std}}.$$

3.5 Validation and calibration

After the model provides expected half-hourly CO2, and the derivative $\partial\hat{Y}/\partial Q$ yields the marginal signal, we can then evaluate whether the marginal signal produced matches the actual changes experienced by the system. To perform this evaluation, we construct ramp-pairs: consecutive half hour periods between t and t+1; and then compute the empirical slope (s_i) and the midpoint of the marginal estimate (ME_i^{mid}):

$$s_i = \frac{\Delta Y}{\Delta Q} = \frac{Y_{t+1} - Y_t}{Q_{t+1} - Q_t} \quad , \quad ME_i^{mid} = \frac{1}{2}\left(\widehat{ME}_t + \widehat{ME}_{t+1}\right).$$

To measure how close our estimates are to the observed change, we can fit a simple linear model:

$$s_i = a + bME_i^{mid} + \varepsilon_i.$$

If the model's marginal signal were perfectly scaled and unbiased, we would have $a \approx 0$ *and* $b \approx 1$ *so that* $s_i \approx ME_i^{mid}$. Since our model produced $a \approx -0.6$ *and* $b \approx 2.2$ (varying across ramp sizes ΔQ) calibration was needed. To determine which ramp size to use for our final calibration coefficients, we ran a sweep of ramp size thresholds on the validation set. At each threshold, the linear model described above was fit with weighted least squares (weights being ramp sizes). The configuration with the highest R^2 (0.29) and lowest RMSE at $\Delta Q \approx 4{,}255$ MW yielded $\hat{a} = -0.583$ *and* $\hat{b} = 2.226$. Finally, we then apply the transformation to the entire series, adjusting each value by the fitted intercept and slope:

$$\widehat{ME}_{cal} = \hat{a} + \hat{b}\widehat{ME}$$

Note that while this step will not change or improve the association as measured by the Pearson and Spearman correlations, it significantly reduces unit error by moving the OLS slopes towards 1.0 and the intercepts towards 0.0.

3.6 Confidence Labels and Measure Consolidation

To make the series usable with the raw meter-readings, we make two final adjustments. First, we attach a confidence label to each timestamp based on the out-of-sample ramp-pair Pearson r correlation in its $|\Delta Q|$ band: r < 0.40 → "low"; 0.40 ≤ r < 0.60 → "medium"; r ≥ 0.60→ "high" (none occur in our data). In analysis, these labels help identify whether any proposed shifts are using predictions with low or medium confidence in their directional accuracy. The second adjustment we make is a unit conversion from tons CO_2 per MW per 30-min to grams CO_2 per kWh to match the meter data. Since 1 MW sustained for 30 min equals 0.5 MWh, we multiply the $\widehat{ME}_{cal}$ value by 2,000 to make this adjustment.

4. METHODS: Emissions-Aware Load Shifting

4.1 Objective and solver choice

With a marginal emissions factor available for each half-hour, the next task was to build an optimiser that checks whether emissions savings are possible on a given day and identifies concrete shifts to realise them.

We prototyped three approaches in parallel: a continuous relaxer, a windowed linear program (LP), and a greedy scheduler, but we report results here for the greedy method only. Greedy was the first to succeed in development, runs quickly, has minimal dependencies, and proved robust across shards. While LP/continuous formulations can reach globally optimal allocations (whereas greedy is local), they were costlier to set up and debug within project timelines. Given the goal of a lightweight, reproducible tool, greedy is the main model profiled here.

4.2 Greedy scheduler: idea and core equation

A greedy search heuristic makes the single best feasible move available, updates all limits, and repeats until no improving move remains. In our setting, a "move" reallocates $q_{i,t\rightarrow s}$ kWh from a higher-MEF slot t to a lower-MEF slot s within a short time window $|\, t \rightarrow s| \leq W$. The model for this scheduler can be written as a minimisation or an equivalent maximisation problem:

$$\min_{q\,\geq\,0} \sum_{i\,\in\,I} \sum_{t} \sum_{s} m_s q_{i,t\,\rightarrow\,s} \qquad or \qquad \max_{q\geq0} \sum_{i\,\in\,I} \sum_{t} \sum_{s} (m_t\; - \; m_s) q_{i,t\,\rightarrow s}$$

where $i\; \in\; I$ indexes customers (households), $t, s\; \in\; T\; =\; \{0, \ldots, 47\}$ index half-hour slots in a day with t as the source time slot and s as the destination time slot, and $m_{s\; or\; t}$ is the marginal emission factor at that slot in grams CO_2/kWh. This project implements the maximisation view by at each step, scoring every feasible local move (t → s) by $\Delta CO_2\; =\; (m_t\; -\; m_s)\; \times\; q_{max}$, executing the best move, updating caps, and repeating until no improving move remains. These moves are also subject to constraints, described in the next section.

4.3 Constraints

To make the implementation more applicable to real world use cases, the search is restricted to operational and comfort limits. Perhaps the most important constraint is the conservation of energy constraint which enforces that each customer still uses the same amount of energy, but just moves the time of usage. This energy is restricted to each household and within each 24-hour period. The equation for this constraint can be written as a combination of slot level conservation:

$$\underbrace{x_{i,t}}_{after} \; = \; \underbrace{u_{i,t}}_{before} \; - \underbrace{\sum_{\substack{s\,\in\,T \\ s\,\neq\,t,\;\; |t-s|\,\leq\,W}}}_{outflow\; from\; t} q_{i,t\rightarrow s} \; + \underbrace{\sum_{\substack{r\,\in\,T \\ r\,\neq\,t,\;\; |r-t|\,\leq\,W}}}_{inflow\; into\; t} q_{i,r\rightarrow t} \qquad with \quad x_{i,t}\; \geq\; 0, \qquad \forall i \in I, \qquad t\; \in T$$

and daily conservation:

$$\sum_{t \in T} x_{i,t} \; = \sum_{t \in T} u_{i,t} \quad with\; \forall i \in I\;,$$

where $u_{i,t}$ is the original energy use in kWh for the customer i in slot t before any shifting, $x_{i,t}$ is the post-shift energy use for customer i in slot t after the optimisers moves, and r is a dummy time slot index for inbound flows.

The rest of the constraints implemented are illustrated in Table 1 below, and full equations are documented in the appendix. It is important to note the values below are used in the main experiment, but they could be modified to test alternative policies.

Constraint	Description	Value used
Maximum Shift Time Window	Limit the time window for each shift to avoid excessive disruption	±2 hours
Maximum Shifts per Customer per Day	Limit number of slot shifts per customer in a day to avoid significant disruption to their routine	1
Maximum Shifts per Customer per Week	Limit number of slot shifts per customer for a week to avoid significant disruption to their routine	3
Maximum Shift out of Peak Hours	Limit shifts that move usage out of peak hours to avoid disruption to routine	≤ 25% of baseline in peak hours
City destination anti-spike	Avoid crowding a single city half-hour	≤ 25% uplift vs baseline
Maximum Regional Shift Percentage per Day	Limit the maximum percentage of a region's total load that can be shifted in a day (city-scale)	≤ 10% of daily city load
Minimum Usage per Slot	Ensure a minimum level of usage in each time slot to represent always-on loads in a household	max{hour-of-day/day-of-week minimum, 10% × robust maximum, **ε**}

Table 1: Optimiser Constraints

5 . RESULTS

5.1 Model Scores

Although Ridge/Huber/OLS polynomials achieved slightly higher level fit (R^2) than GAM , that metric is insufficient for our purpose because marginal emissions (ME) = derivative of the level model. High R^2 does not guarantee a stable or well-scaled derivative. We therefore evaluate models on ramp-pair metrics: Pearson (linear agreement between predicted ME and realized slope $\Delta Y/\Delta Q$) and Spearman (rank ordering) [statstutor.ac.uk]. On 30-min pairs with $|\Delta Q| \geq 100$ MW, GAM delivers the most reliable ME signal (highest Pearson, comparable Spearman) with only a small trade-off in level R^2, so we select GAM as the production model.

Model Family	R^2	RMSE	MAE	Pearson	Spearman
PyGAM	0.9840	754.7759	607.8375	0.20034	0.43128
OLS	0.9866	690.2669	520.8337	0.19631	0.44248
Median	0.8628	2137.5000	1861.5000	0.19100	0.41400
Quantile	0.9366	1487.9000	1253.1000	0.18600	0.48200
Huber	0.9863	697.4679	526.2785	0.17774	0.41410
Ridge	0.9866	690.6372	521.7645	0.17766	0.41352

Table 2: Model Comparison

See Appendix C – Table M5 for full results

5.2 Influence of Ramp Size

The GAM's marginal signal strengthens as ramp magnitude grows. On validation the correlation between our predicted marginal emissions $\widehat{ME}$ and realised slopes $s = \Delta Y/\Delta Q$ rises with |ΔQ|: small ramps (~P10) give Pearson $r \approx 0.36$, Spearman $\rho \approx 0.46$; around the median ramp (|ΔQ| = 2,330 MW) $r \approx 0.53$, $\rho \approx 0.56$. See the figures below for illustration.

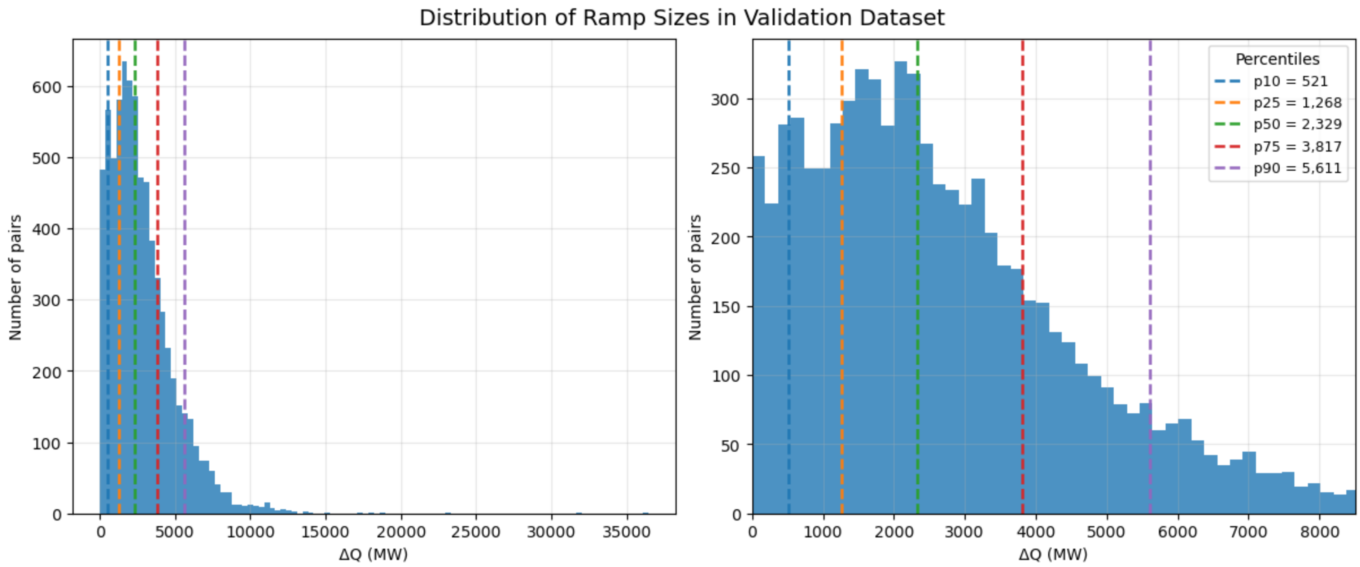

Figure 1: Ramp Sizes Distribution
See Appendix D – Figures M1-3 for full results

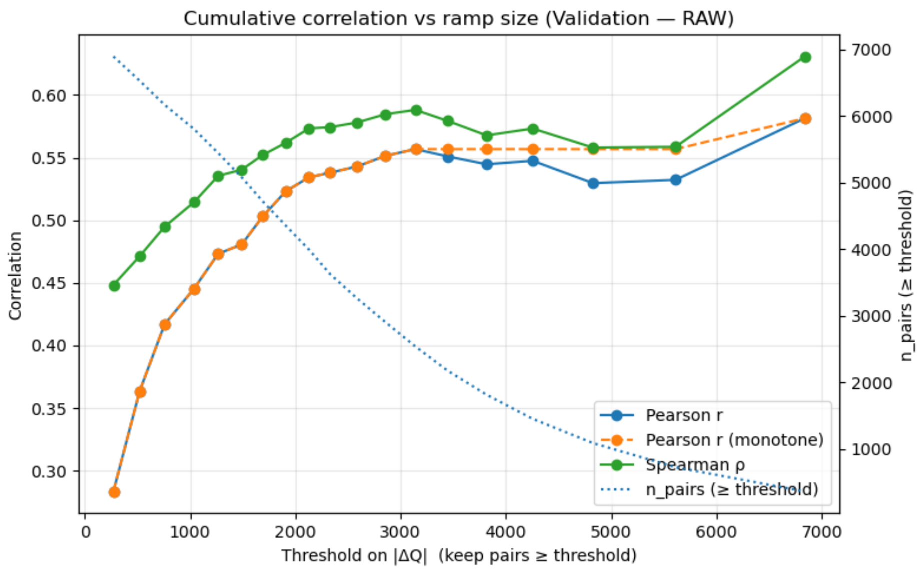

Figure 2: Correlation and Ramp Sizes
See Appendix D – Figures M4 and M5 for full results

5.3 Unit Alignment

Even with good correlation, raw ME magnitudes were under-scaled: regressing s on $\widehat{ME}$ over validation yielded slopes >1 and a small negative intercept. We therefore fit a single linear calibration on validation pairs (weighted by |ΔQ|) in the operationally strong ramp region (≈P80–P90, |ΔQ|≈4.255 GW). This yielded a≈−0.58, b≈2.23, and after applying this calibration to the dataset, the OLS line moves near the identity (slope ≈ 1, intercept ≈ 0) across ramp bins, and distance metrics improve materially (see Appendix C – Tables M8-9). The figure below illustrates how the data transforms, and it is important to note the Spearman value is unchanged after this alignment as it is a simple linear transformation that makes the $\widehat{ME}$ more directly interpretable.

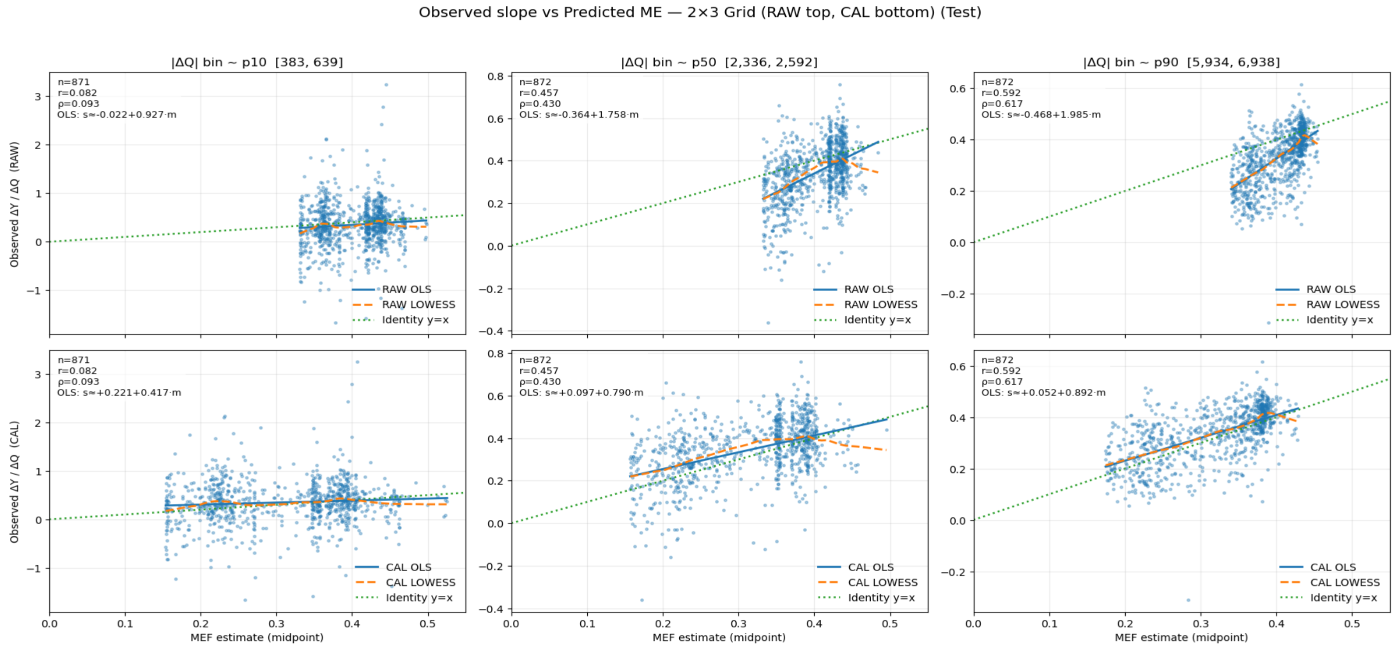

Figure 3: Calibration of Estimates

See Appendix D – Figures M6-10 for full results

5.4 Confidence Labels

To translate the diagnostics into user-facing guidance, we assign confidence to each timestamp based on the expected Pearson r at its ramp magnitude (estimated from the validation curves). Ramps with Pearson scores below 0.4 are assigned 'low' confidence and those between 0.4 and 0.6 are assigned 'medium' confidence. No scores above 0.6 were observed to justify a 'high' confidence band. These cut points are more conservative than the widely used correlation effect-size conventions (small/medium/large around 0.1/0.3/0.5 [8], but still manage to capture most of our timestamps as shown in figure 4 below.

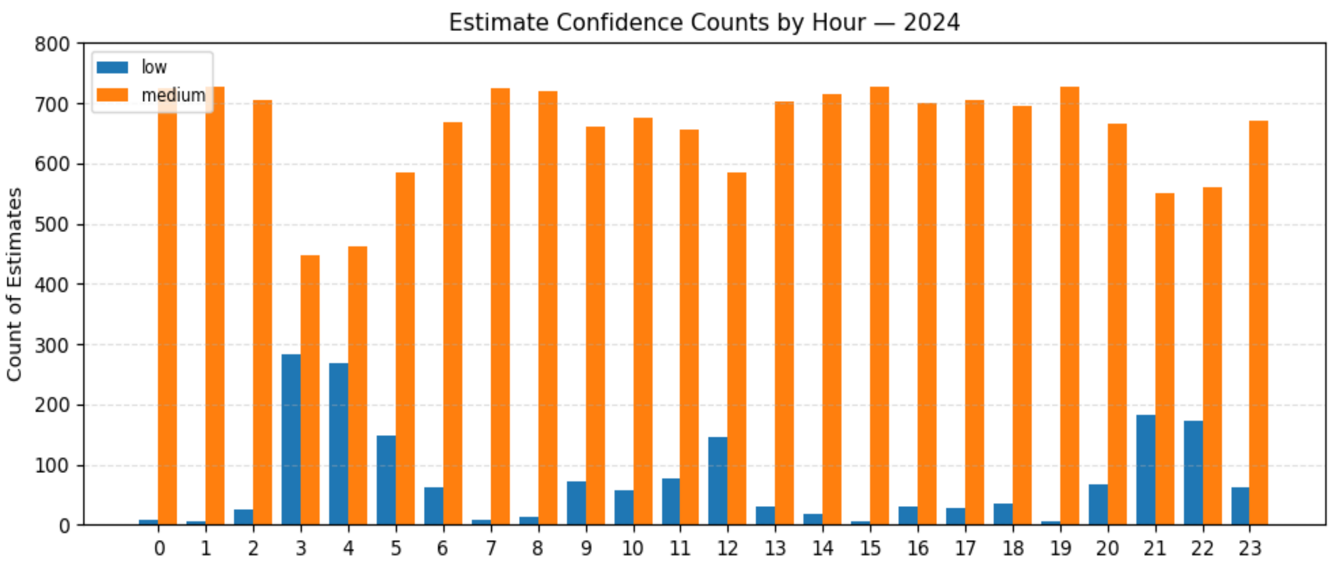

Figure 4: Confidence Bands of Emissions Estimates

See Appendix D – Figures M15-16, and Appendix C – Table O5 for more detail

5.5 Distribution of marginal factors over time

Year-by-year distributions of calibrated ME show a widening range from 2021→2025. Median ME drifts modestly down (consistent with cleaner energy), but the spread increases, reflecting a more variable system as solar/wind penetration rises. This is expected: average emissions can fall even as marginal emissions become more volatile if low-carbon supply saturates during certain hours and fossil units set the margin for increments/decrements [34].

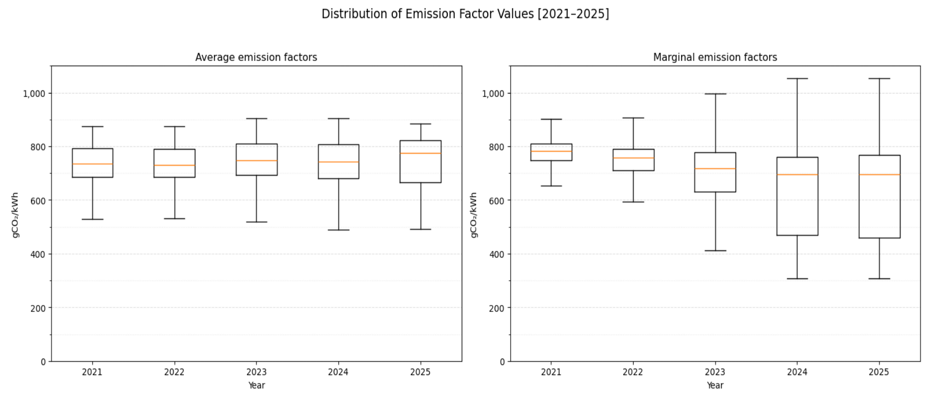

Figure 5: Boxplot Distribution of Emission Factors, 2021-2025

See Appendix D – Figures M12-14 and S8-11 and Appendix C – Table S1 and S2 for more detail

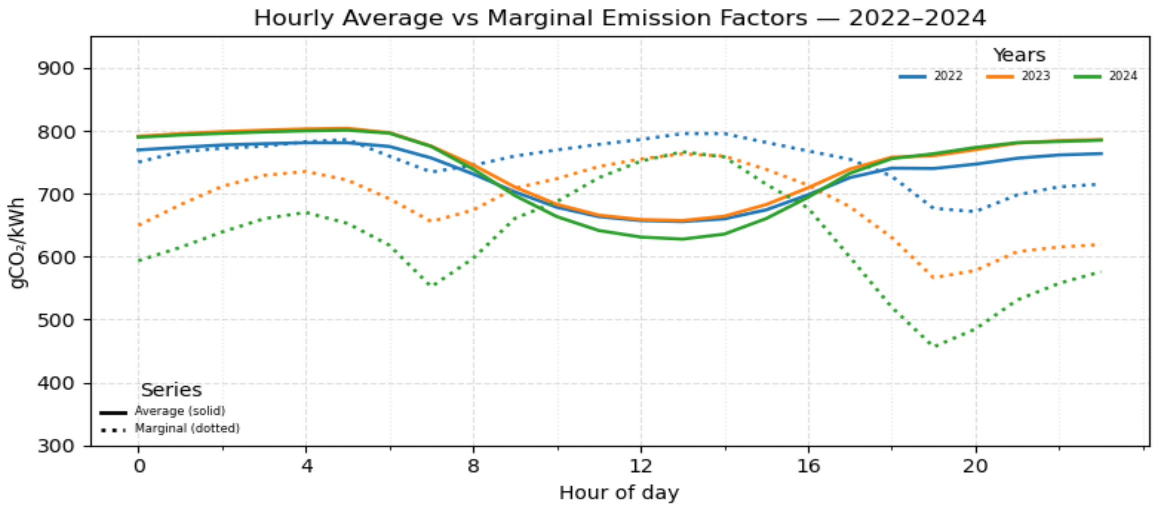

Figure 6: Hourly Marginal (calibrated) and Average Emission Factors, 2022-2024

5.6 Scheduler outcomes (household vs aggregate)

The scheduler respected conservative behavioral and technical constraints, but the results were not groundbreaking. Per-household impact is small. The average action moved about 0.40 kWh (roughly one efficient laundry cycle or several kettle boils [10, 21], so the annual avoided CO_2 per home is only a fraction of a percent of a typical household footprint. Illustrative stats below (all shifts, 30-min slots):

Region	Move Count	Average CO_2 per move (g)	Median CO_2 per move (g)	Average kWh per move	Median kWh per move
All	14,996,536	44.1995	26.1329	0.4286	0.3230
Delhi	12,394,970	44.3293	25.9728	0.4191	0.3185
Mumbai	2,601,566	43.5812	26.7853	0.4738	0.3425

Table 3: Shifts and Savings by Region
See Appendix C – Table O4-7 for full results

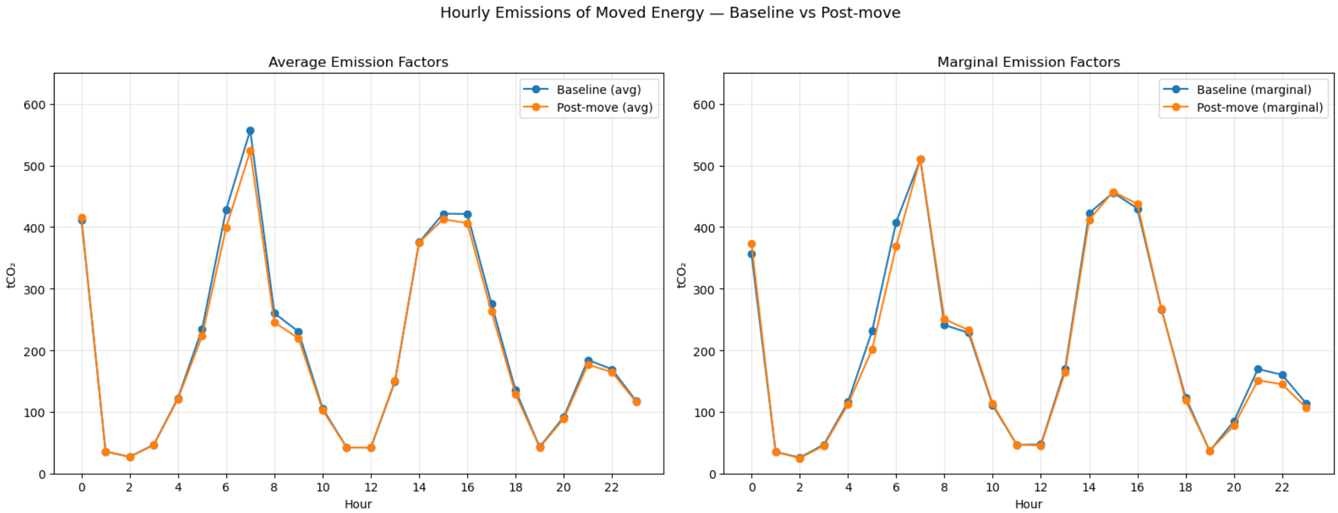

Figure 7: Hourly Emissions Profile pre- and post-Optimisation Moves
See Appendix D – Figure O1 and Appendix C – Tables O4-9 for more detail

Though no extreme opportunities for savings were uncovered, comparisons of the same household schedule with emissions calculated with average emission factors versus calibrated marginal emission factors highlighted the differences between the two. Where AEF-based calculations yielded savings of 160 t CO_2, MEF-based calculations yielded 663 t CO_2, 313% higher. See Appendix C - Table O9.

6. DISCUSSION

6.1 Interpreting the varying MEFs

A persistent feature of the results is higher MEFs than AEFs at midday/early-afternoon, even as AEFs dip (Figure 6). A plausible explanation is renewable saturation: PV (and some wind) already run near their instantaneous limits in these hours, so incremental load is met by coal/gas units with higher marginal rates; unit-commitment and ramp limits amplify this effect [33]. Hydro share enters as a consistent (likely negative) predictor of marginal CO_2 across model families. We interpret this as an association rather than causation: India's reservoir hydro is dispatchable and its operation (releases/peaking) likely co-moves with ramps, renewable availability, and seasonal inflows [22, 24, 27]. But without plant-level dispatch or reservoir data, we cannot identify the causal pathway. The literature is mixed on whether MEFs are typically above or below AEFs, with outcomes depending on fuel mix and operating regime [19, 34]. India's coal-heavy system with fast-growing solar fits squarely within those mixed findings.

Year-by-year, the calibrated MEF distribution widens from 2021→2025, while its median drifts modestly down (Figure 5) i.e., cleaner on average yet more variable at the margin as renewables saturate in some hours and fossil units set the incremental response—exactly what MEFs are meant to capture.

6.2 Pearson, Spearman, and calibration

R^2 alone isn't sufficient for our purpose because MEF is the derivative of the level model—great level fit can still yield a noisy or mis-scaled derivative. We therefore evaluate on ramp-pair diagnostics: Spearman (do we rank ramps correctly?) and Pearson (do magnitudes line up?). Ramp size matters: on validation, small ramps (~P10) give $r \approx 0.36$, $\boldsymbol{\rho} \approx 0.46$; around the median ($|\Delta Q| \approx 2.3$ GW) $r \approx 0.53$, $\boldsymbol{\rho} \approx 0.56$; top-decile ramps ($\gtrsim 5.6$ GW) reach ~0.5–0.6. Raw ME was systematically under-scaled: regressing realized slopes ($\Delta Y/\Delta Q$) on ME yielded slopes >1 and a small negative intercept. A simple linear calibration on validation pairs (WLS, weight $|\Delta Q|$, using the P80–P90 ramp region) produced $a \approx -0.58$ and $b \approx 2.23$; applying s_cal = a + b·ME moves the fit close to the identity line (slope≈1, intercept≈0) and improves distance metrics. Because calibration is linear, Spearman is unchanged; the goal is unit alignment, not re-ranking.

6.3 Practical implications (households and. system)

With conservative behavioral/technical constraints, per-home impact is small. The average action shifted ~0.40 kWh (roughly one efficient laundry cycle or several kettle boils), so annual avoided CO_2 per home is <1% of a typical footprint (Appendix C – Table O8). This is not a compelling standalone lever unless there are regulatory programs where every kilogram counts. We also note most chosen moves were ≥ 90 minutes from the original slot (within the ±2 h window), suggesting the algorithm consistently found cleaner windows just beyond the immediate neighborhood.

Even when per-home gains are small, MEF-aware scheduling remains informative at scale and for program design (when to ask, whom to target, how aggressively to relax constraints for opt-in cohorts). Additionally, using calibrated MEFs instead of AEFs changes the arithmetic: on the same schedule, MEF-based avoided CO_2 was +313% relative to AEF-based calculation (Appendix C – Table O9). Depending on the emission factor used, you can severely under- or over-state impact.

6.4 Limits and Next Steps

We kept the method lean and auditable by design, which brings clear limits: weather inputs come from two metros while the target is national (blurring local dynamics, especially on small ramps); utility operations we don't observe (outages, curtailment, imports/exports, fuel constraints) add noise; a single linear calibration aligns units but can't fix structural bias; and the greedy scheduler is transparent but not globally optimal—regional caps didn't bind here because we modeled a subset of customers, so a real roll-out must couple to true system caps to avoid crowding. Next steps could include: adding richer utility signals and broader weather coverage; allowing varying-coefficient effects (e.g., ME varying with hydro), plus holiday/heat-wave flags and light day-ahead forecasts; replacing the greedy heuristic with a globally aware solver (with fairness/crowding controls) and testing wider opt-in shift windows; and validating externally against third-party MEFs.

6.5 Takeaways

The calibrated GAM provides a directionally correct, operationally useful MEF signal. It tracks short-horizon system movements, ranks hours reliably (especially at normal-to-large ramps), and—after a simple calibration—expresses marginal effects in usable units. It's best used to prioritize hours and guide small shifts, not as a substitute for operator-grade inventories. The midday finding (MEF > AEF while AEF dips) is plausible in a coal-heavy system with growing solar and aligns with prior evidence. Given today's constraints, bigger wins likely lie upstream—hydro dispatch, transmission, curtailment reduction, and commitment strategies that minimize marginal, not just average, emissions.

7. CONCLUSION

We set out to turn public grid and weather data into actionable guidance. The work delivers (i) a portable estimator whose derivative with respect to net load serves as a robust marginal-emissions (MEF) signal, (ii) a simple linear calibration that aligns that derivative to realized short-horizon changes, and (iii) a constraint-aware scheduler that translates the signal into small, acceptable household shifts. A GAM produced the most reliable MEF for operational ramps; correlations strengthen with ramp size, and calibration corrected scale without changing ranks. Ramp-based confidence labels make reliability explicit—most hours fall in the medium band—so the series is best for ranking rather than meter-grade accounting. Midday MEFs often exceed AEFs even as AEFs dip, and the MEF distribution widens from 2021–2025 while medians drift down.

Operationally, per-home effects are small under conservative constraints (~0.4 kWh moves within ±2 h), but the scheduler consistently finds cleaner windows, and using MEFs instead of AEFs materially changes avoided-CO_2 (+313% on the same schedule). Bigger gains likely sit upstream—hydro dispatch, curtailment reduction, transmission, and commitment strategy. The framework is clear, reproducible, and portable; with regionalisation, utility signals, and day-ahead forecasts, it can scale while remaining auditable.

REFERENCES

[1] Aditya Tomar. (2024). *margE_India.Rmd* [R]. Imperial College London. https://github.com/ese-ada-lovelace-2024/irp-dbk24/blob/main/emission_rate_methodologies/marginal%20emissions/margE_India.Rmd

[2] Ajay Shankar, A K Saxena, & Taruna Idnani. (2025). *Roadmap to India's 2030 Decarbonization Target*. The Energy and Resources Institute. https://www.energy-transitions.org/publications/roadmap-to-indias-2030-decarbonization-target/

[3] Bentsen, L. Ø., Warakagoda, N. D., Stenbro, R., & Engelstad, P. (2023). Spatio-temporal wind speed forecasting using graph networks and novel Transformer architectures. *Applied Energy*, *333*, 120565. https://doi.org/10.1016/j.apenergy.2022.120565

[4] Bracht, F., Khanna, S., Martin, R., & Muûls, M. (2025). *Waiting for carbon reductions—A smart IV approach*.

[5] Buchholz, K. (2022, November 9). *Infographic: India's Electricity Mix*. Statista Daily Data. https://www.statista.com/chart/27963/india-energy-mix

[6] Center for Social and Economic Progress. (2020, September 22). *The CSEP Electricity and Carbon Tracker*. CSEP Electricity and Carbon Tracker. https://carbontracker.in/ (Accessed on 14-07-2025)

[7] Clean Energy Buyers Institute. (2022). *GUIDE TO SOURCING MARGINAL EMISSIONS FACTOR DATA* (Next Generation Carbon-Free Electricity Procurement Activation Guide) Clean Energy Buyers Institute. https://cebi.org/wp-content/uploads/2022/11/Guide-to-Sourcing-Marginal-Emissions-Factor-Data.pdf

[8] Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (2nd edn). Routledge. https://doi.org/10.4324/9780203771587

[9] Council on Foreign Relations. (2024, December 5). *How Electrification Can Reduce Emissions*. CFR Education from the Council on Foreign Relations. https://education.cfr.org/learn/reading/electrification-climate-change

[10] Dhesi, Ben (2023, January 18). *How much does it cost to boil a kettle?*. Hugo. https://hugoenergyapp.co.uk/cost-to-boil-a-kettle/

[11] EIA. (2023, June 21). *As solar capacity grows, duck curves are getting deeper in California—U.S. Energy Information Administration* (EIA) [Governmental]. Today in Energy. https://www.eia.gov/todayinenergy/detail.php?id=56880

[12] EIA. (n.d.). *Hourly Electric Grid Monitor: U.S. Electricity Overview (U.S. Lower 48)* [Governmental]. U.S. Energy Information Administrtion Independen Statistics and Analysis. Retrieved 27 August 2025, from https://www.eia.gov/electricity/gridmonitor/index.php

[13] Elecenergy. (2013). *Interval Energy Data, or Itemized Energy Data?: Understanding the Benefits of Analyzing & Improving a Buildings Energy Usage & Cost Profile using one or both techniques*. https://elecenergy.co.uk/assets/white-paper-interval-energy-data-or-itemized-energy-data.pdf

[14] ElectricityMaps. (2025). *Carbon Intensity* (Versions 2025-01-27) [Dataset]. https://github.com/electricitymaps/electricitymaps-contrib/tree/master. https://www.electricitymaps.com

[15] IEA. (2025, August 13). *Real-Time Electricity Tracker – Data Tools—IEA*. Real-Time Electricity Tracker. https://www.iea.org/data-and-statistics/data-tools/real-time-electricity-tracker?from=2025-7-23&to=2025-8-22&category=demand

[16] Energy Dashboard. (2025, August 12). *Live Generation, Demand and Emissions*. Live UK Electricity Generation, Carbon Intensity & Demand – Energy Dashboard. https://www.energydashboard.co.uk/live

[17] Fleschutz, M., Bohlayer, M., Braun, M., & Murphy, M. D. (2022). Demand Response Analysis Framework (DRAF): An Open-Source Multi-Objective Decision Support Tool for Decarbonizing Local Multi-Energy Systems. *Sustainability*, *14*(13). https://doi.org/10.3390/su14138025

[18] Green, A. (2024, December 5). *Average versus Marginal Carbon Emissions: Many professionals are getting this wrong*. [Personal website]. ADG Efficiency. https://adgefficiency.com/energy-basics-average-vs-marginal-carbon-emissions/

[19] Hawkes, A. D. (2010). Estimating marginal CO_2 emissions rates for national electricity systems. *Energy Policy*, 38(10), 5977–5987. https://doi.org/10.1016/j.enpol.2010.05.053

[20] Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2023): ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.adbb2d47 (Accessed on 16-07-2025)

[21] *How Many Watts Does an Electric Kettle Use? (*2023, August 2). BLUETTI-PH. https://www.bluettipower.ph/blogs/news/how-many-watts-does-an-electric-kettle-use

[22] *IHA Sediment Management Case Studies: India—Teesta V*. (2025, August 27). [Organisational]. International Hydropower Association. https://www.hydropower.org/sediment-management-case-studies/india-teesta-v

[23] International Energy Agency (IEA). (2025, February). *Emissions – Electricity 2025 – Analysis*. IEA. https://www.iea.org/reports/electricity-2025/emissions

[24] LotusArise. (2024, February 17). *Indian Geography Mapping: Hydroelectric Power Plants in India* [Organisational]. *LotusArise*. https://lotusarise.com/hydroelectric-power-plants-in-india-upsc/

[25] Manembu, P., Kewo, A., Bramstoft, R., & Nielsen, P. S. (2023). *A Review of Residential Electricity Load Shifting at the Appliance Level*. Engineering. https://doi.org/10.20944/preprints202310.0908.v1

[26] Muñoz Sabater, J. (2019): ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.e2161bac (Accessed on 14-07-2025)

[27] Naga Gangadhar. (2024, December 12). Top 7 Largest Hydroelectric Power Plants In India (2025) [Organisational]. *Blackridge Research & Consulting*. https://www.blackridgeresearch.com/blog/top-seven-hydroelectric-power-plants-in-india

[28] Open-Meteo.com. (n.d.). *Free Open-Source Weather API | Open-Meteo.com* [Organisational]. Open-Meteo.Com. Retrieved 27 August 2025, from https://open-meteo.com/

[29] Oracle. (n.d.). *Interval Data File Specifications* [Organisational]. Data Standards Documentation. Retrieved 27 August 2025, from https://docs.oracle.com/en/industries/energy-water/opower-platform/data-transfer/interval_AMI_spec/DataFileSpecifications_AMI_.htm

[30] Ritchie, H., Roser, M., & Rosado, P. (2020). Renewable Energy. *Our World in Data*. https://ourworldindata.org/renewable-energy

[31] Roser, M. (2020). Why did renewables become so cheap so fast? *Our World in Data*. https://ourworldindata.org/cheap-renewables-growth

[32] Scott Cunningham. (2021). 7 Instrumental Variables. *In Causal Inference: The Mixtape*. Yale University Press. https://mixtape.scunning.com/07-instrumental_variables

[33] Sengupta, S., Spencer, T., Rodrigues, N., Pachouri, R., Thakare, S., Adams, P. J., Tongia, R., & Azevedo, I. M. L. (2022). Current and Future Estimates of Marginal Emission Factors for Indian Power Generation. *Environmental Science & Technology*, *56*(13), 9237–9250. https://doi.org/10.1021/acs.est.1c07500

[34] Siler-Evans, K., Azevedo, I. L., & Morgan, M. G. (2012). Marginal Emissions Factors for the U.S. Electricity System. Environmental Science & Technology, 46(9), 4742–4748. https://doi.org/10.1021/es300145v

[35] statstutor.ac.uk. (2011, December 16). *Pearson's correlation*. statstutor.ac.uk. https://www.statstutor.ac.uk/resources/uploaded/pearsons.pdf

[36] Vikram Singh & Kingsmill Bond. (2024). *Powering Up the Global South: The Cleantech Path to Growth*. Rocky Mountain Institute (RMI). https://rmi.org/insight/powering-up-the-global-south/

[37] WattTime. (2022). *Marginal Emissions Modeling: WattTime's approach to modeling and validation*. WattTime. https://watttime.org/wp-content/uploads/2023/11/WattTime-MOER-modeling-20221004.pdf

[38] WattTime. (n.d.). *Average vs. Marginal Emissions* [Organisational]. WattTime. Retrieved 30 May 2025, from https://watttime.org/data-science/data-signals/average-vs-marginal/

Appendix A – Disclosures and Acknowledgements

i. Use of Artificial Intelligence (AI).

I acknowledge that ChatGPT was used throughout the course of this project when debugging code, clarifying subject matter related to the energy field and climate data, assisting in learning the python libraries of polars and SQLAlchemy, how to effectively use High Performance Computing resources through parallelisation and scheduling, and generally brainstorming to help guide the project direction. Specific models used include GPT-5, GPT-4o, and GPT-4o mini (OpenAI, https://chatgpt.com/) . I confirm that no content generated by this tool has been presented as my own work.

I acknowledge that DeepSeek (deepseek, https://chat.deepseek.com/) was used throughout the course of this project when debugging code, particularly that relating to shell scripts and the usage of the PBS scheduler for Imperial's High Performance Computing Centre. I confirm that no content generated by this tool has been presented as my own work.

I acknowledge that Github Copilot (Github, https://github.com/features/copilot) was used throughout the course of this project to assist in the development of code, primarily through its autocomplete functionality. Specifically Copilot was used to draft the docstrings for functions. I confirm that no content generated by this tool has been presented as my own work.

I acknowledge that NotebookLM (Google, https://notebooklm.google/) was used when researching to explore new literature, and summarise topics in order to prioritise which resources to focus on. I confirm that no content generated by this tool has been presented as my own work.

ii. Data Access

I would like to thank the Hitachi-Imperial Centre for Decarbonisation and Natural Climate Solutions and the Data Science Institute at Imperial College London for providing access to the data. Special thanks is needed to Brython Caley-Davies who provided a database for easier access to cleaned data.

Appendix B – Project Evolution vs Plan

The core objective—quantifying emissions-reduction potential from temporal load shifting—remained unchanged. However, several sub-objectives shifted:

- Dropped PCA/K-means dimensionality reduction. After building robust ETL for large data (Polars, HPC) and stabilising the MEF pipeline, PCA added limited value for nine core predictors; simpler collinearity checks and selective pruning were sufficient
- Deferred household demand forecasting. Given time constraints and mature literature, bespoke forecasting contributed less to the primary objective than MEF estimation and optimisation; the MEF task itself required a forecasting-like level model.
- Abandoned dispatch prototype. Prioritised a portable, public-data method usable across regions; a dispatch layer is future work.

Appendix C – Results Tables

Note to reader: appendix C and D contain tables and Figures related to the marginal emisssions factor estimation, the optimisation, and supplementary information. The subject that a Figure or table belongs to is denoted as M for MEF estiamation, O for opimisation, and S for supplementary information.

Table M1. Testing Scaling Methods

Q value	Degree	Scaling Method	Alpha	R^2	RMSE	Pearson	Spearman
demand_met	4	standard	0.001	0.9952	409.6967	0.03263	0.16034
demand_met	4	robust	0.001	0.9952	409.6967	0.03263	0.16034
demand_met	3	robust	30.000	0.9952	409.6417	0.03248	0.16014
demand_met	3	standard	0.001	0.9952	409.6441	0.03247	0.16014
demand_met	2	robust	30.000	0.9953	407.6685	0.03130	0.15921
demand_met	2	standard	0.001	0.9953	407.7014	0.03130	0.15921
demand_met	1	robust	30.000	0.9952	412.6967	0.03095	0.15787
demand_met	1	standard	0.001	0.9952	412.7413	0.03095	0.15787
demand_met	5	standard	30.000	0.9952	410.5312	0.03000	0.15932
demand_met	5	robust	30.000	0.9952	410.6298	0.02999	0.15932

Notes for Table M1. Testing Scaling Methods	
Variables tested:	surface_net_solar_radiation_kWh_per_m2_log1p, wind_speed_mps, thermal_share, gas_share, hydro_share, nuclear_share,
Other Notes	To test whether standard scaling or robust scaling would be a better option for our variables set, a ridge model of varying degrees was run on the same variable set with different processing to observe any potential differences. The table below illustrates that no meaningful differences were observed. As a result, standard scaling was kept for simplicity in following models and testing. Results show validation set metrics for ΔQ ≥ 100 (MW) and are sorted by Pearson score.

Table M2. Q as Demand Met Minus Renewables v Demand Met

Q value	Scaling Method	Degree	Alpha	R^2	RMSE	Pearson	Spearman
demand_met_minus_renewables	standard	2	30.0	0.9910	565.6546	0.12952	0.29520
demand_met	standard	4	0.001	0.9952	409.6967	0.03263	0.16034
demand_met	robust	4	0.001	0.9952	409.6967	0.03263	0.16034
demand_met	robust	2	30.0	0.9953	407.6685	0.03130	0.15921
demand_met	standard	2	0.001	0.9953	407.7014	0.03130	0.15921
demand_met_minus_renewables	standard	4	10.0	0.9899	598.0801	0.00143	0.07091

Notes for Table M2. Q as Demand Met Minus Renewables v Demand Met	
Variables tested:	surface_net_solar_radiation_kWh_per_m2_log1p, wind_speed_mps, thermal_share, gas_share, hydro_share, nuclear_share,
Other Notes	To test whether a q value of demand met or demand met minus renewables was a better predictor, additional models were run testing on degrees 2 and 4, with standard scaling, and auto-selected best alpha. The table illustrates that when using demand_met_minus renewables there are large variations depending on the degree used, but that degree 2 yields notable gains. Results show validation set metrics for ΔQ ≥ 100 (MW) and are sorted by Pearson score.

Table M3. Varying Share Generation Predictors

Model, Degree, Alpha	Shares included in model	R^2	RMSE	MAE	Pearson	Spearman
Ridge, 2, 100	hydro	0.9866	690.1893	521.4062	0.17772	0.41359
Ridge, 2, 100	hydro, nuclear	0.9881	650.8535	496.8325	0.17283	0.39630
Ridge, 2, 100	hydro, renewable	0.9869	682.8594	526.3622	0.17084	0.39085
Ridge, 2, 100	gas, hydro	0.9893	616.8194	476.3836	0.16391	0.39735
Ridge, 2, 100	hydro, nuclear, renewable	0.9901	593.1144	467.8615	0.16196	0.35777
Ridge, 2, 100	gas, hydro, renewable	0.9896	608.0846	485.1469	0.16126	0.37851
Ridge, 2, 100	thermal, hydro	0.9872	672.9140	529.2815	0.15983	0.36324
Ridge, 2, 100	thermal, gas hydro	0.9894	612.8507	489.0780	0.15463	0.36149
Ridge, 2, 10	(all)	0.9917	543.3570	424.3982	0.13093	0.29910
Ridge, 2, 100	*(none)*	0.8361	2415.7196	2073.3357	0.08380	0.19672

Notes for Table M3. Varying Share Generation Predictors	
Constant Variables:	***Weather:*** surface_net_solar_radiation_kWh_per_m2_log1p_std, wind_speed_mps_std, temperature_celsius_std, wind_dir_sin, wind_dir_cos, is_sunny ***Time features:*** doy_sin, doy_cos, hour_sin, hour_cos, is_weekend
Varied Variables	***Generation features****:* thermal_share_std, gas_share_std, hydro_share_std, nuclear_share_std, renewable_share_std,
Other Notes	To determine the best set of predictors to use, models were run varying the generation share variables included. The table illustrates that hydro yields the best Pearson score. As a result – this variable was kept in future models. Ad hoc analysis was run to determine the best set of weather predictors and yielded the set described above. Results show validation set for ΔQ ≥ 100 (MW) and are sorted by Pearson score.

Table M4. Initial Pearson and Spearman Scores over varying ΔQ ranges

Model, Degree, Alpha	Ramp Pair Window	Percentile	Ramp pairs	Minimum ΔQ (MW)	Pearson	Spearman
Ridge, 2, 30	30min	0.01	14,358	52.6691	0.15564	0.40841
Ridge, 2, 30	30min	0.10	13,054	521.2500	0.34059	0.45428
Ridge, 2, 30	30min	0.25	10,878	1267.8125	0.44144	0.51176
Ridge, 2, 30	30min	0.50	7,252	2328.8333	0.49735	0.53524
Ridge, 2, 30	30min	0.70	4,352	3445.4166	0.50559	0.52455
Ridge, 2, 30	30min	0.80	2,902	4255.2500	0.49679	0.51267
Ridge, 2, 30	30min	0.90	1,452	5611.7500	0.49357	0.49866
Ridge, 2, 30	30min	0.95	726	6841.1708	0.58566	0.60154
Ridge, 2, 30	30min	0.99	148	10127.1666	0.46160	0.48051

Notes for Table M4. Initial Pearson and Spearman Scores over varying ΔQ ranges	
Variables Included:	***Weather:*** surface_net_solar_radiation_kWh_per_m2_log1p_std, wind_speed_mps_std, temperature_celsius_std, wind_dir_sin, wind_dir_cos, is_sunny ***Time features:*** doy_sin, doy_cos, hour_sin, hour_cos, is_weekend ***Generation features:*** hydro_share_std
Other Notes	This table validates our assumption that as ramp sizes increase, the pearson and spearman scores increase. Though run on the Ridge model, the results present hold true across all models.

Table M5. Model Comparison

Model Family	Q value	Settings	R^2	RMSE	MAE	Pearson	Spearman
PyGAM	demand_met_minus_-renew-ables_std	{splines: 20, lambda: 50}	0.9840	754.7759	607.8375	0.20034	0.43128
OLS	demand_met_minus_-renew-ables_std	{deg: 2}	0.9866	690.2669	520.8337	0.19631	0.44248
*Median	demand_met_minus_-renewables_std & demand_met_minus_-renewables_std^2	{bin_variables: temperature, wind_speed, hydro, min_group_size: 20}	0.8628	2137.5000	1861.5000	0.19100	0.41400
*Quantile	demand_met_minus_-renewables_std & demand_met_minus_-renewables_std^2	{bin_variables: temperature, wind_speed, hydro, bins: 3, min_group_size: 20}	0.9366	1487.9000	1253.1000	0.18600	0.48200
Huber	demand_met_minus_-renew-ables_std	{deg: 2, alpha: 0.10, epsilon: 1.75}	0.9863	697.4679	526.2785	0.17774	0.41410
Ridge	demand_met_minus_-renewables_std	{deg: 2, alpha: 30}	0.9866	690.6372	521.7645	0.17766	0.41352
Old Median	demand_met & demand_met^2	{bin_variables: surface_net_solar_radiation-_kWh_per_m2, wind_speed, min_group_size: 10}	0.7626	2907.1000	2339.3000	0.02400	0.13900
Old Quantile	demand_met & demand_met^2	{bin_variables: surface_net_solar_radiation-_kWh_per_m2, wind_speed, bins: 5, min_group_size: 10}	0.7580	2935.3000	2334.7000	0.00500	0.11200

Notes for Table M5. Model Comparison	
Core Variables Included:	***Weather:*** surface_net_solar_radiation_kWh_per_m2_log1p_std, wind_speed_mps_std, temperature_celsius_std, wind_dir_sin, wind_dir_cos, is_sunny ***Time features:*** doy_sin, doy_cos, hour_sin, hour_cos, is_weekend ***Generation features:*** hydro_share_std
*Model Variables	For these models, the sin/cos encoded time variables were not available, so simple doy and hour encodings were used.
Old Model Variables	For these models, the variables used were different from those mentioned above and not undergo scaling or other processing before being used. Time features: month, hour Weather: surface_net_solar_radiation_kWh_per_m2, wind_speed_mps
Other Notes	This table compares the best of each model family developed. Results show validation set for ΔQ ≥ 100 and are sorted by Pearson score.

Table M6. PyGAM Pearson and Spearman Scores over varying ΔQ ranges

Model	Ramp Pair Window	Percentile	Ramp pairs	Minimum ΔQ (MW)	Pearson	Spearman
PyGAM	30min	0.01	14,358	52.669	0.17140	0.42510
PyGAM	30min	0.10	13,054	521.250	0.36342	0.47117
PyGAM	30min	0.25	10,878	1267.812	0.47319	0.53541
PyGAM	30min	0.50	7,252	2328.833	0.53789	0.57429
PyGAM	30min	0.70	4,352	3445.416	0.55096	0.57935
PyGAM	30min	0.80	2,902	4255.250	0.54742	0.57318
PyGAM	30min	0.90	1,452	5611.750	0.53221	0.55859
PyGAM	30min	0.95	726	6841.170	0.58136	0.63071
PyGAM	30min	0.99	148	10127.166	0.53444	0.61407

Notes for Table M6. PyGAM Pearson and Spearman Scores over varying ΔQ ranges	
Variables Included:	*Weather:* surface_net_solar_radiation_kWh_per_m2_log1p_std, wind_speed_mps_std, temperature_celsius_std, wind_dir_sin, wind_dir_cos, is_sunny *Time features:* doy_sin, doy_cos, hour_sin, hour_cos, is_weekend *Generation features:* hydro_share_std
Other Notes	Following the table M4, this table shows the relationship between ramp sizes and pearson and spearman scores

Table M7. Model Validation Comparison

Model Family	Intercept	Slope	R^2	Ramp pairs	Ramp Pair Window
PyGAM	-0.608	+2.297	0.149	14,222	30 min
Ridge	-1.221	+3.479	0.130	14,222	30 min
Huber	-1.226	+3.484	0.129	14,222	30 min
OLS	-2.599	+6.832	0.152	14,222	30 min

Notes for Table M7. Model Validation Comparison	
Core Variables Included:	*Weather:* surface_net_solar_radiation_kWh_per_m2_log1p_std, wind_speed_mps_std, temperature_celsius_std, wind_dir_sin, wind_dir_cos, is_sunny *Time features:* doy_sin, doy_cos, hour_sin, hour_cos, is_weekend *Generation features:* hydro_share_std
Other Notes	This table compares the best of each model family developed and their validation calibration scores (comparing observed slopes to estimated slopes). Results are sorted their intercept and slope distances from 0 and 1 respectively.

Table M8. Model Calibration – Validation Set

Model	ΔQ (MW) ≥	Raw Intercept	Calibrated Intercept	Raw Slope	Calibrated Slope	Raw RMSE	Calibrated RMSE	Ramp pairs	Ramp Pair Window
PyGAM	100	-0.746	0.217	+2.638	+ 1.493	0.449	0.440	7,111	30 min
PyGAM	1,000	-0.602	0.144	+2.283	+ 1.292	0.183	0.153	5,852	30 min
PyGAM	2,000	-0.605	-0.004	+ 2.295	+ 1.031	0.153	0.119	4,216	30 min
PyGAM	4,000	-0.606	-0.007	+ 2.287	+ 1.027	0.140	0.101	1,653	30 min
PyGAM	4,255	-0.595	-0.003	+ 2.261	+ 1.016·	0.139	0.100	1,451	30 min
PyGAM	4,500	-0.558	+0.011	+ 2.172	+ 0.976	0.136	0.097	1,272	30 min

Notes for Table M8. Model Comparison	
Core Variables Included:	***Weather:*** surface_net_solar_radiation_kWh_per_m2_log1p_std, wind_speed_mps_std, temperature_celsius_std, wind_dir_sin, wind_dir_cos, is_sunny ***Time features:*** doy_sin, doy_cos, hour_sin, hour_cos, is_weekend ***Generation features:*** hydro_share_std
Other Notes	This table compares the best of each model family developed and their validation calibration scores (comparing observed slopes to estimated slopes). Results are sorted their intercept and slope distances from 0 and 1 respectively.

Table M9. Model Calibration – Test Set

Model	ΔQ (MW) ≥	Raw Intercept	Calibrated Intercept	Raw Slope	Calibrated Slope	Raw RMSE	Calibrated RMSE	Ramp pairs	Ramp Pair Window
PyGAM	100	-0.305	+0.017	+1.606	+0.909	0.342	0.336	17,137	30 min
PyGAM	1,000	-0.359	-0.012	+1.734	+0.982	0.149	0.134	14,049	30 min
PyGAM	2,000	-0.605	-0.004	+ 2.295	+ 1.031	0.153	0.119	**10,335**	30 min
PyGAM	4,000	-0.606	-0.007	+ 2.287	+ 1.027	0.140	0.101	5,092	30 min
PyGAM	4,255	-0.595	-0.003	+ 2.261	+ 1.016·	0.139	0.100	4,625	30 min
PyGAM	4,500	-0.558	+0.011	+ 2.172	+ 0.976	0.136	0.097	4,227	30 min

Notes for Table M9. Model Comparison	
Core Variables Included:	***Weather:*** surface_net_solar_radiation_kWh_per_m2_log1p_std, wind_speed_mps_std, temperature_celsius_std, wind_dir_sin, wind_dir_cos, is_sunny ***Time features:*** doy_sin, doy_cos, hour_sin, hour_cos, is_weekend ***Generation features:*** hydro_share_std
Other Notes	This table compares the best of each model family developed and their validation calibration scores (comparing observed slopes to estimated slopes). Results are sorted their intercept and slope distances from 0 and 1 respectively.

Table O1. Average Emissions Per Household

Time Period	Average CO2 emissions per household (kg)
Week	17.227
Month	74.653
Year	895.845

Table O2. Average Customer Usage By Region and Time Period

Region	Time Period	Usage (kWh)	Periods included
All	Hour	0.4717	8,327
All	Week	60.2525	68
All	Month	273.0477	16
All	Year	918.5136	3
Delhi	Hour	0.4077	4,163
Delhi	Week	33.4946	55
Delhi	Month	162.6607	13
Delhi	Year	988.2248	2
Mumbai	Hour	0.5267	4,327
Mumbai	Week	87.0975	27
Mumbai	Month	376.2835	7
Mumbai	Year	863.5501	2

Table O3. Median Customer Usage By Time Period

Region	Time Period	Usage (kWh)	Periods included
All	Hour	0.4836	8,327
All	Week	66.4678	68
All	Month	268.8008	16
All	year	967.2981	3

Table O4. Shifts and Savings by Region

Region	Move Count	Average CO_2 per move (g)	Median CO_2 per move (g)	Average kWh per move	Median kWh per move
All	14,996,536	44.1995	26.1329	0.4286	0.3230
Delhi	12,394,970	44.3293	25.9728	0.4191	0.3185
Mumbai	2,601,566	43.5812	26.7853	0.4738	0.3425

Table O5. Shift Confidence Levels

Confidence Band	Move Count	Share of moves	Total kWh moved	kWh per move
low	554,231	3.7%	237,188	0.4279
medium	14,442,305	96.3%	6,191,403	0.4286

Table O6. Average Annual Savings Per Household

Region	Annual CO_2 savings per household (kg)
All	6.344
Delhi	6.433
Mumbai	5.952

Table O7. Average Weekly Savings Per Household

Region	Household - Weeks	Average CO_2 savings (g)	Median CO_2 savings (g)	Average kWh moved	Median kWh moved	Average moves per week	Median moves per week
All	5,118,766.0	129.492	88.753	1.2558	1.0360	2.9297	3
Delhi	4,203,122	130.726	89.255	1.2361	1.0251	2.9489	3
Mumbai	915,644	123.824	86.931	1.3463	1.0792	2.8412	3

Table O8. Average Emissions Per Household Metrics

Time Period	Average CO_2 emissions per household (kg)	Average CO_2 savings (kg)	Impact
Week	17.227	0.129	0.75% reduction
Year	895.845	6.344	0.71% reduction

Table O9. General Metrics

Statistic	Value
Households analysed	238,479
kWh move rate	≥ 1,100 /second
Household evaluation rate	≥ 40 / second
Total shift savings estimated by marginal emissions (t CO_2)	662.84
Total shift savings estimated by average emissions (t CO_2)	160.51
Difference between average and marginal emissions estimates (t CO_2)	502.33
Difference between average and marginal emissions estimates (%)	Marginal emissions are measured +313%

Table S1. Year over Year Growth of Generation Sources

Year	Thermal	Gas	Nuclear	Hydro	Renewable	Total
2021	13.9%	-22.8%	-1.1%	-1.5%	17.1%	**10.3%**
2022	8.5%	-34.4%	8.8%	8.4%	21.1%	**8.4%**
2023	9.9%	15.2%	4.7%	-15.6%	20.3%	**7.9%**
2024	3.5%	14.3%	11.8%	3.7%	10.1%	**4.7%**
2025	3.2%	-25.4%	4.1%	-25.4%	18.6%	**1.9%**

Table S2. Generation Growth Relative to 2021 (% change)

Year	Thermal	Gas	Nuclear	Hydro	Renewable
2021	0.0%	0.0%	0.0%	0.0%	0.0%
2022	8.5%	-34.4%	8.8%	8.4%	21.1%
2023	19.2%	-24.5%	14.0%	-8.6%	45.7%
2024	23.3%	-13.7%	27.4%	-5.1%	60.5%
2025	27.2%	-35.6%	32.5%	-29.2%	90.3%

Table S3. VIF Scores with all Generation Sources Included

Variable	VIF
Const	89735.003
demand_met	2.635
thermal_generation_share	504.051
gas_generation_share	6.140
hydro_generation_share	208.627
nuclear_generation_share	5.264
renewable_generation_share	393.161
wind_speed_mps	1.196
wind_direction_meteorological	1.142
temperature_celsius	2.234
precipitation_mm	1.209
surface_net_solar_radiation_kWh_per_m2	1.653
total_cloud_cover	1.357

Table S4. VIF Scores with N-1 Generation Sources Included

Variable	VIF
Const	434.498117
demand_met	2.634714
gas_generation_share	1.272634
hydro_generation_share	1.473141
nuclear_generation_share	2.433939
renewable_generation_share	1.548980
wind_speed_mps	1.196071
wind_direction_meteorological	1.142489
temperature_celsius	2.234459
precipitation_mm	1.209362
surface_net_solar_radiation_kWh_per_m2	1.653393
total_cloud_cover	1.357693

Appendix D – Visualisations

Figure M1. Distribution of Ramp Sizes (ΔQ in MW) in the Validation Dataset

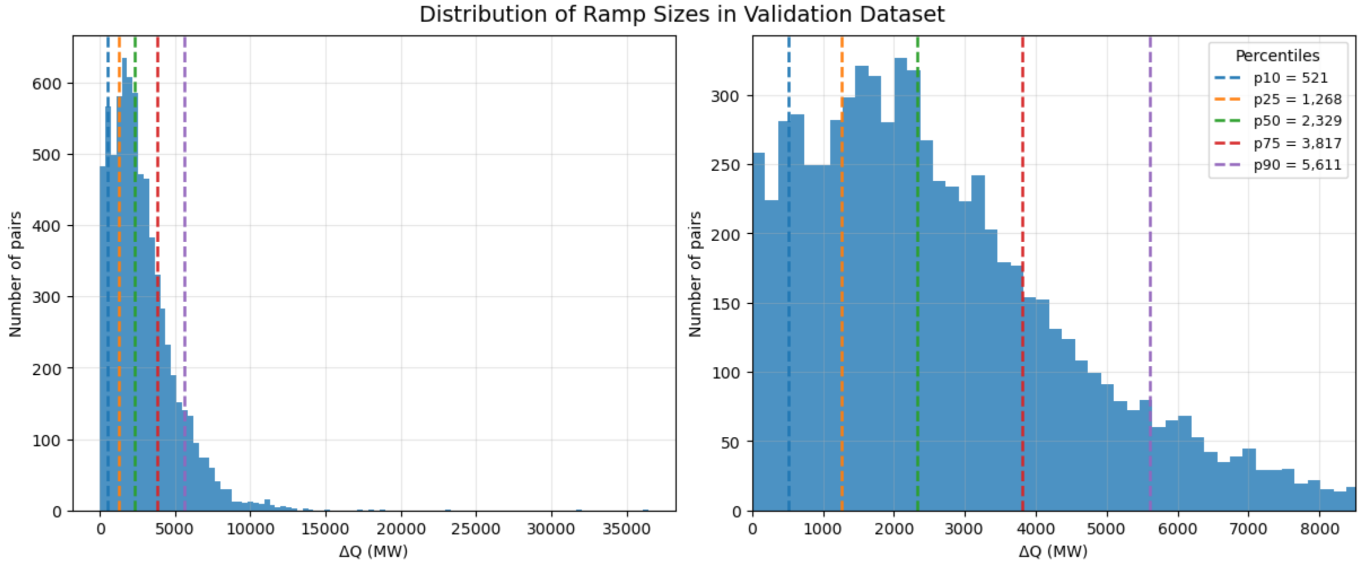

Figure M2. Distribution of Ramp Sizes (ΔQ in MW) in the Test Dataset

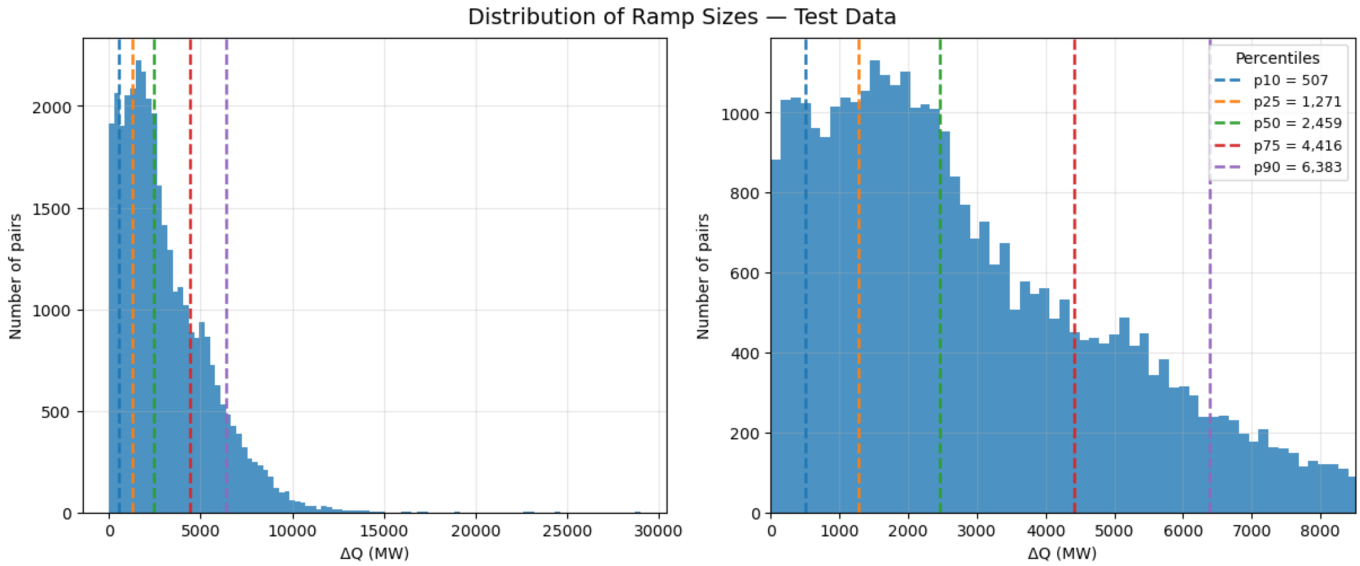

Figure M3. Distribution of Ramp Sizes (ΔQ in MW) in the Full Dataset

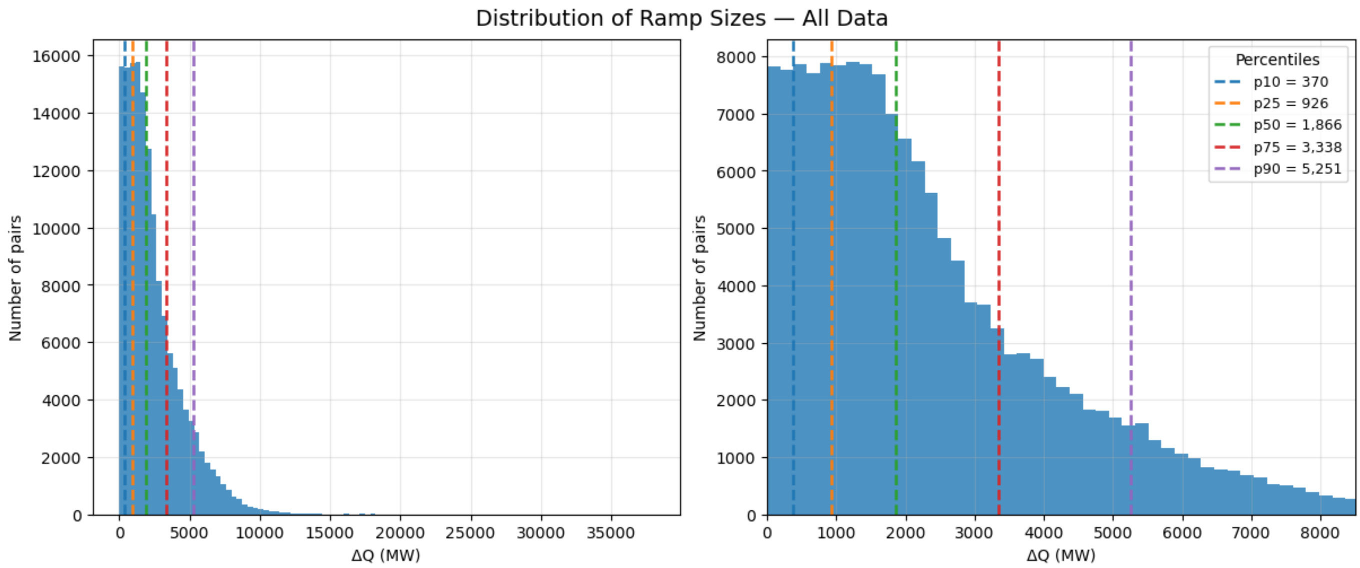

Figure M4. Correlation Scores Relative to Ramp Pair Sizes (ΔQ in MW) on Validation Dataset

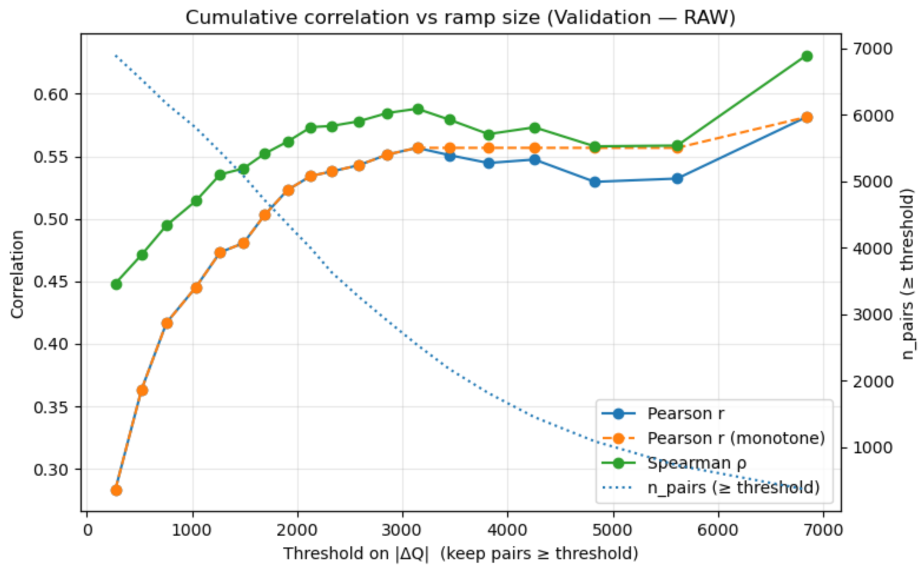

Figure M5. Correlation Scores Relative to Ramp Pair Sizes (ΔQ in MW) on Test Dataset

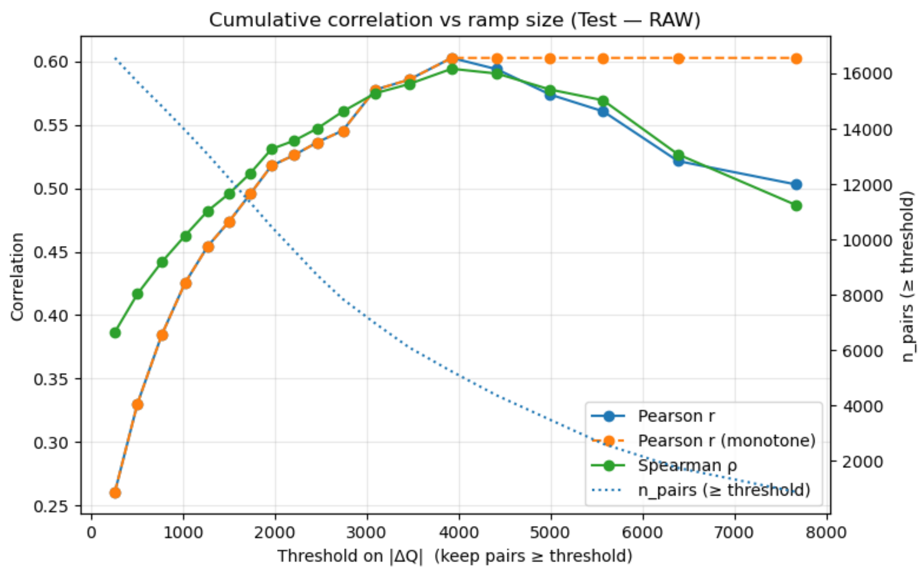

Figure M6. Calibration at Ramp ΔQ = 4,255 MW for Validation Dataset

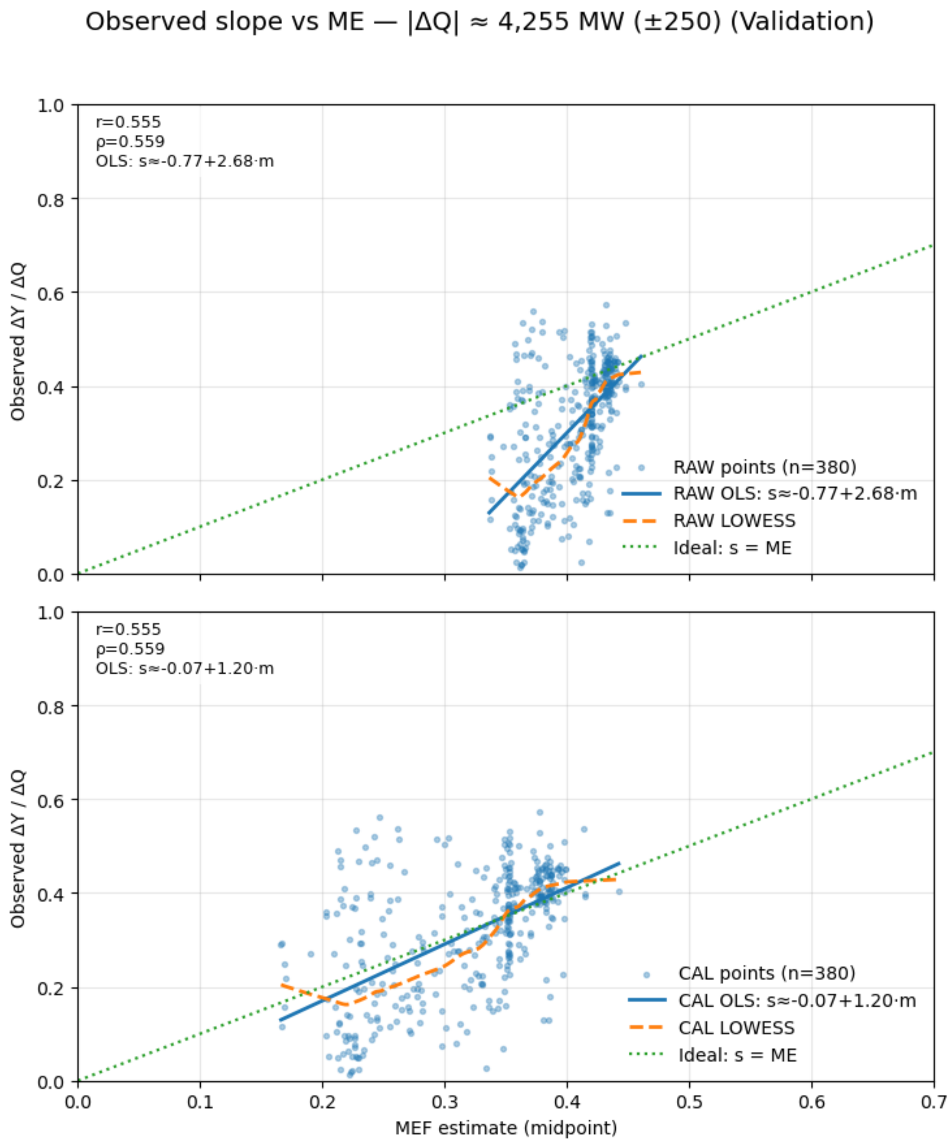

Figure M7. Calibration at 50th percentile Ramp (ΔQ in MW) for Validation Dataset

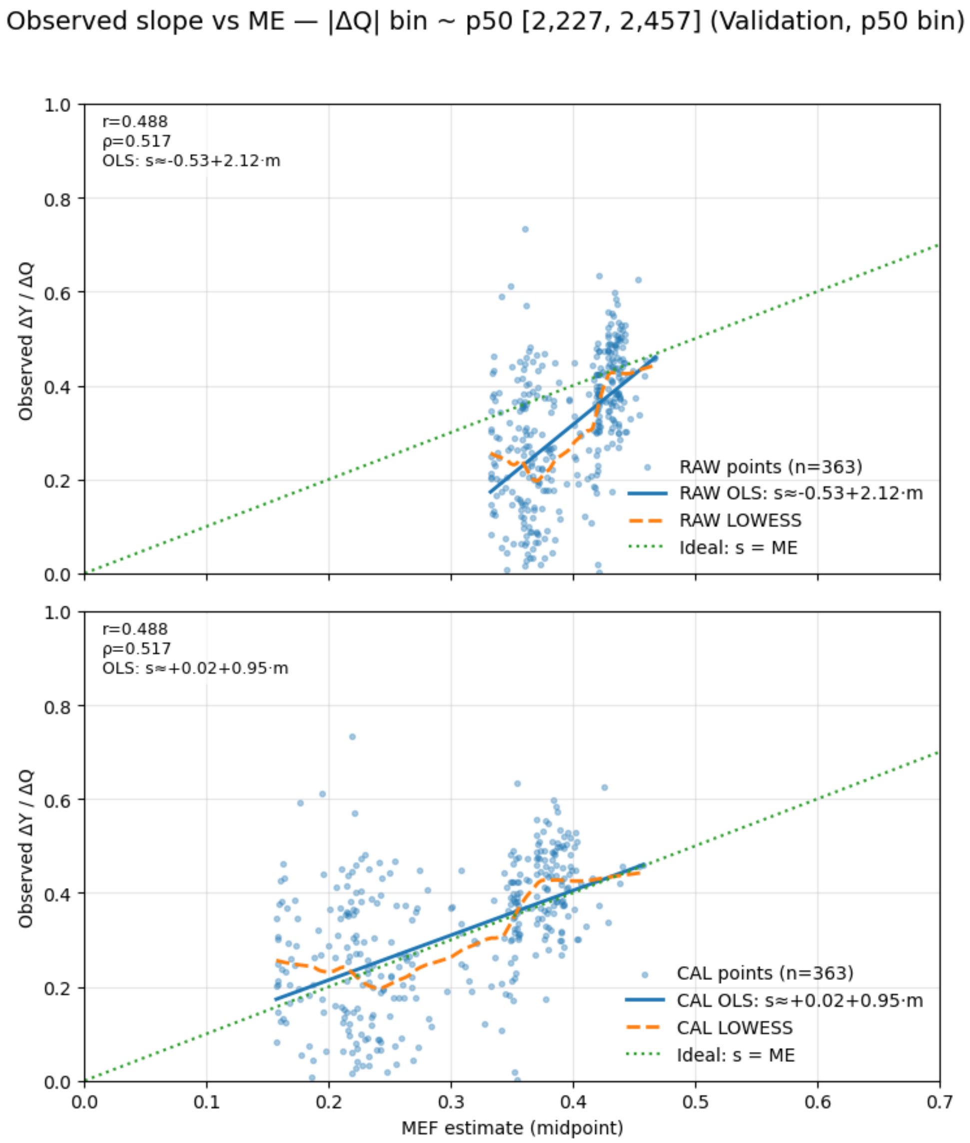

Figure M8. Calibration at 50th percentile Ramp (ΔQ in MW) for Test Dataset

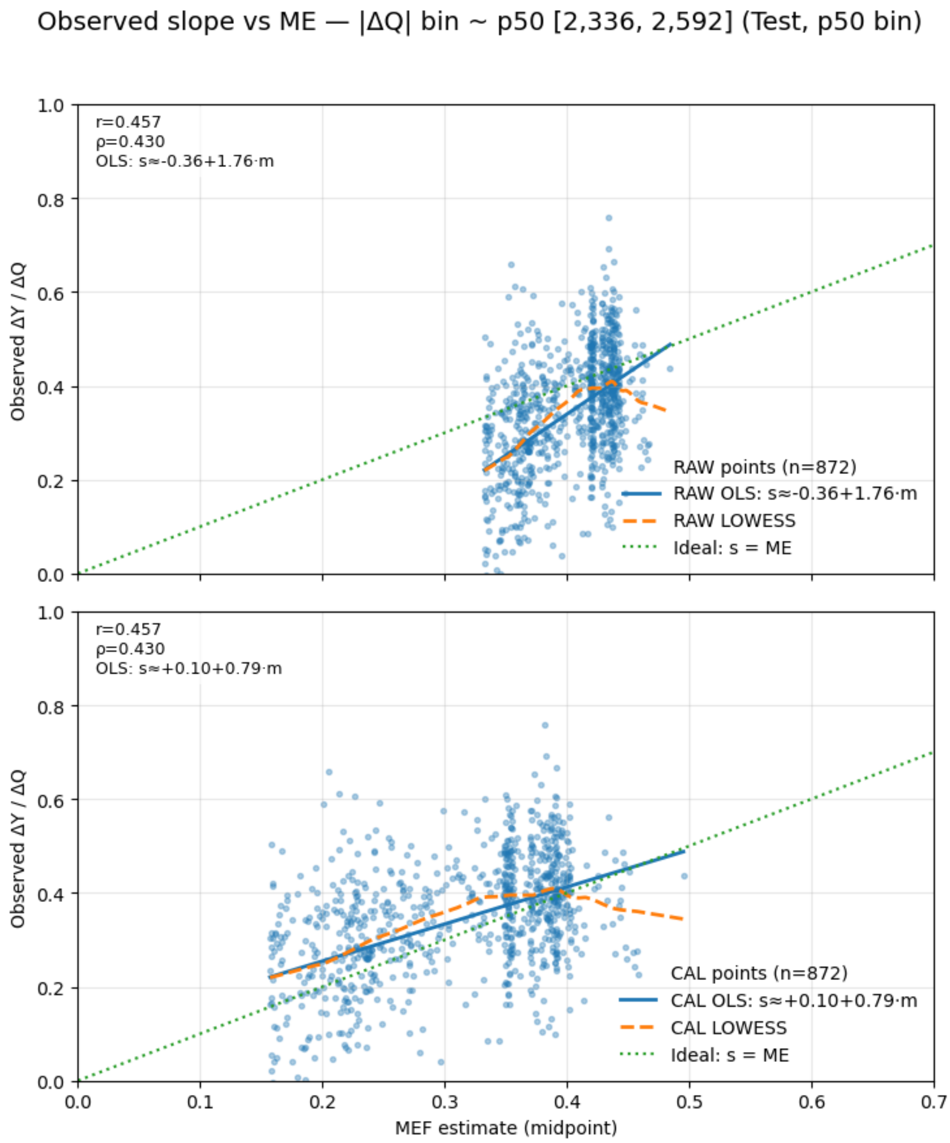

Figure M9. Calibration at 10th, 50th, 90th percentile ramps (ΔQ in MW) for Validation Dataset

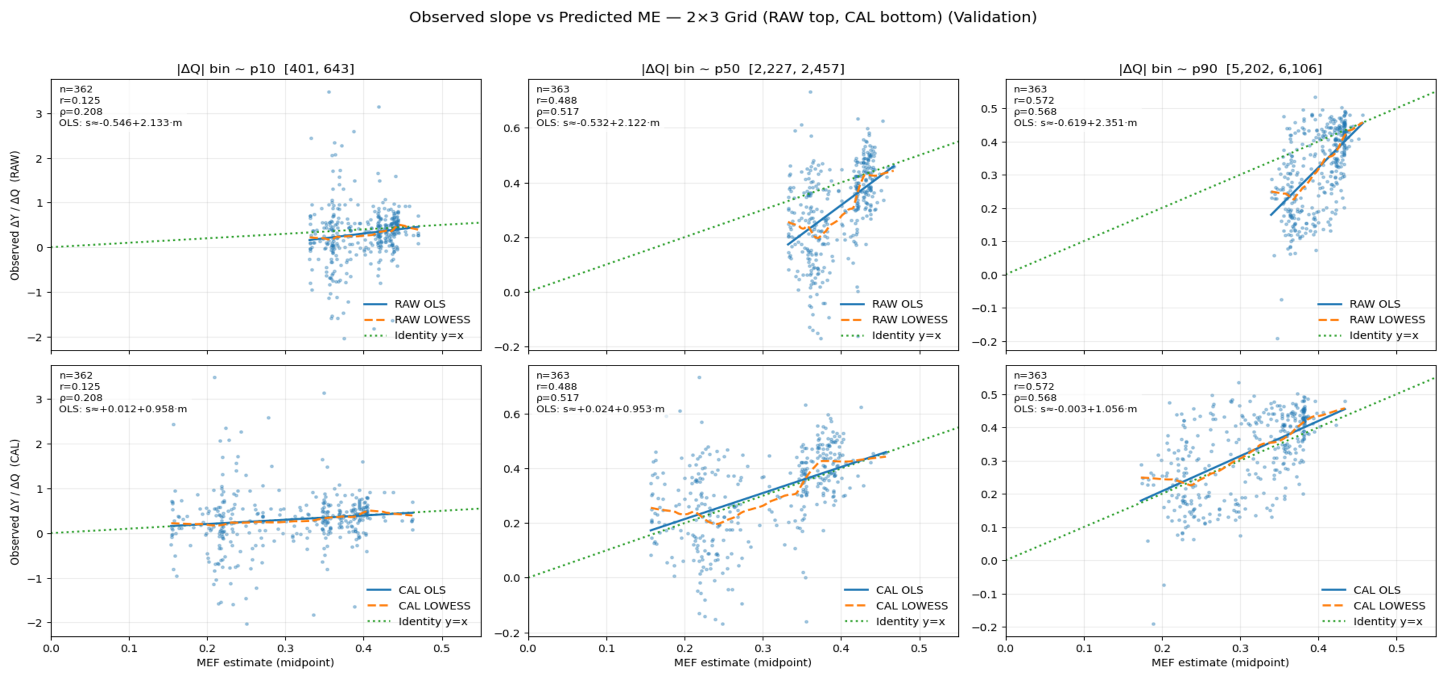

Figure M10. Calibration at 10th, 50th, 90th percentile ramps (ΔQ in MW) for Test Dataset

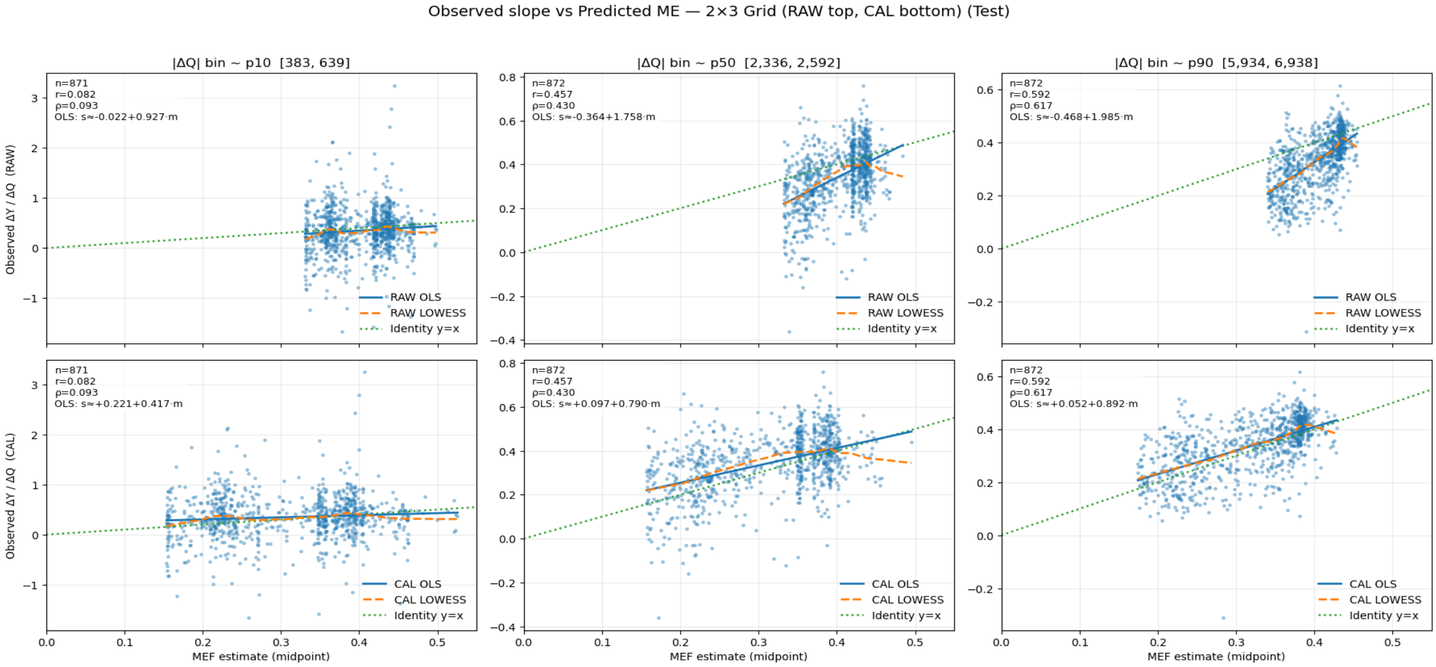

Figure M11. Hourly Average and Marginal Emission Factors by Estimation Method, 2024

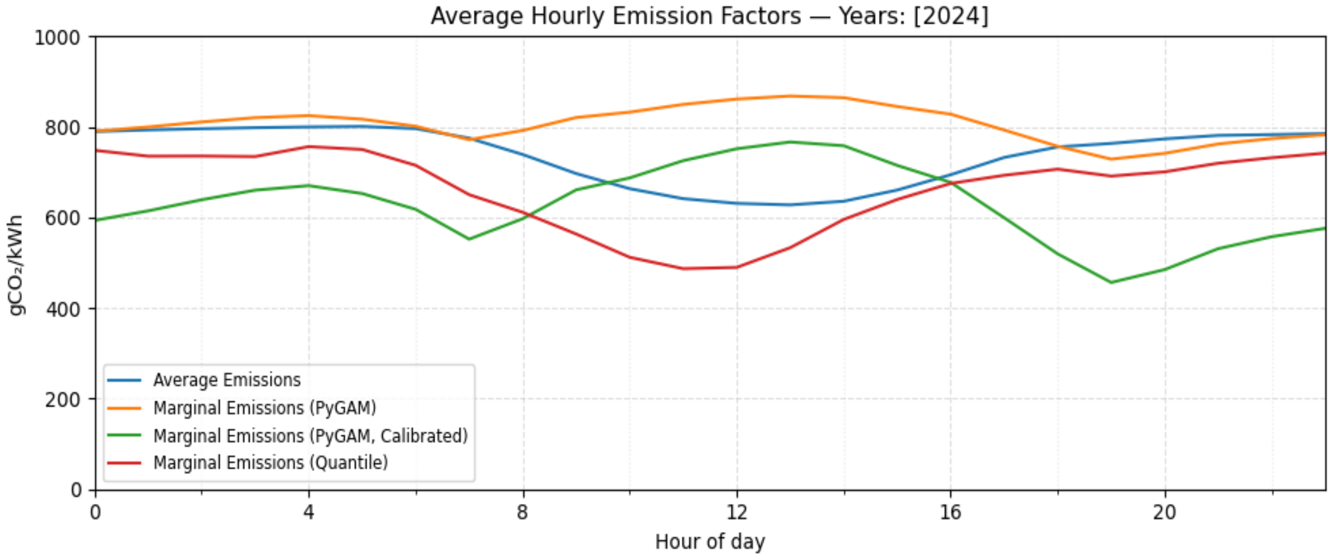

Figure M12. Hourly Average and Marginal Emission Factors by Estimation Method (reduced), 2024

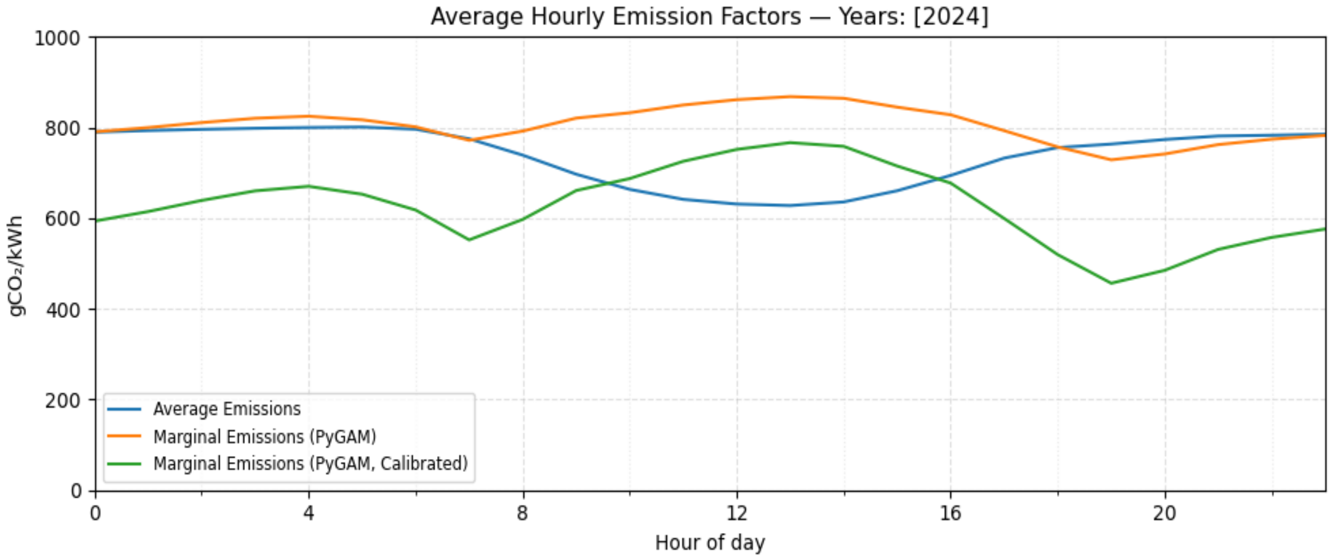

Figure M13. Hourly Marginal (calibrated) and Average Emission Factors, 2022-2024

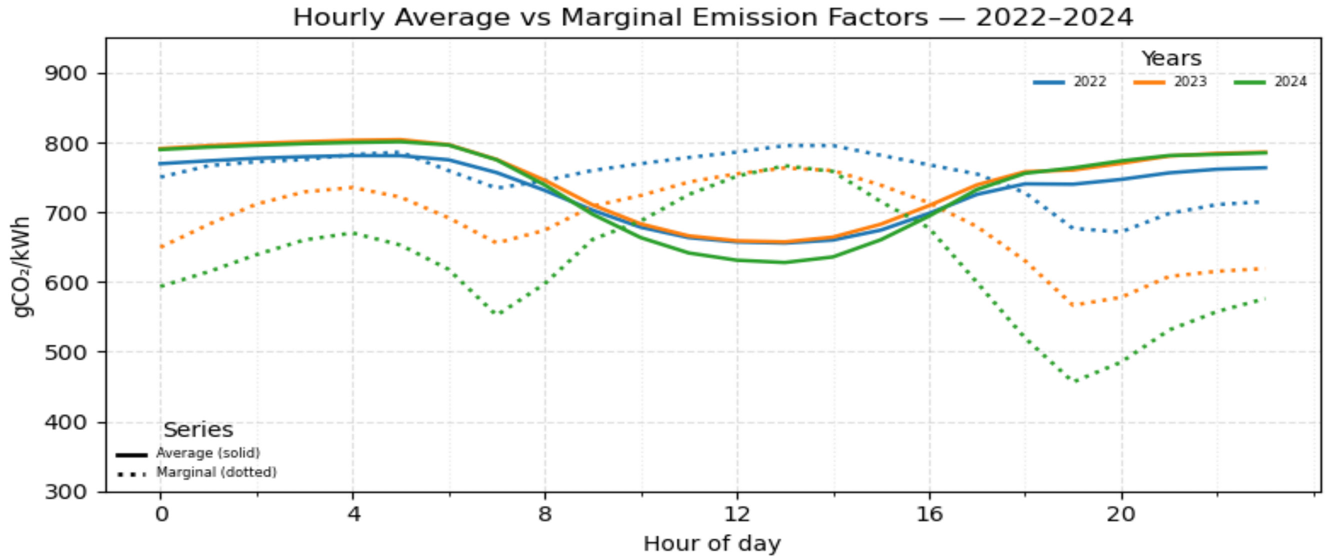

Figure M14. Distribution of Average and Marginal Emission Factor Values by Year

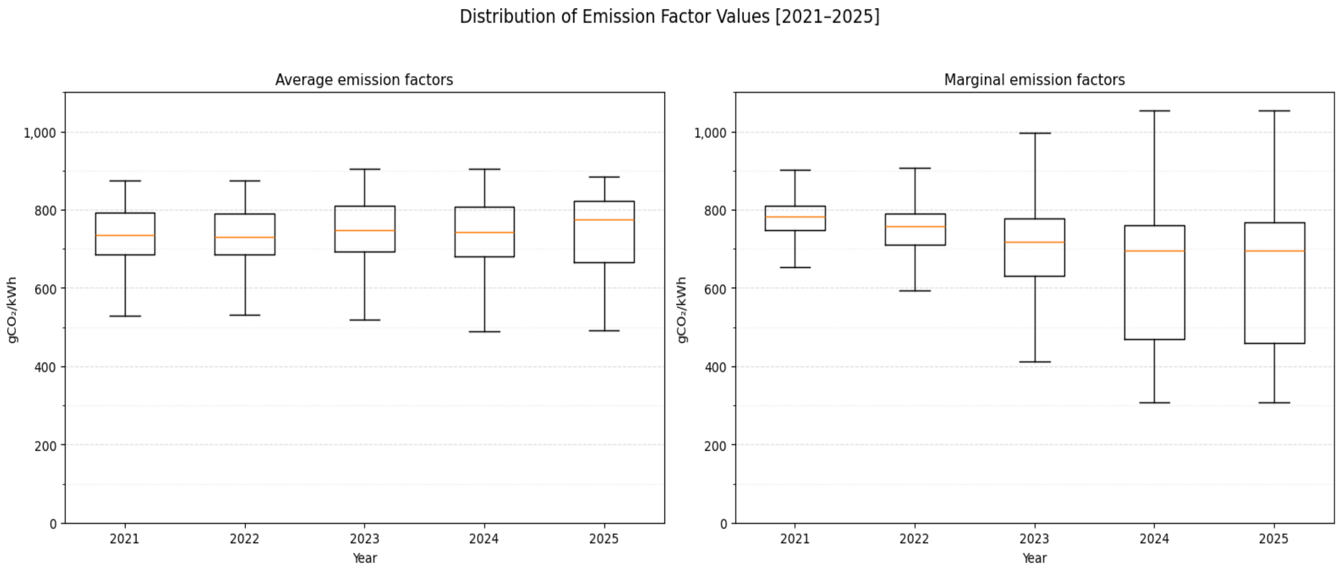

Figure M15. Marginal Emission Factor Estimate Confidence Label Counts by Year

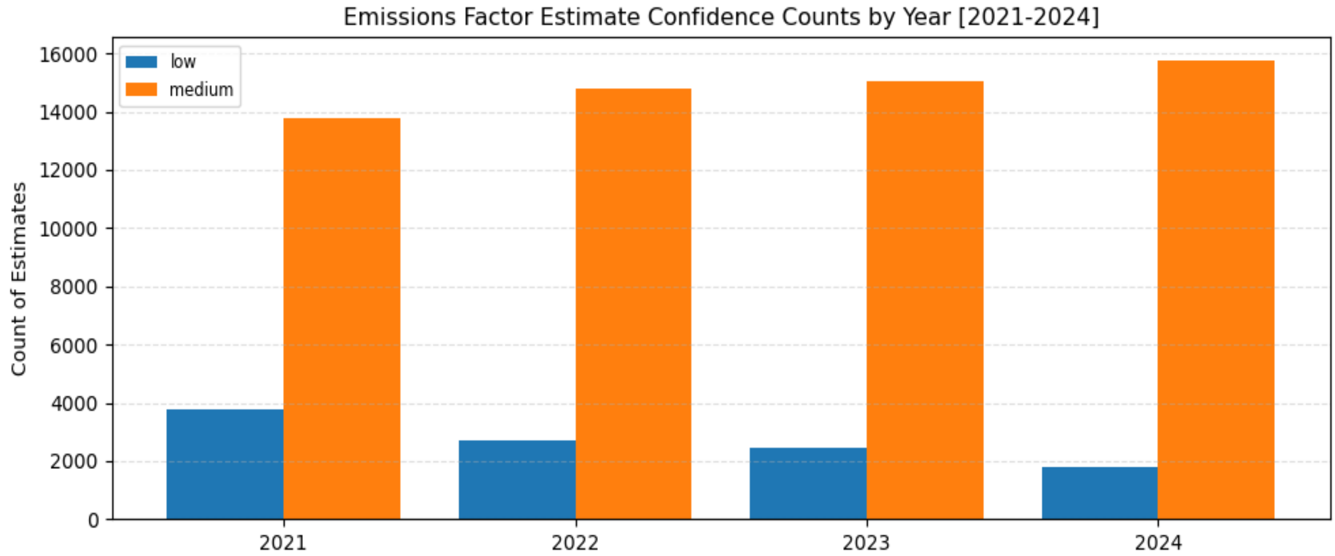

Figure M16. Marginal Emission Factor Estimate Confidence Label Counts by Hour, 2024

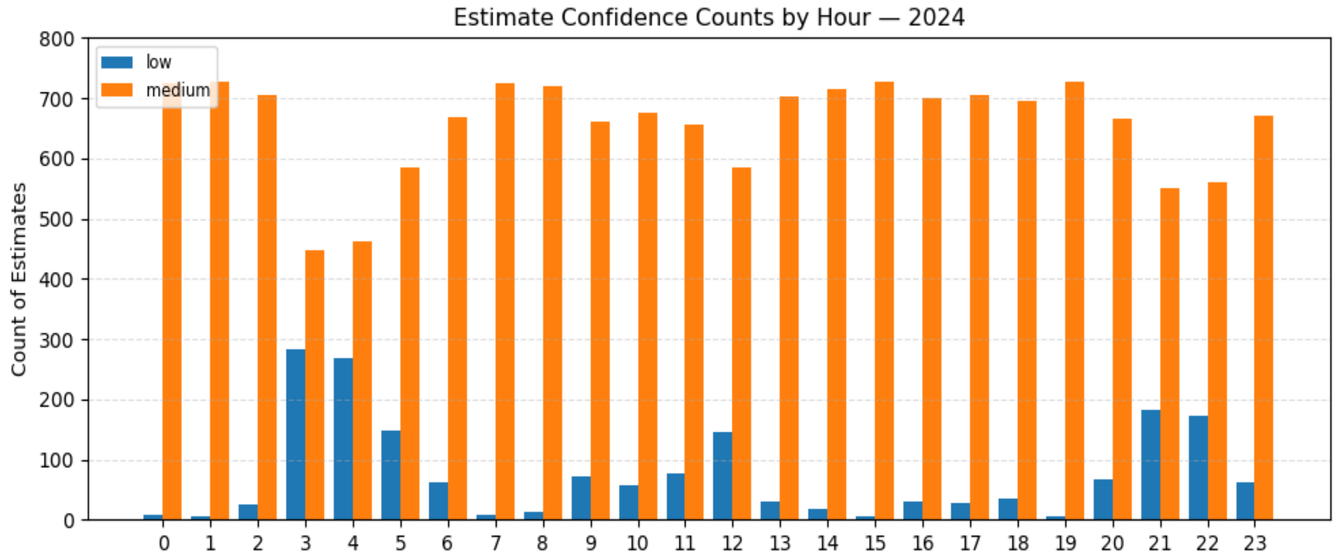

Figure M17. Hourly Average and Marginal Emission Factors with Average Usage (kWh), 2023

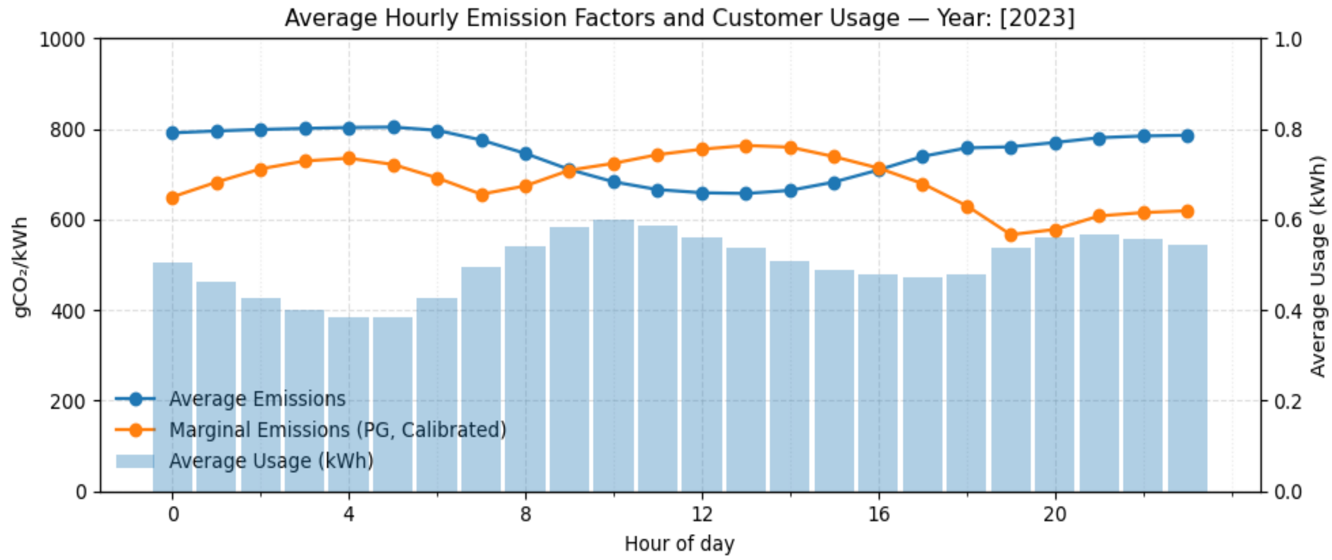

Figure M18. Hourly Average and Marginal Emission Factors with Average Usage (kWh), 2022

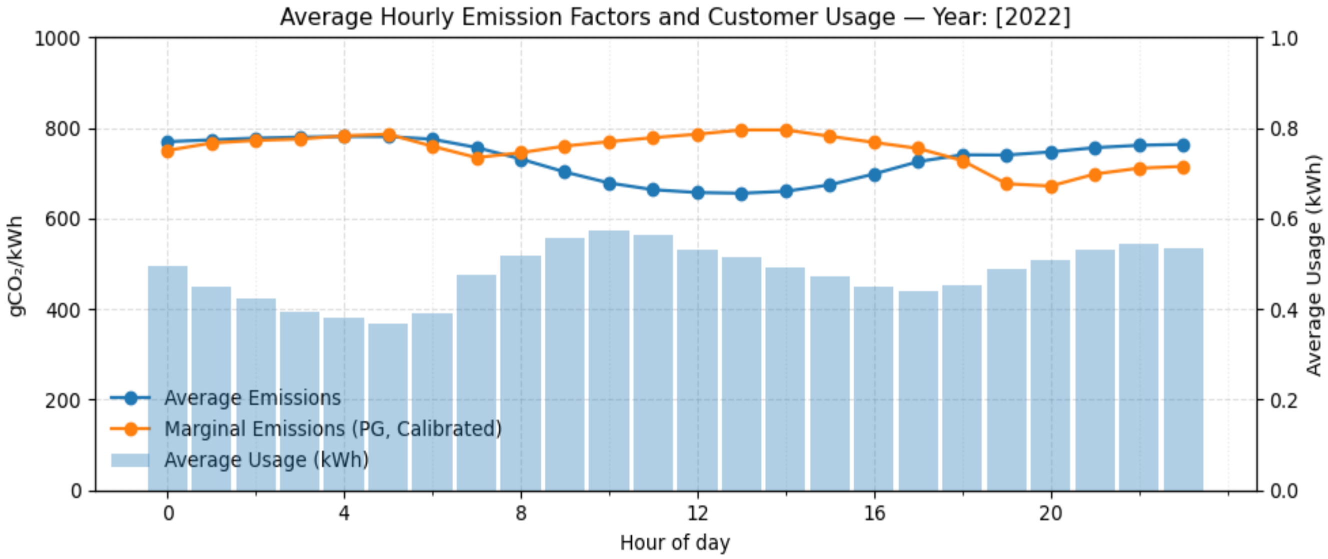

Figure O1. Hourly Energy Profile pre- and post-Optimisation

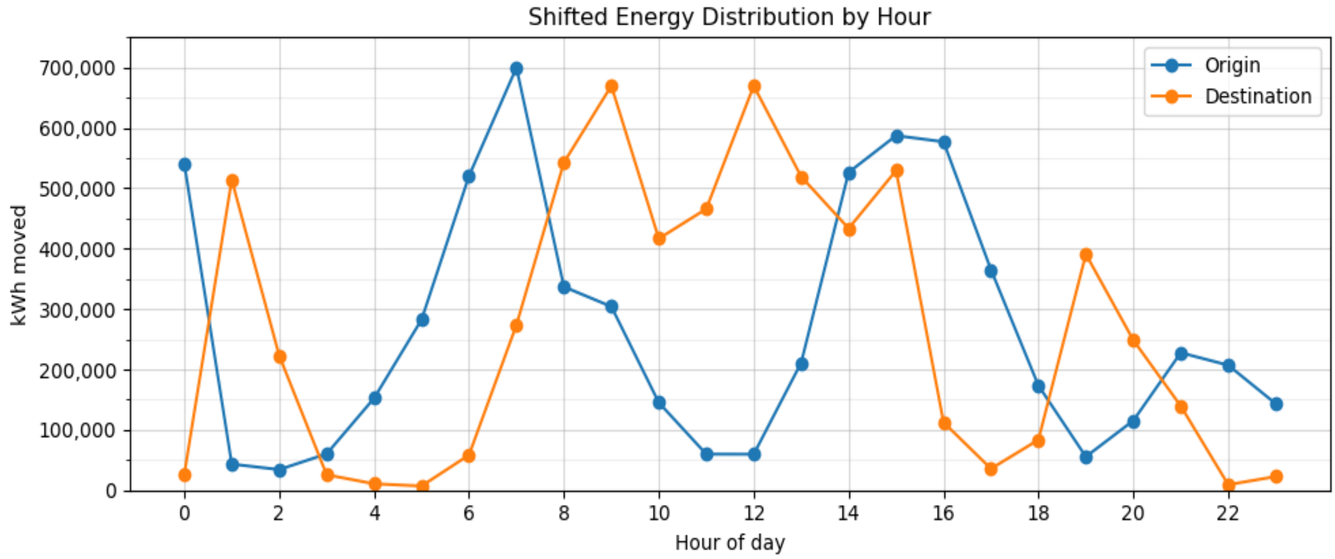

Figure O2. Hourly Emissions Profile pre- and post-Optimisation, Average and Marginal Emissions

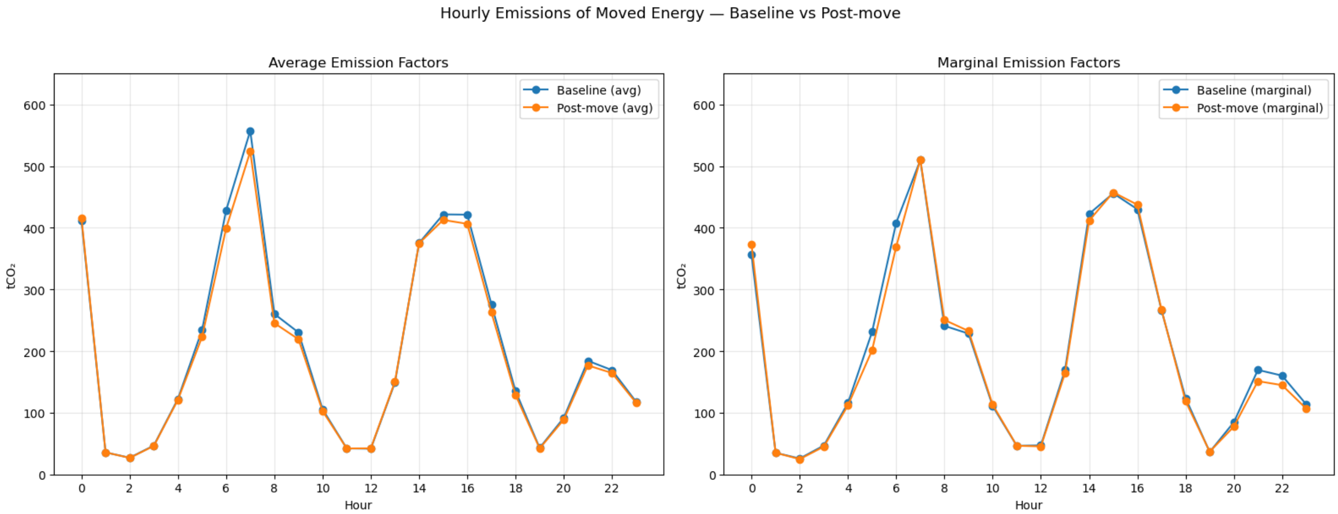

Figure O3. Distribution of Optimisation Shift Temporal Sizes

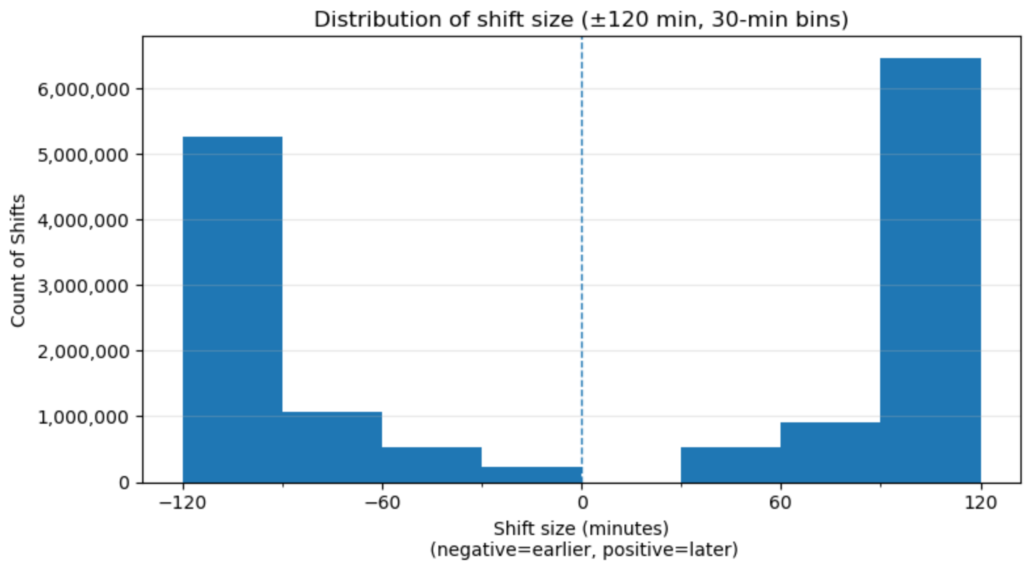

Figure O4. Direction of Optimisation Shifts

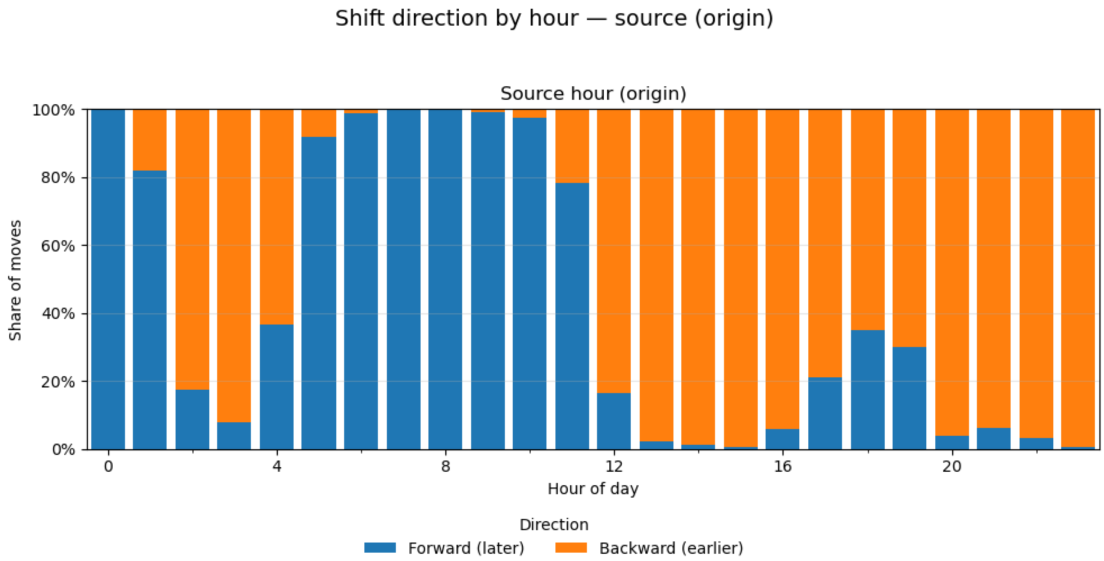

Figure S1. Histogram Distribution of Energy Usage per Hour for all Customers (kWh)

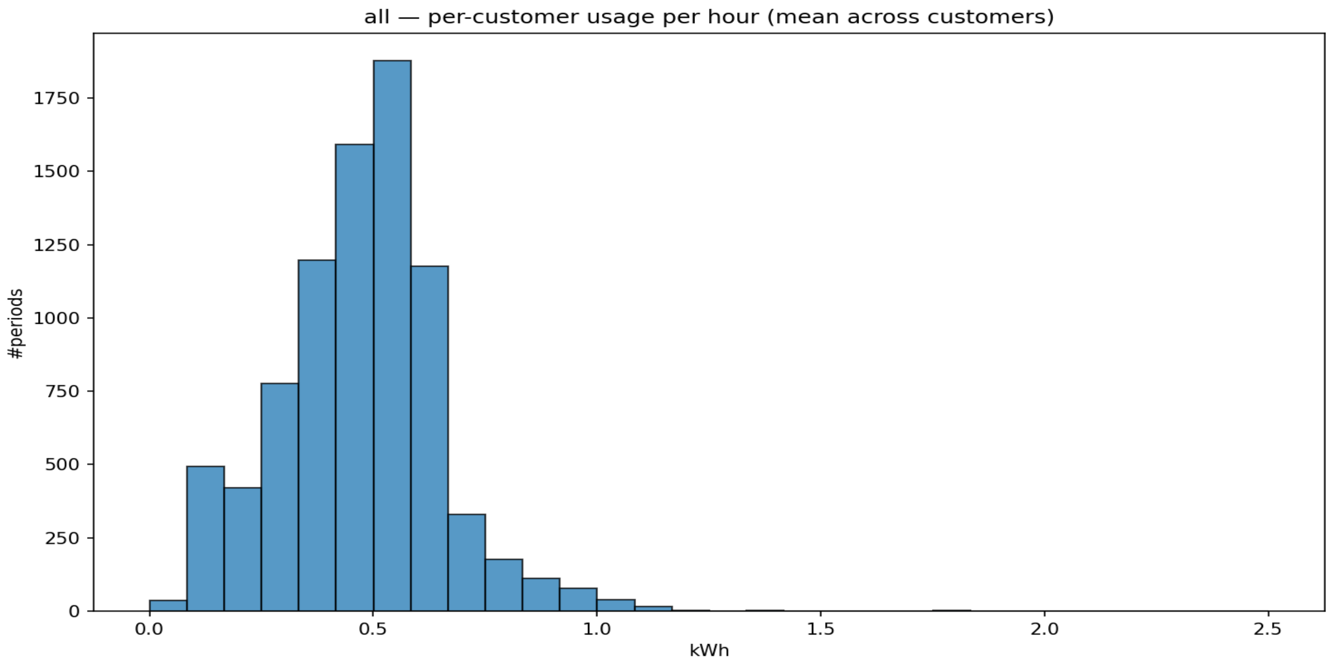

Figure S2. Boxplot Distribution of Energy Usage per Hour for all Customers (kWh)

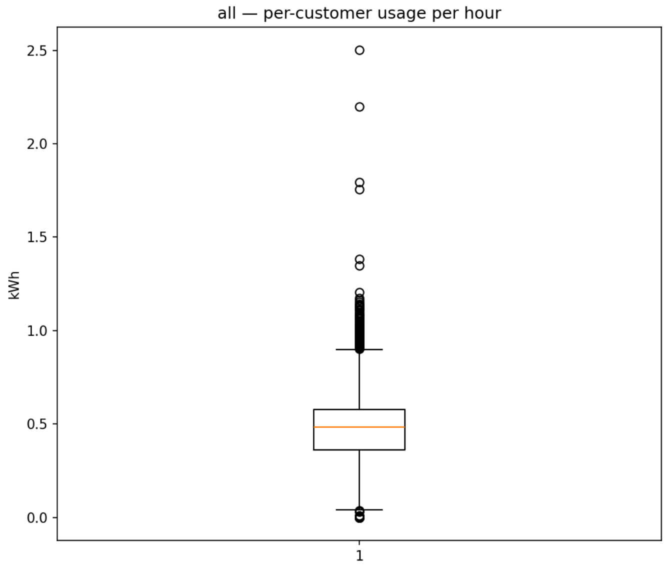

Figure S3. Hourly Energy Use Profiles for Weekday v Weekend Patterns.

Figure S4. Boxplot Distribution of Annual Energy Usage per Customer (kWh)

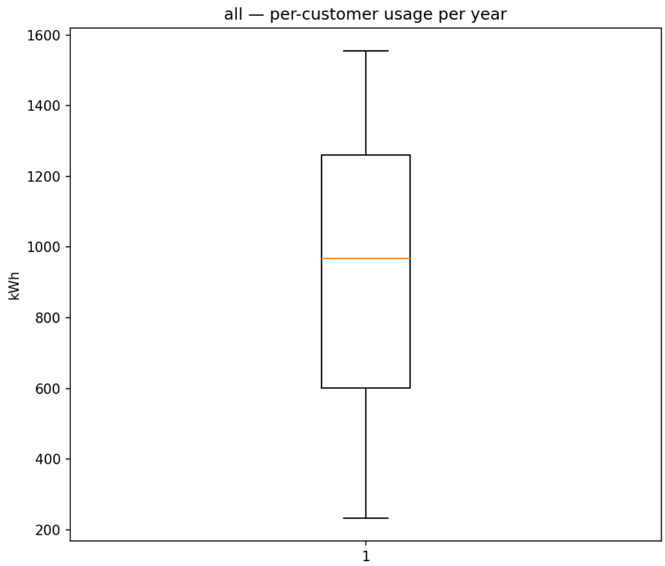

Figure S5. Location of Customers in Delhi

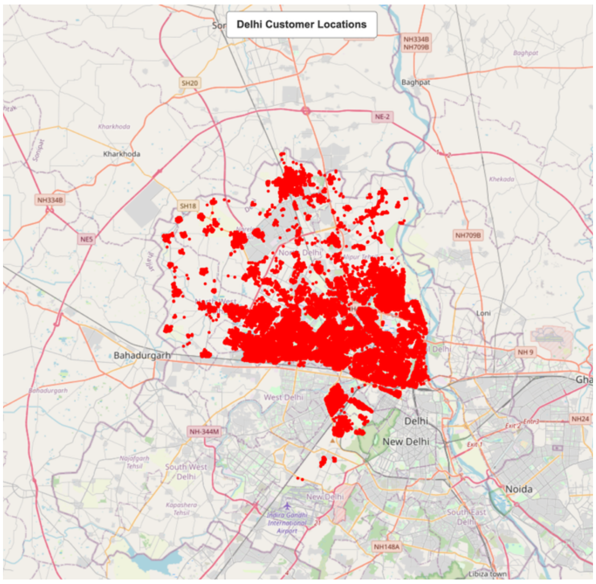

Figure S6. Location of Customers in Mumbai

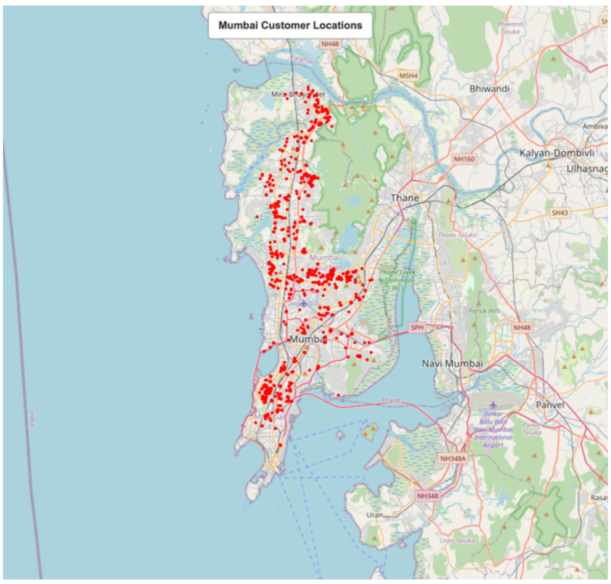

Figure S7. Daily Usage Profiles by Hour, Mumbai

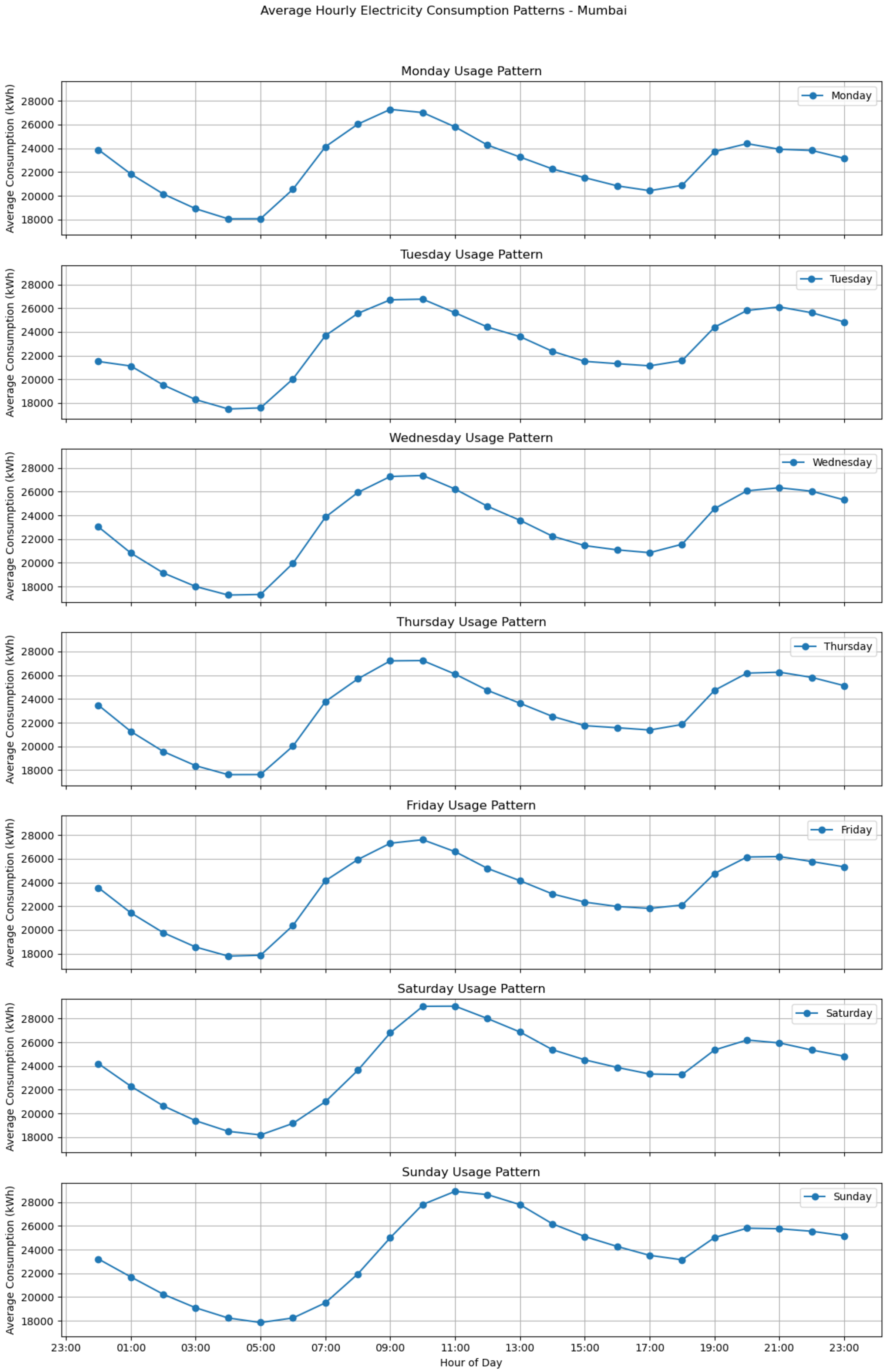

Figure S8. Quarterly Generation Shares and Trends, 2020-2025

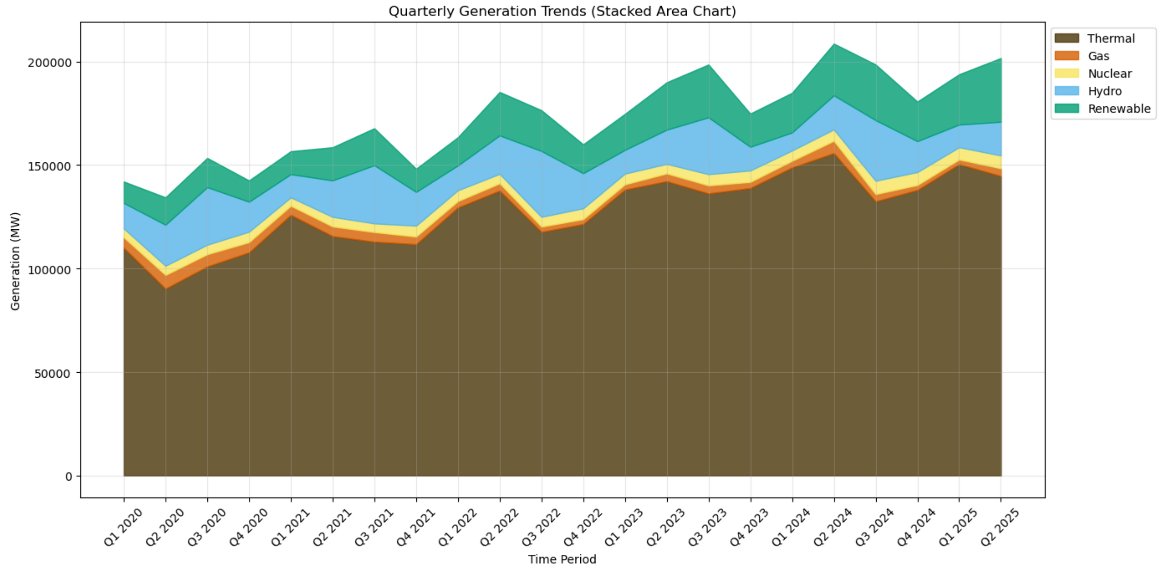

Figure S9. Generation Share Growth Relative to 2021

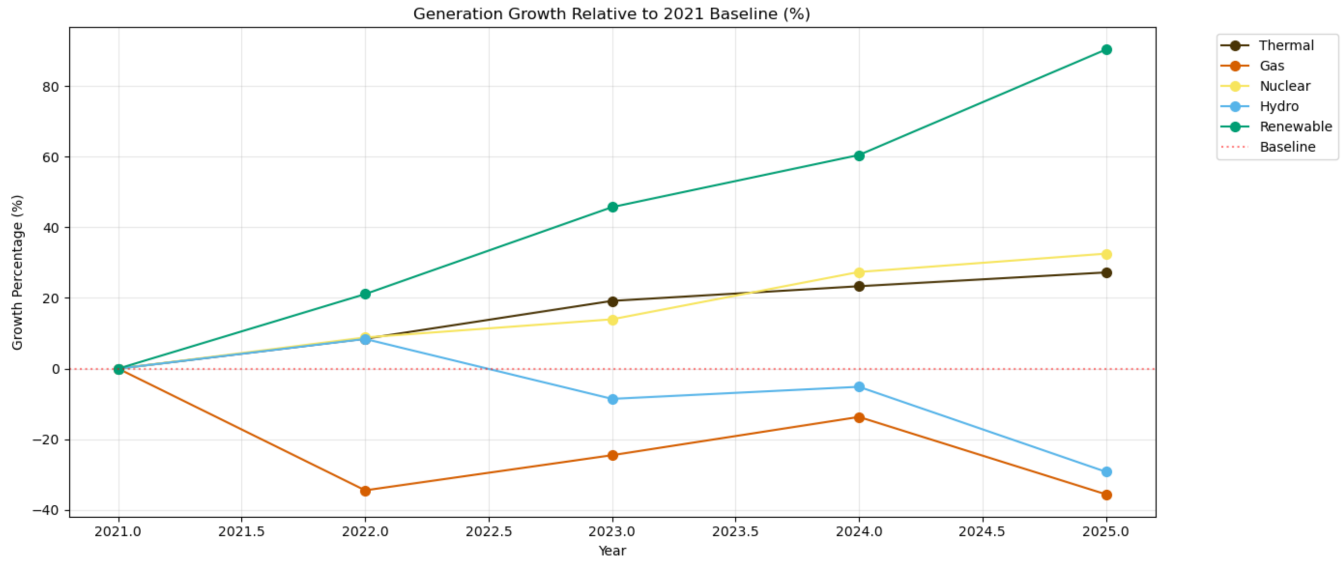

Figure S10. Year over Year Growth of Generation Sources, 2021-2025

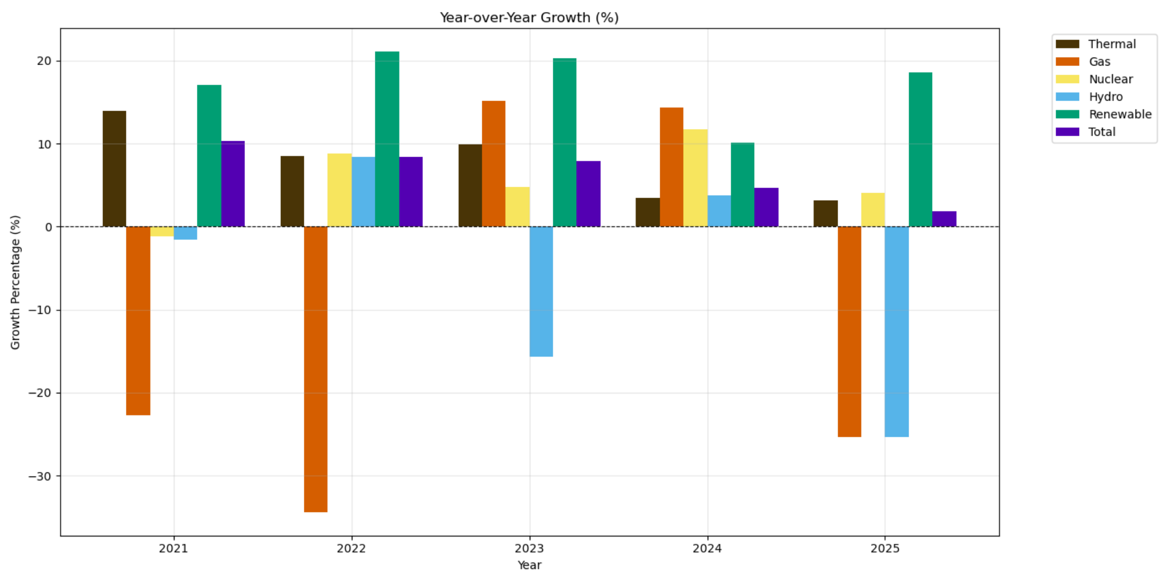

Figure S11. Contribution of Generation Sources to Overall Growth, 2021 Baseline

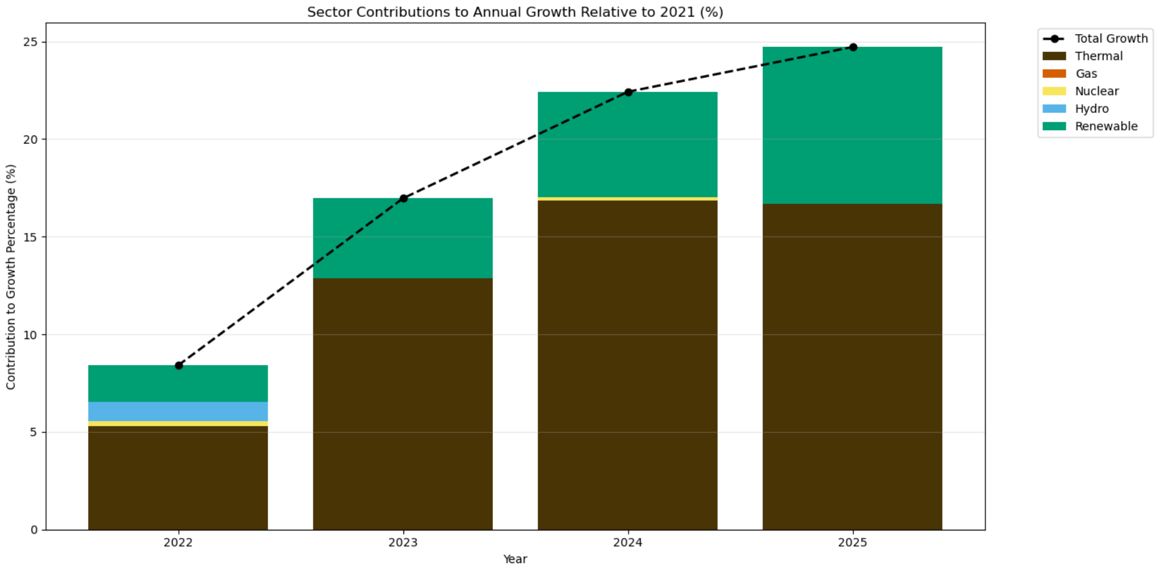

Figure S12. Total Generation Daily Profile, 2022-04-01

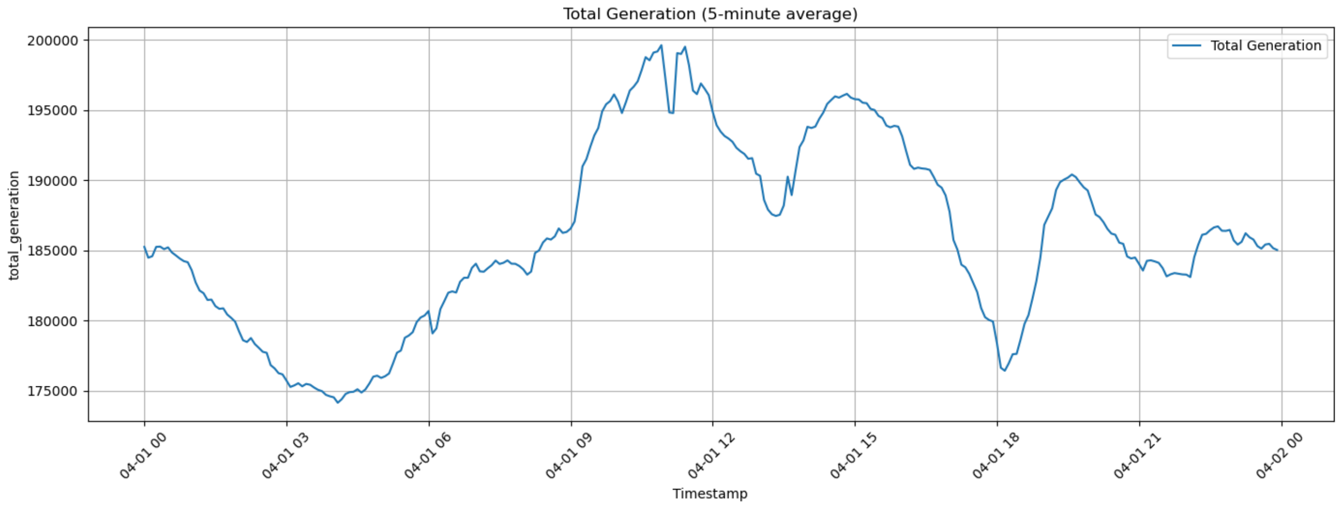

Figure S13. Thermal Generation Daily Profile, 2022-04-01

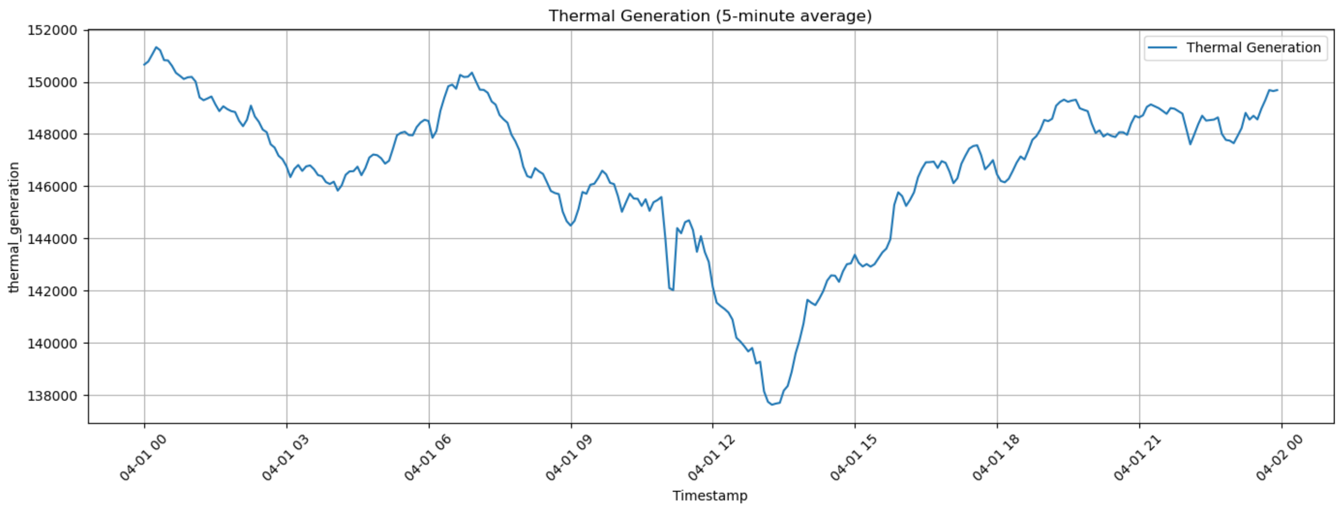

Figure S14. Gas Generation Daily Profile, 2022-04-01

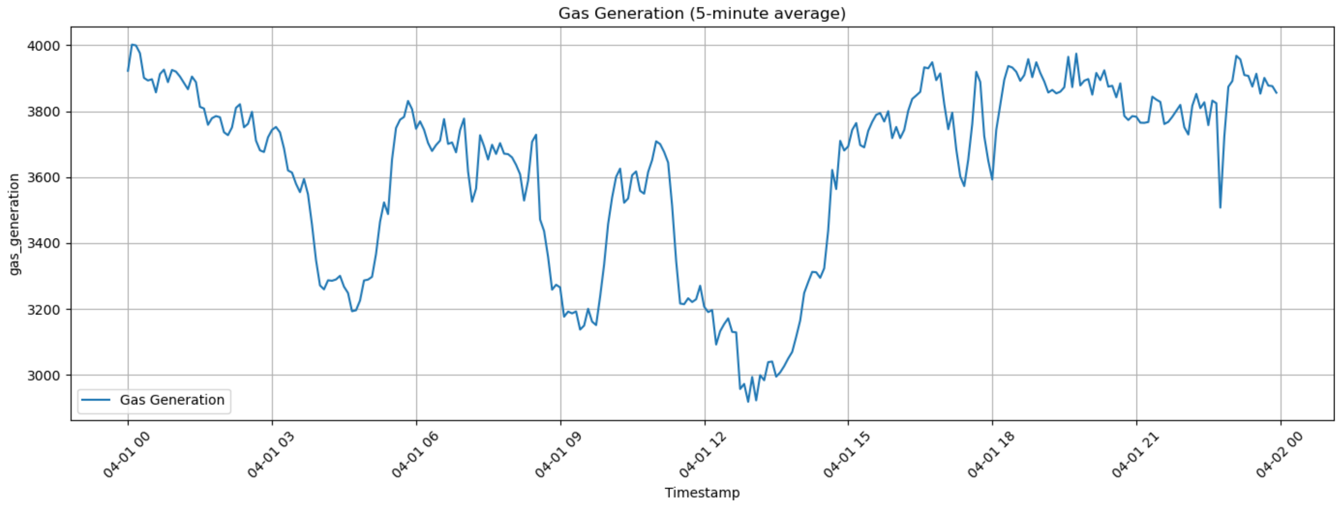

Figure S15. Hydro Generation Daily Profile, 2022-04-01

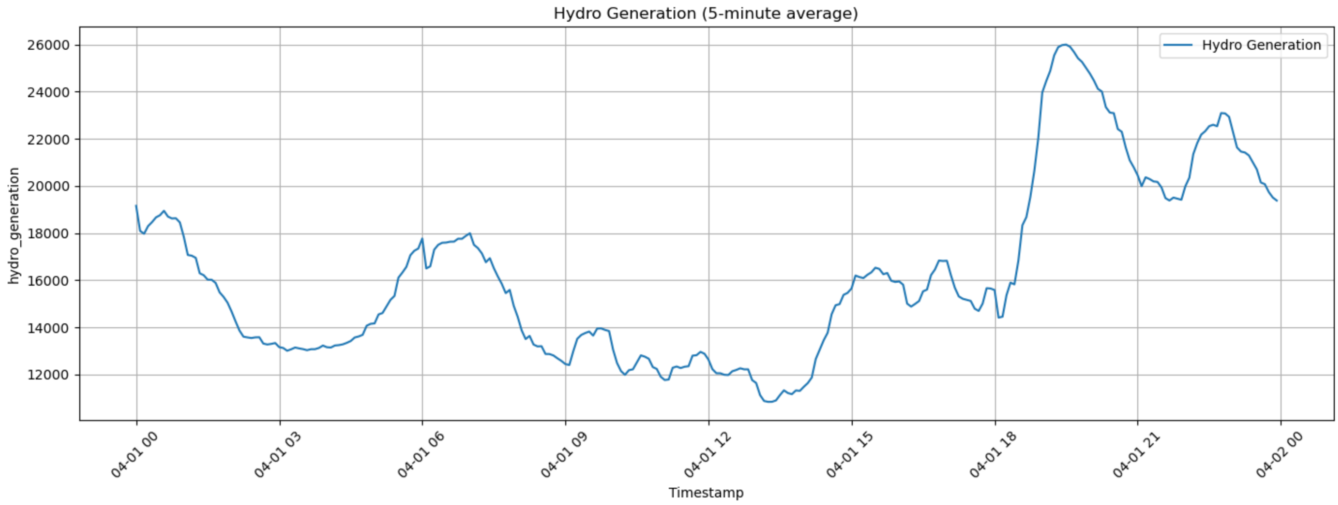

Figure S16. Nuclear Generation Daily Profile, 2022-04-01

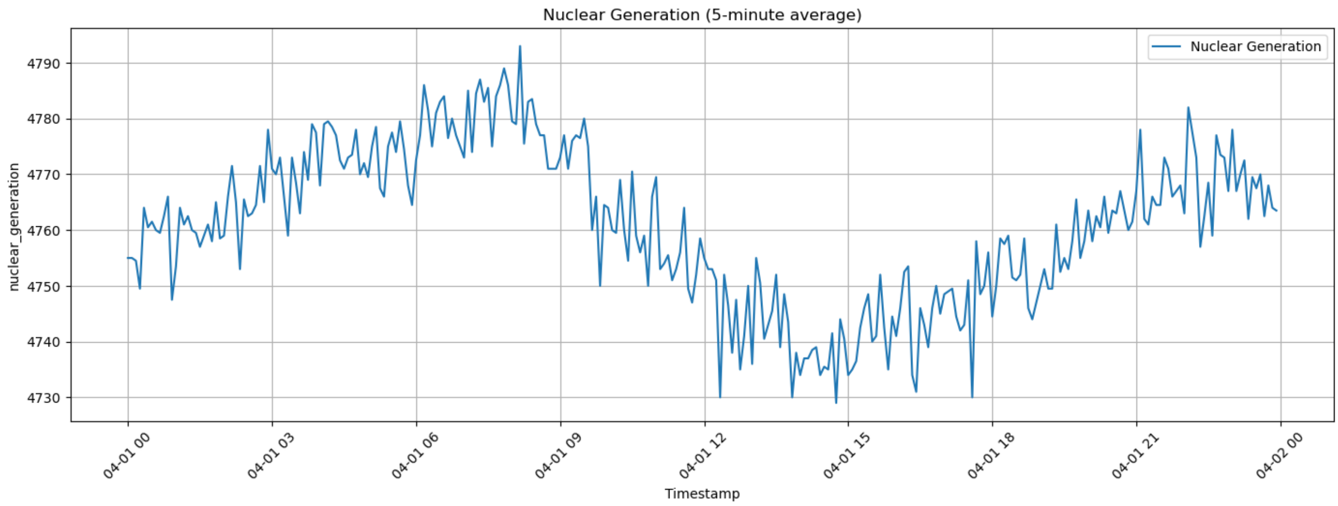

Figure S17. Renewable Generation Daily Profile, 2022-04-01

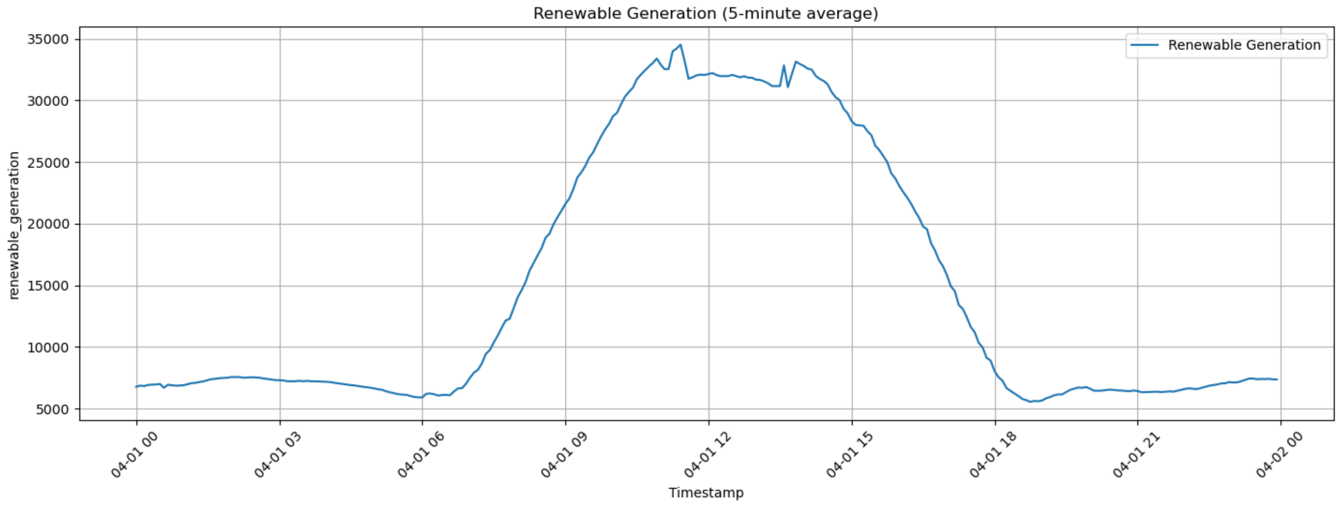

Figure S18. Gap Filling Generation Data, 2020-03-24

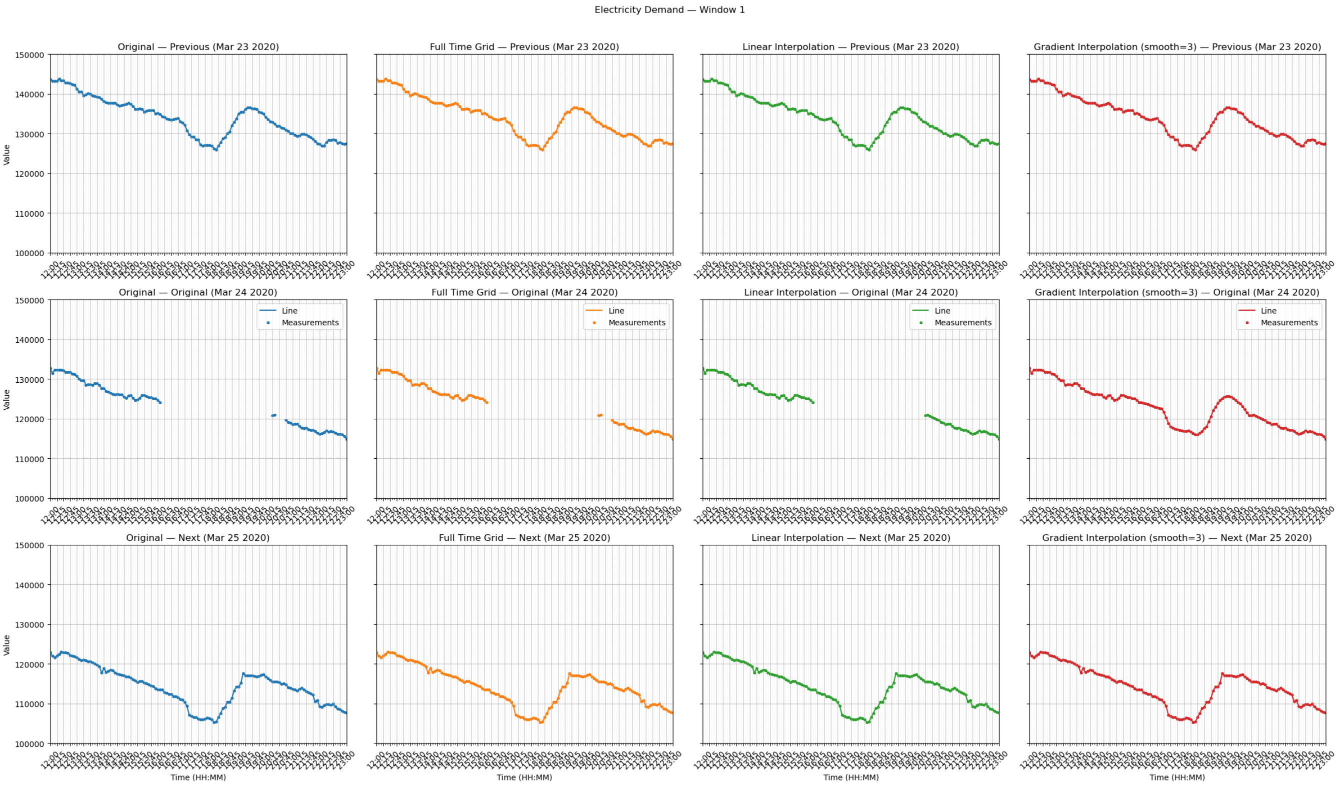

Figure S19. Gap Filling Generation Data, 2022-04-25

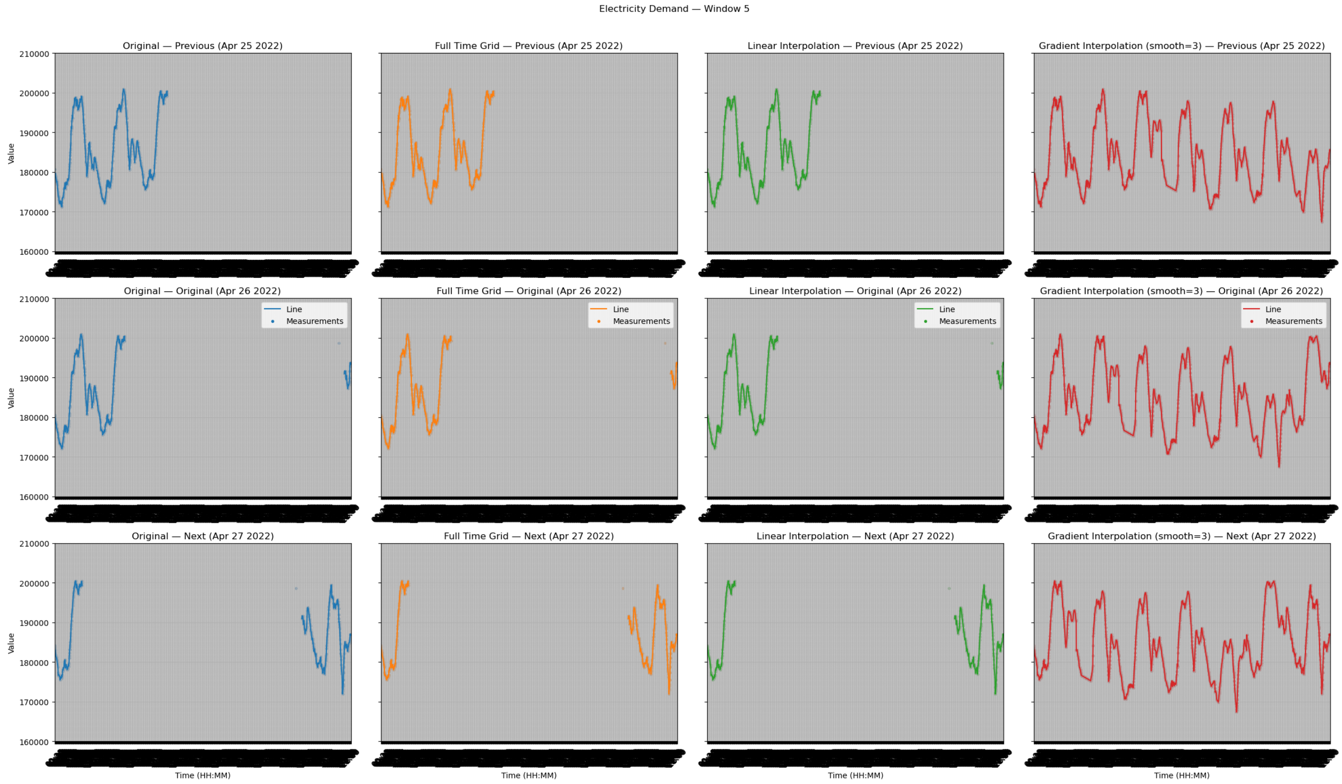

Figure S20. Gap Filling Generation Data, 2024-04-05

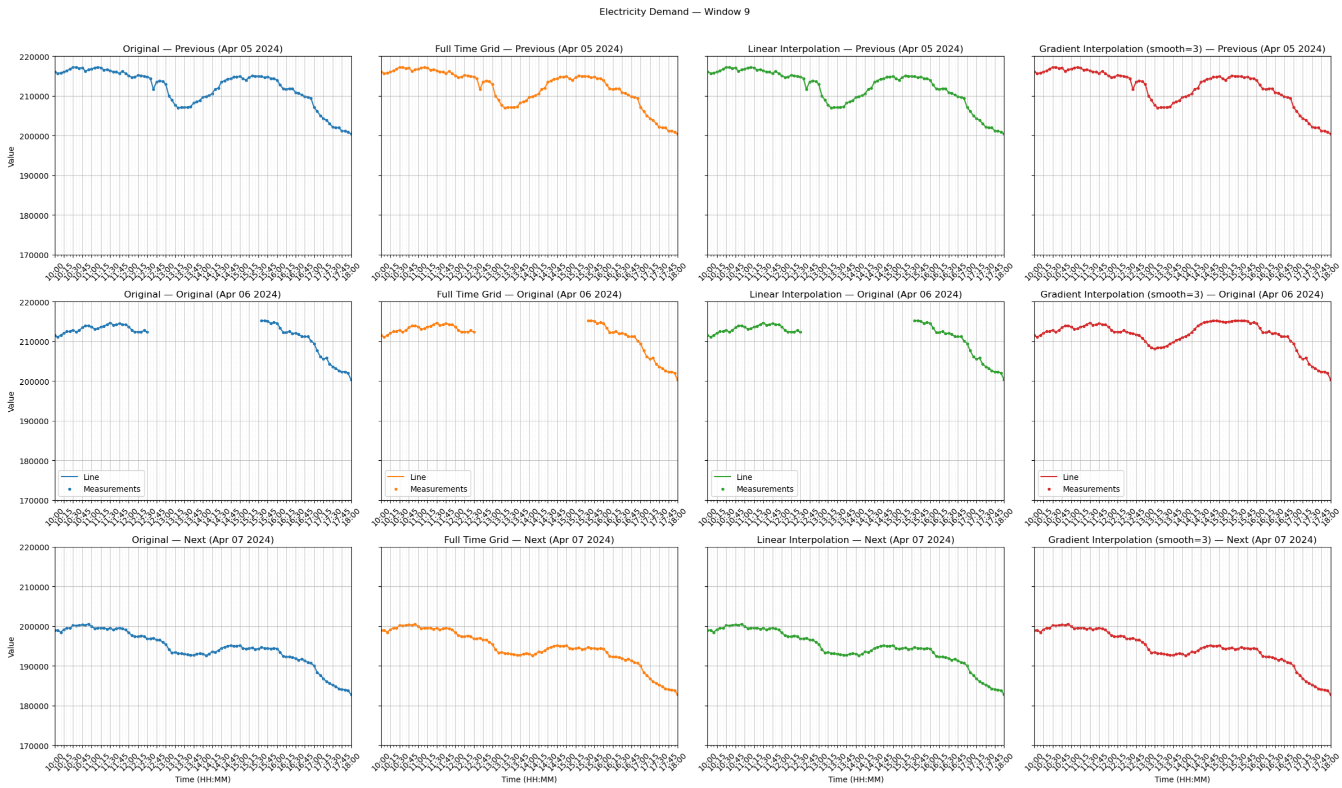

Figure S21. Histogram Distributions of Generation Data

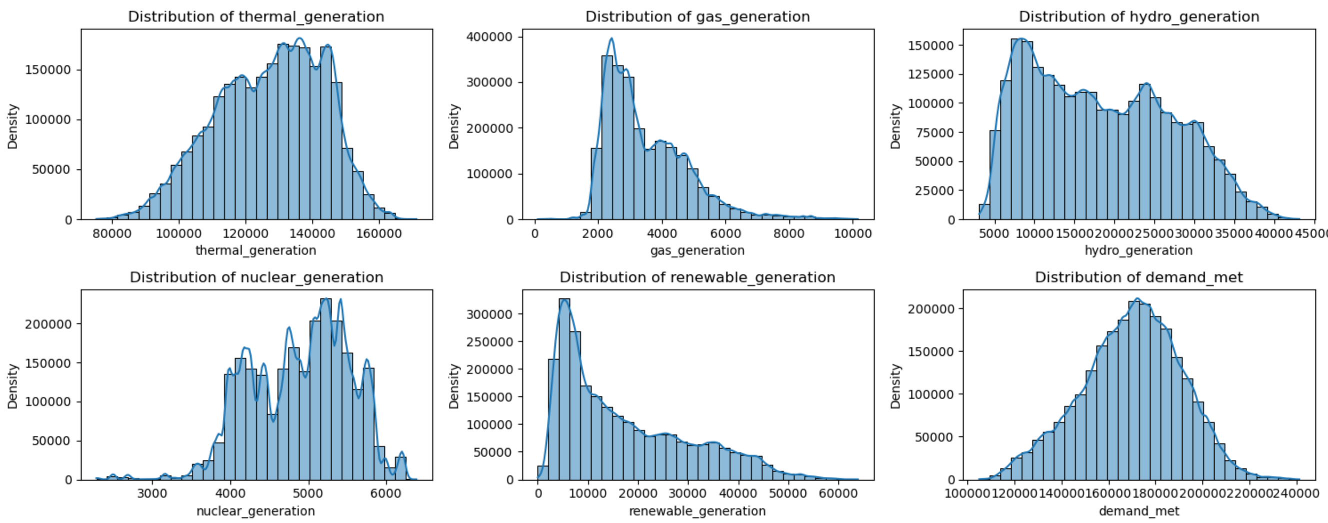

Figure S22. Histogram Distributions of ERA5 weather data

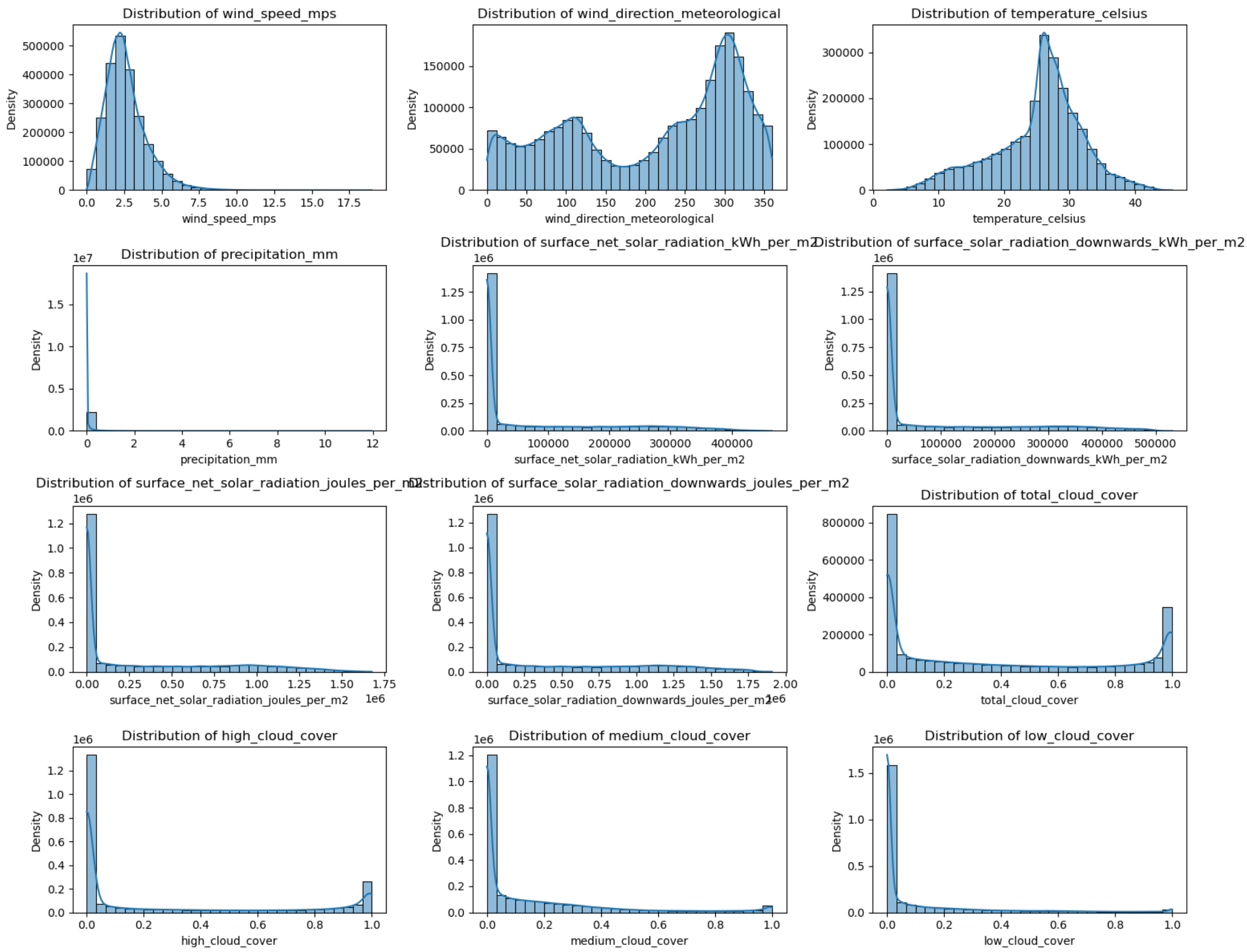

Figure S23. Histogram Distribution of Tons of CO_2 (Generation)

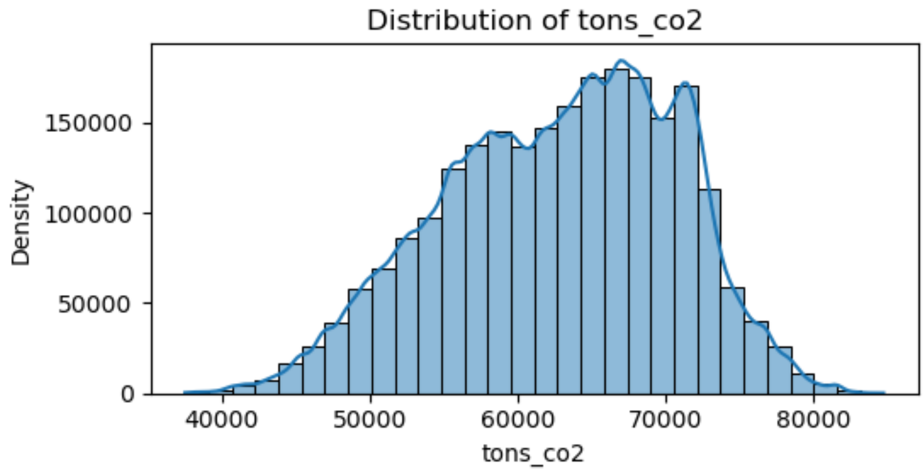

Figure S24. Histogram Distribution of Wind Speed post Log-Transform

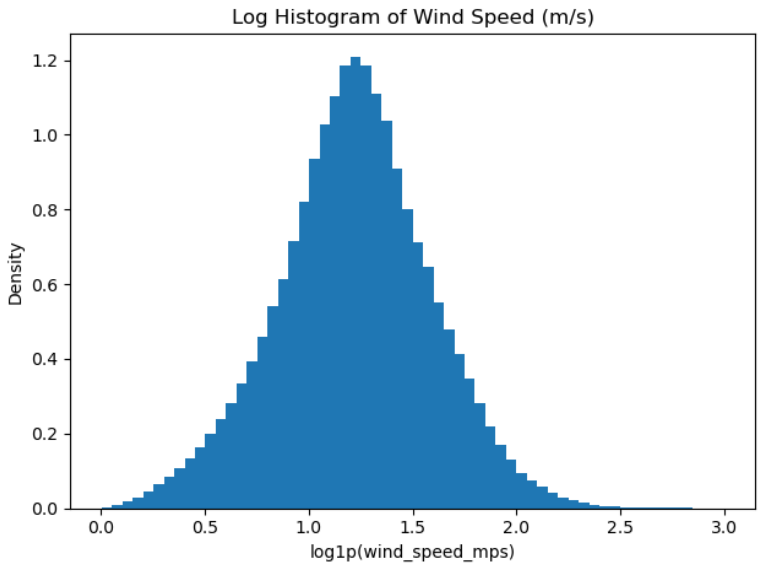

Figure S25. Seasonal Patterns of Temperature – Average Daily Profile

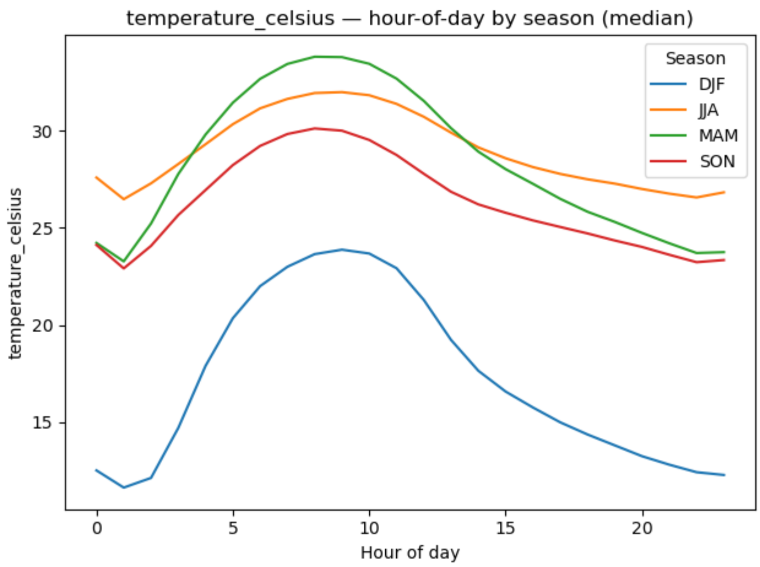

Figure S26. Seasonal Patterns of Wind Speed – Average Daily Profile

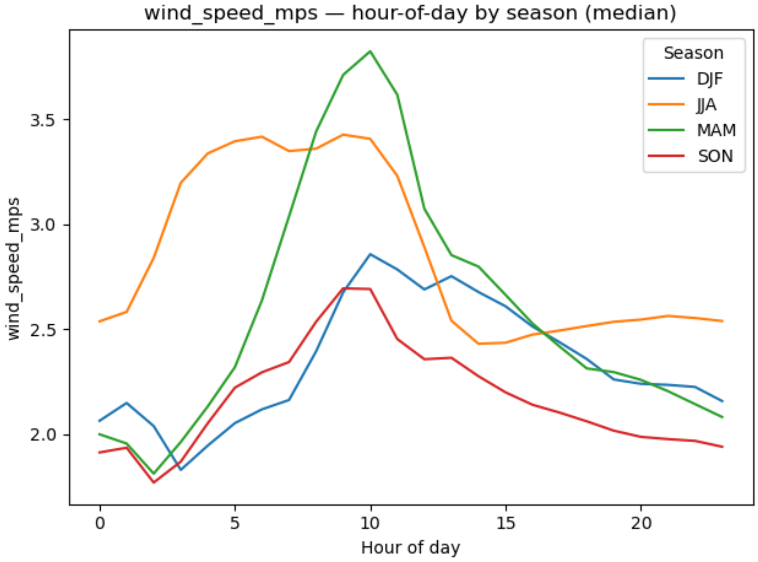

Figure S27. Seasonal Patterns of Solar Radiation – Average Daily Profile

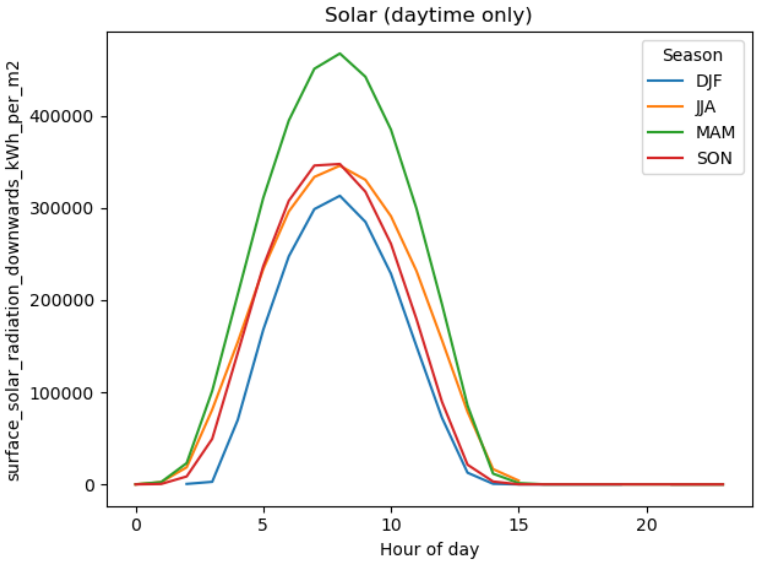

Figure S28. Seasonal Patterns of Wind Direction – Average Daily Profile

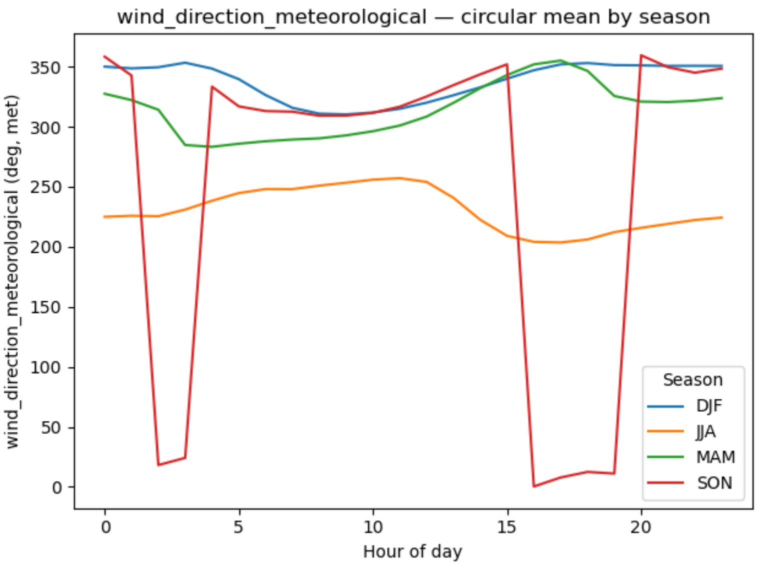

Figure S29. Covariance Matrix of Core Predictor Variables (untransformed)

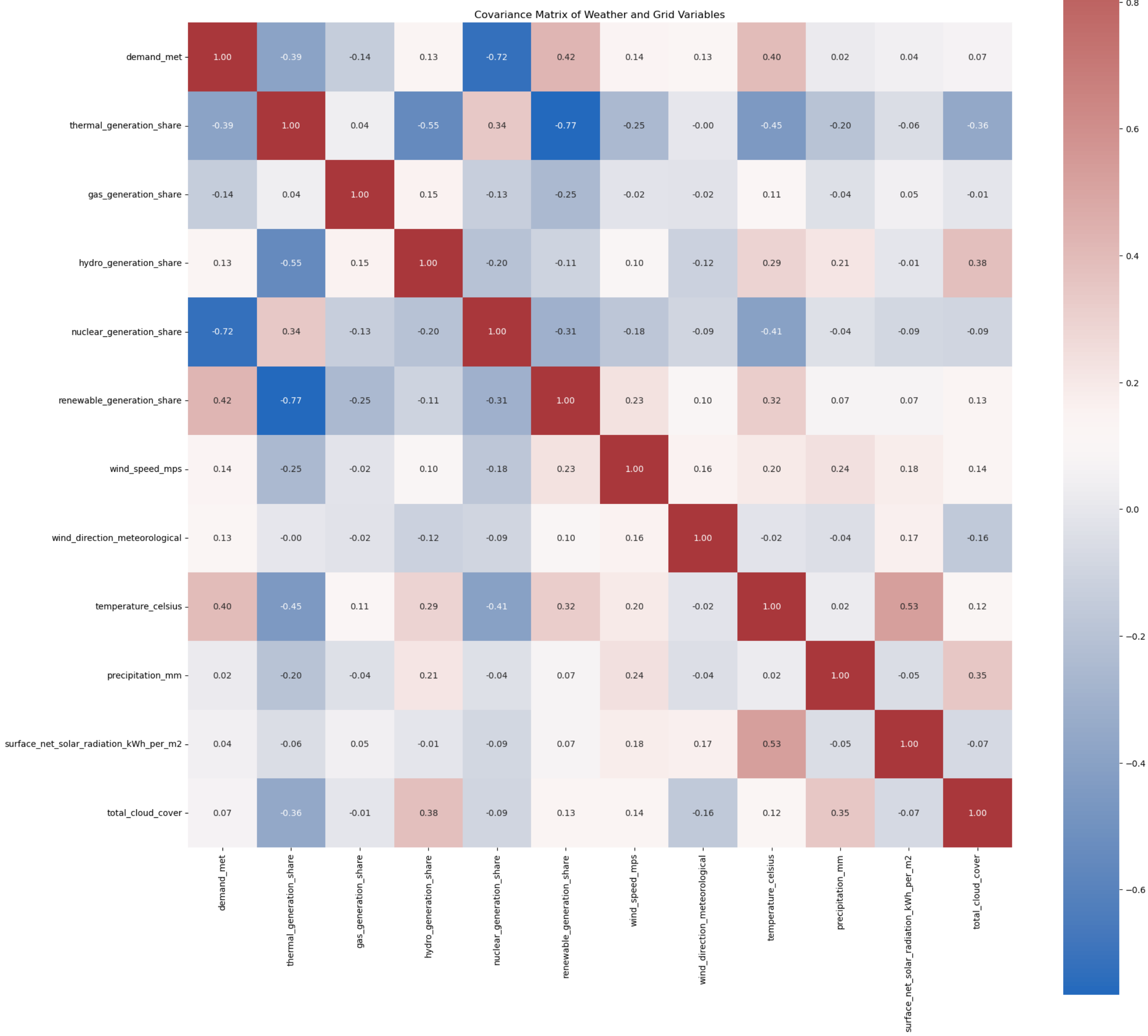

Figure S30. Hitachi Database Schema

weather	
timestamp	TIMESTAMP
wind_speed	NUMERIC
wind_direction	NUMERIC
temperature	NUMERIC
city	VARCHAR(6)
precipitation	NUMERIC
location	NULL
surface_net_solar_radiation	NUMERIC
surface_solar_radiation_downwards	NUMERIC

meter_readings	
ca_id	VARCHAR
date	TIMESTAMP
value	DOUBLE PRECISION
city	VARCHAR

grid_readings	
timestamp	TIMESTAMP
thermal_generation	DOUBLE PRECISION
gas_generation	DOUBLE PRECISION
g_co2_per_kwh	DOUBLE PRECISION
hydro_generation	DOUBLE PRECISION
nuclear_generation	DOUBLE PRECISION
renewable_generation	DOUBLE PRECISION
tons_co2	DOUBLE PRECISION
total_generation	DOUBLE PRECISION
tons_co2_per_mwh	DOUBLE PRECISION
demand_met	DOUBLE PRECISION
net_demand	DOUBLE PRECISION

carbon_readings	
timestamp	TIMESTAMP
thermal_generation	NUMERIC
gas_generation	NUMERIC
g_co2_per_kwh	NUMERIC
hydro_generation	NUMERIC
nuclear_generation	NUMERIC
renewable_generation	NUMERIC
tons_co2	NUMERIC
total_generation	NUMERIC
tons_co2_per_mwh	NUMERIC
demand_met	NUMERIC
net_demand	NUMERIC

alembic_version	
version_num	VARCHAR(32)

readings	
id	INTEGER
date	DATE
ca_id	VARCHAR(15)
value	DOUBLE PRECISION
city	VARCHAR(6)

ca_id → id

customers	
id	VARCHAR(15)
location	NULL
city	VARCHAR(6)

spatial_ref_sys	
srid	INTEGER
auth_name	VARCHAR(256)
auth_srid	INTEGER
srtext	VARCHAR(2048)
proj4text	VARCHAR(2048)