electricity

MDPI

Article

Synthetic Load Profile Generation for Residential and Commercial Loads: A Comparative Study of Stochastic Models

Juan Jiménez , Ricardo Isaza-Ruget and Javier Rosero-García *

Electrical Machines & Drives Research Group (EM&D), Department of Electrical and Electronic Engineering, Universidad Nacional de Colombia, Bogotá 111321, Colombia
* Correspondence: jaroserog@unal.edu.co

Abstract

Synthetic load profiles are essential for distribution network planning, protection sizing, and demand-side management studies. However, most generation methods are validated on a single load typology, and their transferability remains unexamined. Two broad paradigms dominate the generation literature: data-driven approaches that learn the demand distribution directly from historical records (generative adversarial networks, diffusion models, Markov-chain generators) and bottom-up, physically motivated approaches that reconstruct demand from the superposition of discrete appliance ON/OFF events. Same-data, same-metric comparisons across these two families for structurally distinct load typologies remain absent from the literature, and this gap is the one this paper addresses. This paper presents a systematic comparison of three stochastic models (a first-order autoregressive (AR(1)) profile, a physically constrained ON/OFF event model, and a nonlinear-least-squares (NLS) calibrated variant) applied to two fundamentally different load typologies measured with a Class A power-quality recorder at 10-min resolution: a single-family residential dwelling (13 days, 1860 samples) and an institutional commercial building (9 days, 1333 samples). Evaluation spans six distributional statistics (mean, standard deviation, and the percentiles p_{50}, p_{90}, p_{95} and p_{99}), the Kolmogorov–Smirnov (KS) statistic, root mean square error (RMSE), and hourly variance profiles. In this two-site study, load typology, not model sophistication, emerges as the dominant factor shaping fit quality. The simple AR(1) reproduces all percentiles within 7% for the near-Gaussian commercial load (skewness 3.47). By contrast, no model reproduces the centre, the dispersion and the upper tail of the highly skewed residential load (skewness 6.56) simultaneously to within 10%. On the residential tail the AR(1) ensemble is the least biased (p_{99}: -5.8%) but by far the most dispersed across realizations (CV $= 15.3\%$), whereas the event-based models are more stable but biased, so model choice on skewed loads is a bias–stability trade-off rather than an accuracy ranking. The NLS-calibrated model attains near-exact residential p_{95} reproduction (-0.3%) and commercial upper-tail errors below 2%, at a computational cost roughly three orders of magnitude above the AR(1). The dynamic characterization of demand obtained from these models, including the magnitude and frequency of the detected events, provides elements that may be of interest for the sizing and operation of photovoltaic systems in the context considered. Based on these two cases, a preliminary recommendation matrix mapping models to engineering applications is proposed, motivating typology-specific model selection rather than universal approaches; broader validation on additional sites is identified as future work.

Academic Editors: Andreas Sumper and Farhad Shahnia

Received: 20 May 2026
Revised: 13 August 2026
Accepted: 1 September 2026
Published: 14 September 2026

Keywords: synthetic load profile; stochastic load model; residential demand; commercial demand; ON/OFF event model; nonlinear least squares calibration; distribution network planning; distributed photovoltaic generation; battery energy storage system; hosting capacity

1. Introduction

The global clean energy transition is accelerating, driven by national policies and international commitments that promote distributed energy resources, electric vehicles, and demand-response programs [1]. This transition is reshaping the operating conditions of low-voltage distribution networks, where rising adoption of photovoltaic generation and electric mobility introduces new sources of uncertainty in demand profiles [2]. Accurate planning of transformer capacity, protection coordination, and voltage regulation increasingly depends on realistic representations of end-user demand that capture not only average consumption but also the temporal variability and extreme peaks that drive equipment sizing [3,4]. Because multi-year field campaigns at every service point are impractical, synthetic load profile generators have become an essential tool for distribution engineers and researchers alike [5].

Synthetic load profiles are particularly critical for the design and optimization of distributed solar energy systems. Hosting capacity studies for rooftop photovoltaic (PV) installations require thousands of stochastic demand scenarios to evaluate reverse power flow and overvoltage conditions at the point of common coupling. The sizing of battery energy storage systems (BESS) coupled with PV generation depends on the statistical overlap between the solar irradiance profile and the prosumer load profile, an overlap that can only be characterized through realistic representations of demand variability across hours and seasons. Furthermore, the planning of solar microgrids in communities that lack prolonged measurement campaigns requires representative synthetic profiles for both residential and institutional loads. The present work contributes to this need in the context of the Microgrids project for flexible and efficient energy in Cundinamarca, Colombia, which motivates the development of validated stochastic generators applicable to low-voltage networks with distributed PV penetration. The Colombian wholesale electricity market and the operation of the National Interconnected System are administered by XM S.A. E.S.P., which publishes monthly operational and demand reports that provide the broader context for distribution-level studies such as the present one [6].

Several families of methods have been proposed, broadly falling into two groups depending on whether they learn the demand distribution from data or reconstruct it from the underlying physical process. Among data-driven approaches, generative adversarial networks [7–9] have demonstrated the ability to synthesize realistic profiles from historical datasets. More recently, conditional diffusion models [10] have extended this capability to multi-conditional generation. Markov-chain generators [11,12] offer a lighter alternative that preserves transition dynamics. A complementary line of work follows a physics-informed or bottom-up strategy, originally developed by Capasso et al. [13]. These methods decompose aggregate demand into individual appliance usage patterns and reconstruct the total load through stochastic superposition of ON/OFF events [14–17]. At the boundary between these two paradigms, top-down statistical models fit parametric distributions or autoregressive processes to measured profiles and sample new realizations [18–20]. These models require less data than deep generative approaches while offering more temporal structure than purely empirical resampling. Separately, voltage-dependent ZIP models characterize the static relationship between supply voltage and power consumption [21–23]. However, their deterministic nature limits their use for temporal scenario generation.

Finally, ensemble machine-learning methods and decomposition-based approaches have targeted short-term load forecasting rather than profile generation [24–26], addressing a related but distinct problem. A recent systematic review of 169 works on synthetic energy time series [27] confirms that most published generators fall within the bottom-up or direct-generation families and predominantly target residential consumption, while a consolidated protocol for evaluating fidelity across generator families has yet to be adopted.

Despite this diversity of techniques, two important gaps remain insufficiently addressed in combination, even though each has received partial attention in isolation: the systematic review cited above [27] documents that most published generators are validated on a single typology, and [28] shows that no consensus has emerged on the minimal metric set for comparing generator families; neither work addresses both limitations jointly. First, most studies validate their proposed generator on a single load typology, typically residential, without examining whether the same model performs equally well on structurally different loads such as commercial or institutional buildings. The statistical properties of residential demand (high skewness, heavy tails, sporadic high-power events) differ fundamentally from those of commercial loads (lower skewness, bimodal distribution, schedule-driven patterns), and these differences can alter model rankings in ways that single-typology studies cannot reveal. Second, comparative studies that evaluate multiple model families on the same dataset using the same metrics remain scarce [5,26]. In their absence, practitioners lack quantitative guidance for selecting the appropriate generator for a given engineering application. Even among the comparisons that do exist, similarity is often judged with a limited set of measures, and no consensus has emerged on the minimal set of metrics required to assess representativeness across generator families [28].

This paper addresses both gaps through a systematic, same-data comparison of three stochastic load profile generators applied to two fundamentally different load typologies measured in Bogotá, Colombia. The specific contributions are as follows:

1. We perform a comparative evaluation of three models of increasing complexity (a first-order autoregressive (AR(1)) hourly profile, a physically constrained ON/OFF event model, and a nonlinear-least-squares (NLS) calibrated variant) on residential and commercial loads recorded with a Class A power-quality analyzer at 10 min resolution.
2. Quantitative evidence from two representative sites that load typology, rather than model sophistication, is the dominant factor determining fit quality for these cases: the near-Gaussian commercial load (skewness = 3.47) allows a simple AR(1) to reproduce all percentiles within 7%, whereas the highly skewed residential load (skewness = 6.56) exposes a bias–stability trade-off in which the AR(1) ensemble provides the least biased upper-tail estimate but the event-based models produce more stable, systematically biased realizations.
3. We offer a practical recommendation matrix that maps each model to the engineering applications for which it is best suited, enabling typology-specific model selection for distribution network studies, including hosting capacity assessment for distributed PV and BESS sizing in solar microgrids [29,30].
4. Three methodological elements support this comparison and are transferable to other studies: (i) a formulation of the ON/OFF event model as an inverse optimization problem calibrated by nonlinear least squares with Monte Carlo simulation embedded in each function evaluation, using a parameterization that resolves the arrival process by time block and the event distributions by event class, with a smoothness penalty across adjacent blocks; (ii) an explicit set of seven physical realism constraints for the event-based generator, whose variance penalty is shown to be typology-dependent; and (iii) a multi-dimensional evaluation protocol combining distributional statistics,

the Kolmogorov–Smirnov statistic, RMSE, hourly variance profiles and joint hour–power density heatmaps, applied identically to the three models on both typologies.

The remainder of this paper is organized as follows. Section 2 describes the measurement campaigns, the exploratory data analysis, and the three model formulations. Section 3 presents the results for each typology and the cross-comparison. Section 4 discusses the role of load typology, model trade-offs, practical recommendations, and limitations. Section 5 summarizes the key findings and outlines future work.

2. Materials and Methods

Comparing stochastic generators requires a common empirical basis: the same field data, the same descriptive statistics, and the same evaluation metrics applied to every model. The three generators produce distributional, not predictive, synthetic series: individual realizations are not time-synchronized with the measured record, so point-wise error metrics such as RMSE are of order σ by construction and are reported here only as a secondary indicator of fit, not a primary one. To make this explicit, two random reference generators without any temporal structure are included as a lower bound throughout the evaluation: Baseline A resamples the empirical marginal of P_{total} with replacement (i.i.d., no notion of hour or day), and Baseline B draws $\mathcal{N}(\mu, \sigma)$ white noise using the measured mean and standard deviation, clipped at $P \geq 0$. Because both baselines can reproduce the marginal almost exactly while carrying no temporal information, two metrics sensitive to temporal structure are also reported: the hourly variance ratio (synthetic-to-real variance in each hour of day) and the distribution of ramps $\Delta P = P(t) - P(t-1)$; both are used in Section 3.4 to show what a purely marginal fit does, and does not, guarantee. Two measurement campaigns conducted in Bogotá, Colombia, provide this basis; their exploratory analysis reveals the distributional contrasts that motivate the choice of three progressively complex formulations: an AR(1) hourly profile, a physically constrained ON/OFF event model, and an NLS-calibrated variant. The three formulations produce dynamic representations of demand that, beyond the comparative study reported here, may also serve as inputs for subsequent analyses of photovoltaic generation in low-voltage distribution settings.

2.1. Measurement Campaign

The purpose of the two campaigns is not to sample a population of residential or commercial demand, nor to derive an average behavior representative of a demand class. It is to obtain two well-characterized reference signals with structurally different distributional properties, against which the three generators can be calibrated and evaluated on a per-load basis. Averaging across sites would smooth out the discrete switching events and heavy-tailed excursions whose faithful reproduction is the central difficulty for the models compared here, so the parameters reported below characterize these two specific loads and are not offered as representative of residential or commercial populations.

Two measurement campaigns were conducted in Bogotá, Colombia, using a DRANETZ HDPQ Visa Class A power-quality analyzer compliant with IEC 61000-4-30 [31]. The instrument recorded three-phase RMS voltage (V_{rms}), RMS current (I_{rms}), fundamental active power (P_{Fnd}), fundamental reactive power (Q_{Fnd}), and displacement power factor at a 10 min aggregation interval. Table 1 summarizes both campaigns. The wiring configuration adopted for each campaign is shown in Figures 1 and 2.

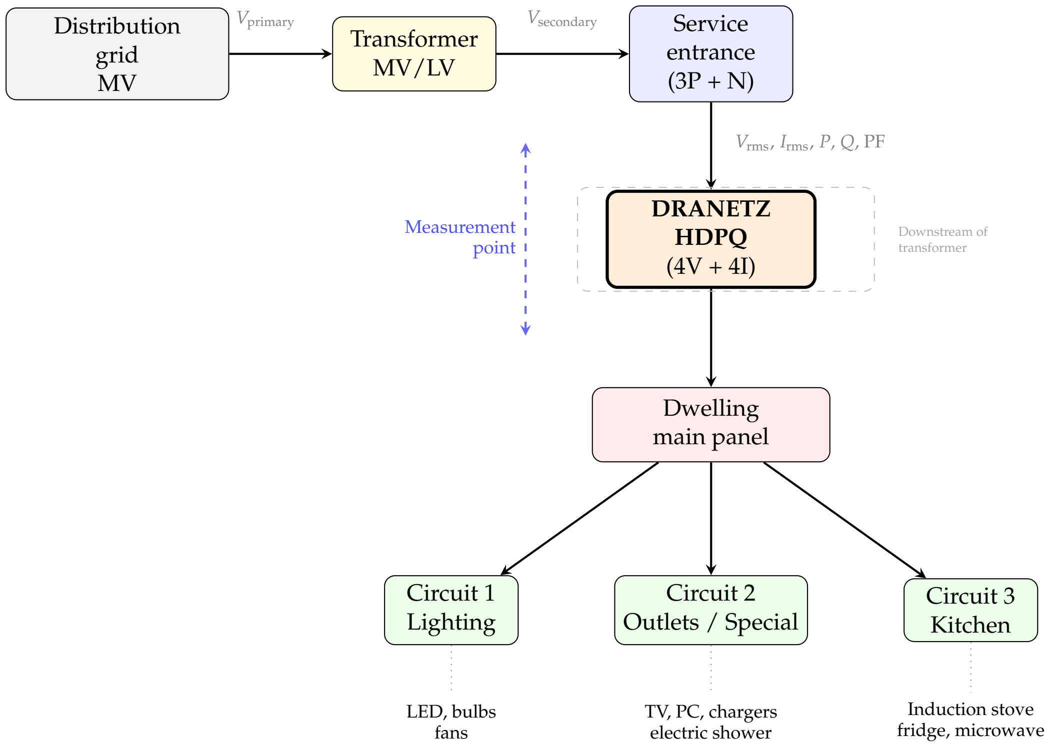

Figure 1. Single-line diagram of the residential measurement setup. The DRANETZ HDPQ recorder is placed at the dwelling main panel, downstream of the distribution transformer, capturing the aggregate total demand. Solid arrows indicate power flow, dotted grey lines link each circuit to its loads, and the dashed blue arrow marks the measurement point; box colours only distinguish element types, and the dashed grey rectangle delimits the zone downstream of the transformer.

Table 1. Measurement campaign parameters.

Parameter	**Residential**	**Commercial**
Location	Service entrance, single-family dwelling	LV transformer, LABE laboratory (UNAL)
Period	14–27 November 2025	20–29 October 2025
Duration	13 days	∼9 days
Samples (N)	1860	1333
Phases	A, B, C	A, B, C

The DRANETZ HDPQ Visa is a Class A instrument per IEC 61000-4-30, with a manufacturer-specified voltage accuracy of ±0.1% of reading and an active-power accuracy of class 0.2S. The dominant term in the combined uncertainty budget is the current probe (Dranetz TR2500 flexible Rogowski coil, ±1% of reading); propagating the voltage, current and power-factor channel uncertainties in quadrature yields a combined standard uncertainty of approximately 1% and an expanded uncertainty of approximately 2% ($k = 2$). The expanded instrumental uncertainty is approximately 2%. This is small relative to the largest model discrepancies, which reach 10–40%, but comparable to several of the smallest commercial errors. It therefore does not alter the broad typology-level conclusions, although differences below approximately 2% should not be interpreted as instrument-resolved rankings. All series analyzed in this study are field measurements acquired with a Class A instrument, not simulated inputs; access to the underlying data is described in the Data Availability Statement.

Figure 1 shows the single-line diagram of the residential measurement setup. The DRANETZ HDPQ recorder was installed at the dwelling main panel, immediately downstream of the distribution transformer and upstream of all branch circuits, so that it captures the aggregate total demand including all loads connected to the derived circuits.

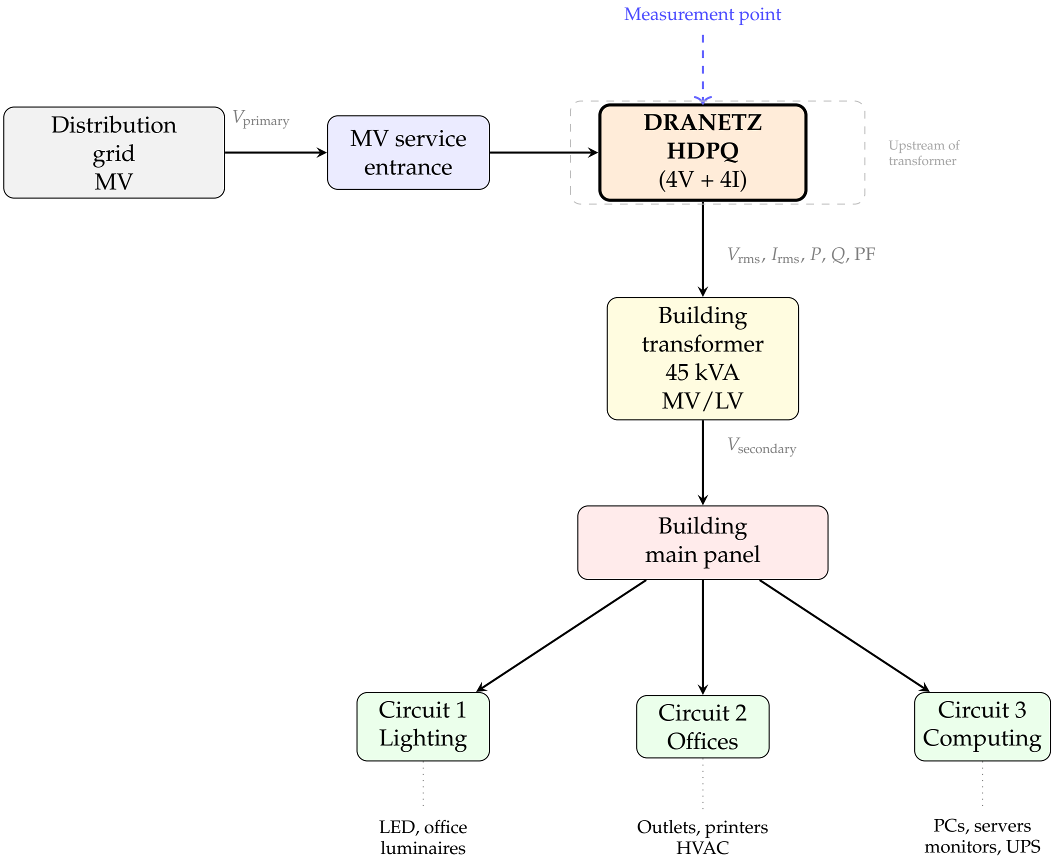

Figure 2. Single-line diagram of the commercial measurement setup. Unlike the residential case, the DRANETZ HDPQ recorder is placed upstream of the building's own 45 kVA transformer, measuring the aggregate total demand on the medium-voltage side and including transformer losses. Solid arrows indicate power flow, dotted grey lines link each circuit to its loads, and the dashed blue arrow marks the measurement point; box colours only distinguish element types, and the dashed grey rectangle delimits the zone upstream of the transformer.

Figure 2 presents the single-line diagram of the commercial measurement setup. Unlike the residential case, the DRANETZ HDPQ recorder was installed upstream of the building's own 45 kVA transformer, capturing the aggregate total demand on the medium-voltage side and including transformer losses as well as all loads supplied by the building's internal circuits (offices, computing equipment, lighting, and auxiliary services).

The residential site represents a typical single-family household in Bogotá, with occupancy patterns driven by individual appliance switching (cooking, water heating, lighting). The commercial site is an institutional building housing laboratories, offices, and server rooms at the Universidad Nacional de Colombia; its consumption follows fixed business hours (approximately 07:00–18:00) with a permanent base load from servers, security, and emergency lighting.

From the per-phase measurements, three aggregate variables were computed for all subsequent analyses:

$$P_{\text{total}}(t) = P_A(t) + P_B(t) + P_C(t) \quad [\text{kW}], \tag{1}$$

$$V_{\text{avg}}(t) = \frac{V_A(t) + V_B(t) + V_C(t)}{3} \quad [\text{V}]. \tag{2}$$

The neutral/fourth phase was excluded from all calculations.

2.2. Exploratory Data Analysis

Table 2 compares the descriptive statistics of P_{total} for both loads. The two typologies exhibit fundamentally different distributional properties.

Table 2. Descriptive statistics of P_{total}.

Statistic	Residential	Commercial
Mean [kW]	0.33	2.46
Std [kW]	0.57	2.65
Min [kW]	0.05	0.77
Max [kW]	7.98	16.59
p_{50} [kW]	0.22	1.37
p_{90} [kW]	0.63	5.00
p_{95} [kW]	1.17	6.23
p_{99} [kW]	2.69	15.63
Skewness	6.56	3.47
Excess kurtosis	61.41	13.84
V_{avg} [V]	125.56	117.34
V_{avg} range [V]	123.39–127.81	114.50–119.57

The residential load has a mean of 0.33 kW but a median of only 0.22 kW. This difference confirms extreme positive skewness. The household operates in a low-consumption regime most of the time (standby, refrigerator), punctuated by short, high-power events (electric shower, oven, iron) that reach up to 8.0 kW. The resulting skewness of 6.56 and excess kurtosis of 61.41 indicate a distribution that is far from Gaussian, with very heavy right tails. The hourly profile follows a characteristic double-hump pattern with morning (05:00–07:00) and evening (17:00–20:00) peaks.

The commercial load has a mean of 2.46 kW, which is 7.4× larger than the residential mean. The minimum of 0.77 kW reflects the permanent server and security base load. The commercial distribution is bimodal, with a narrow peak at $\sim$1 kW (night/weekend base) and a broader plateau at 2–6 kW (daytime operation). The skewness (3.47) and kurtosis (13.84) are substantially lower than those of the residential case. The superposition of many simultaneously operating office devices produces a distribution closer to Gaussian. The hourly profile shows a daytime plateau (08:00–17:00) governed by the institutional schedule, in contrast to the residential double-hump pattern.

These structural differences have direct implications for model selection, as discussed in Section 4.

Before formalizing the three stochastic models, a scoping decision deserves explicit justification. A voltage-dependent ZIP formulation was also considered but is excluded from the present comparison because its deterministic, static nature addresses a different question (the $P = f(V)$ relationship for power-flow studies) rather than temporal profile generation. In addition, the narrow voltage range observed at both sites (3.5–4.3% of nominal) prevents reliable identification of the ZIP coefficients, as discussed in Section 4.4.

2.3. Model 1: Hourly Profile with AR(1) Noise

The simplest generator constructs a synthetic series by sampling from the empirical hourly profile with temporally correlated noise. For each timestamp t falling in hour h and day-of-week d, the synthetic power is

$$P_{\text{syn}}(t) = \mu(h, d) + \sigma(h, d) \cdot z_t, \tag{3}$$

where $\mu(h,d)$ and $\sigma(h,d)$ are the sample mean and standard deviation of P_{total} computed over all observations in the (h,d) cell, forming a $24 \times 7 = 168$-element lookup table. A hierarchical fallback is applied when a cell has fewer than three observations: first to the hourly marginal, then to the global mean.

The noise process z_t is a stationary AR(1):

$$z_t = \rho\, z_{t-1} + \sqrt{1-\rho^2}\, \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1), \tag{4}$$

where $\rho = 0.85$ controls the temporal autocorrelation. The parameter ρ was not fitted to the measured autocorrelation, which is considerably lower (lag-1 autocorrelation of the deseasonalized residual: 0.29 residential, 0.41 commercial); it acts as a temporal-smoothness parameter rather than a distributional one, since the stationary AR(1) noise of Equation (4) has unit variance for any ρ and the marginal distribution of P_{syn} is therefore invariant to it. A 20-seed sensitivity analysis over $\rho \in [0.30, 0.92]$ confirms this: for every distributional statistic and both typologies, the range of the ensemble means across ρ is smaller than the dispersion between seeds at fixed ρ (ratios 0.1–0.8), so no value of ρ in this range is distinguishable from another in terms of marginal fit. The value $\rho = 0.85$ was retained as a mid-range reference. Matching the measured short-lag autocorrelation would be a relevant refinement for applications in which trajectory dynamics matter, such as storage operation studies. The factor $\sqrt{1-\rho^2} \approx 0.53$ ensures unit stationary variance. A physical clipping constraint $P_{\text{syn}} \geq 0$ is enforced.

This model has the advantage of reproducing the hourly mean and standard deviation by construction, since the profile is copied directly from the data. Its main limitation is the Gaussian innovation assumption. The symmetric, light-tailed nature of Gaussian noise prevents the generation of the sharp, isolated peaks observed in residential demand.

2.4. Model 2: Physically Constrained ON/OFF Events

Whereas the AR(1) model captures the temporal autocorrelation of demand under a Gaussian-noise assumption, the second formulation departs from this assumption by reconstructing the load from its discrete generating process. Following the bottom-up paradigm pioneered by Capasso et al. [13], the ON/OFF model decomposes aggregate demand into a continuous base load and a superposition of discrete rectangular events representing appliance activations:

$$P(t) = P_{\text{base}}(t) + \sum_k A_k \cdot \operatorname{rect}(t - t_k,\ D_k), \tag{5}$$

where $P_{\text{base}}(t) = \mu(h,d)$ is the same hourly profile used in the AR(1) model, A_k is the amplitude of event k (in kW), D_k is its duration (in samples), and t_k is its start time drawn from a non-homogeneous Poisson process with hourly rate λ_h.

2.4.1. Event Detection

Events are detected on the deseasonalized residual $r(t) = P_{\text{total}}(t) - \mu(h,d)$. Contiguous segments where $r(t)$ exceeds the 90th percentile threshold are identified. Adjacent segments separated by fewer than two samples are merged, and segments shorter than two samples (20 min) are discarded. The detected events are classified into three amplitude categories, small (S), medium (M), and large (L), using tercile boundaries computed from the observed amplitude distribution. Table 3 summarizes the detection results.

The residential load produces frequent, short, high-amplitude events (household appliances switched on briefly), whereas the commercial load exhibits fewer but longer events (laboratory equipment or HVAC operating for hours). Figure 3 visualizes the detected events as a scatter plot of amplitude versus duration, colored by category (S/M/L). The figure illustrates the contrasting event signatures of both typologies.

Table 3. Event detection summary for both load typologies.

Parameter	Residential	Commercial
Detection threshold [p_{90} of $r(t)$]	0.21 kW	0.54 kW
Events detected	55	14
Amplitude range	0.30–5.30 kW	0.56–2.28 kW
Amplitude median	0.60 kW	1.26 kW
Duration median	2 samples (20 min)	5 samples (50 min)
Duration max	9 samples (90 min)	44 samples (7.3 h)
Classification: S/M/L	35/16/4 (63.6/29.1/7.3%)	3/8/3 (21.4/57.1/21.4%)

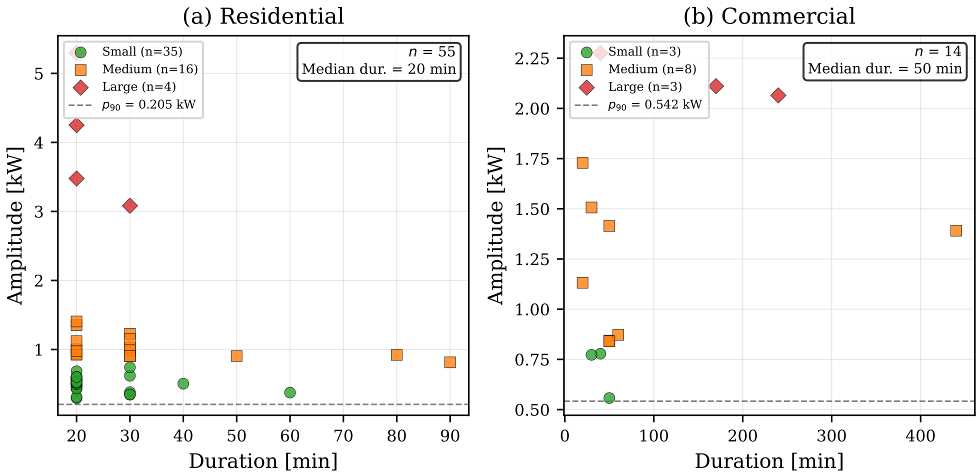

Figure 3. Detected events for both load typologies, amplitude versus duration, colored by category (Small, Medium, Large). The dashed line indicates the p_{90} detection threshold. Panels correspond to (**a**) the residential load and (**b**) the commercial load; the inset in each panel reports the number of detected events and the median event duration.

2.4.2. Physical Constraints

Seven physical realism constraints are incorporated in the model: (i) the Poisson arrival rate λ_h is smoothed across adjacent hours to eliminate spurious fluctuations; (ii) nocturnal suppression sets $\lambda_h \approx 0$ during quiet hours (00:00–04:00 for residential); (iii) event amplitudes are capped per category using the observed maximum of each class; (iv) event durations follow a lognormal distribution fitted to the detected events, with parameters $\mu_{\ln}$ and $\sigma_{\ln}$; (v) amplitudes are scaled proportionally to the base-load level at the event's start time; (vi) a Monte Carlo ensemble of $K = 50$ realizations is generated to produce confidence bands; and (vii) advanced validation diagnostics (hourly variance ratio, joint hour–power heatmaps) are computed.

The lognormal duration parameters are $\mu_{\ln} = 0.88$, $\sigma_{\ln} = 0.35$ for the residential load (median duration $e^{0.88} \approx 2.4$ samples ≈ 24 min) and $\mu_{\ln} = 1.74$, $\sigma_{\ln} = 0.88$ for the commercial load (median ≈ 57 min).

2.5. Model 3: NLS Batch Calibration

The empirical parametrization of the ON/OFF model produces physically plausible realizations but does not formally optimize any global fit criterion. The third formulation closes this gap by recasting the same generative structure as a calibration problem. The NLS model retains the ON/OFF event structure of Equation (5) but replaces the direct empirical parametrization with a formal optimization that minimizes the discrepancy between a vector of target statistics computed from real data and the corresponding statistics averaged over K Monte Carlo simulations of the generative model.

Let $\boldsymbol{\theta} \in \mathbb{R}^{24}$ denote the parameter vector. It resolves the arrival process in time and the event distributions by class: six Poisson rates λ_b and six mixing probabilities $p_{\text{med}}^{(b)}$, indexed by six 4-h blocks b covering the day (00–03, 04–07, …, 20–23), together with three pairs of lognormal amplitude parameters $(\mu_A^{(c)}, \sigma_A^{(c)})$ and three pairs of lognormal duration parameters $(\mu_D^{(c)}, \sigma_D^{(c)})$, indexed by the event category $c \in \{\text{small}, \text{medium}, \text{large}\}$, for a total of $6 + 6 + 6 + 6 = 24$ free scalars. This split keeps the dimensionality low while preserving both the intraday variation of the arrival process and the distinct amplitude and duration signatures of the three event classes. The remaining mixing probability $p_{\text{large}}^{(b)}$ is not optimized: it is estimated once per block from the empirical proportion of large events detected there and held fixed throughout the calibration, which reduces the free dimension from 30 to 24.

The target vector $\boldsymbol{S}_{\text{real}} \in \mathbb{R}^{59}$ contains 54 distributional statistics (hourly means, hourly variances, global percentiles (p_{50}, p_{90}, p_{95}, p_{99}), the overall mean and standard deviation) plus 5 regularization terms ($\lambda_{\text{reg}} = 0.5$) that penalize the differences between the arrival rates of adjacent blocks, $\sqrt{\lambda_{\text{reg}}}\,(\lambda_{b+1} - \lambda_b)$ for $b = 1, \ldots, 5$, so that the fitted diurnal profile of the arrival process remains smooth.

The calibration solves

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \left\| \boldsymbol{S}_{\text{real}} - \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{S}_{\text{sim}}^{(k)}(\boldsymbol{\theta}) \right\|^2, \tag{6}$$

using the Trust Region Reflective algorithm implemented in `scipy.optimize.least_squares` [32] with box constraints on all parameters. During optimization, $K = 15$ Monte Carlo runs are averaged per function evaluation to reduce gradient noise. For the final evaluation, $K = 50$ runs are used. The optimizer is allowed up to 300 function evaluations.

This approach is related to indirect inference [33] and the method of simulated moments [34,35], where model parameters are estimated by matching simulated auxiliary statistics to their empirical counterparts. The key advantage of the NLS formulation is that it directly optimizes the upper-tail percentiles (p_{95}, p_{99}) that are critical for protection sizing and transformer dimensioning. The cost is higher computational expense and a tendency to overestimate the median due to the event-superposition structure shared with the ON/OFF model.

3. Results

Each model was applied to both load typologies using the same model structures, calibration procedures and evaluation metrics. The evaluation combines multiple families of similarity measures—central tendency, dispersion, distributional distance, and temporal correlation—because no single metric captures all dimensions of representativeness [28]. All three models are reported as the median of $K = 50$ Monte Carlo realizations generated with seeds 42–91, so that no result is conditioned on a single random draw. The residential and commercial cases are presented separately to expose the typology-dependent behavior before being consolidated in a unified error matrix.

3.1. Residential Load

Table 4 compares the distributional statistics of the three synthetic series against the measured residential demand.

https://doi.org/10.3390/electricity7030105

The AR(1) model overestimates the mean by 12.9% and underestimates the standard deviation by 12.4%, and its Gaussian innovations cause it to overestimate p_{90} by 28.7%. The model generates too many moderately high values and fails to produce the isolated extreme peaks that characterize the real data. Its ensemble median nonetheless reproduces the upper tail with the smallest bias of the three models (p_{95}: −0.8%; p_{99}: −5.8%), because the marginal of each hourly cell is copied from the data by construction. This accuracy is unstable: across the 50 realizations the p_{99} estimate ranges from 1.74 kW to 3.56 kW (CV = 15.3%), so an individual AR(1) trace may misrepresent the tail by more than 25% in either direction. The synthetic time series (Figure 4) shows smooth, continuous fluctuations that follow the circadian pattern. In contrast, the measured series exhibits sharp, discrete transitions that the AR(1) cannot reproduce.

Table 4. Comparison of distributional statistics: residential load. All model columns report the median of $K = 50$ Monte Carlo realizations (seeds 42–91).

	Real	**AR(1)**	**Δ%**	**ON/OFF**	**Δ%**	**NLS**	**Δ%**
Mean [kW]	0.33	0.37	+12.9	0.36	+10.0	0.42	+25.8
Std [kW]	0.57	0.50	−12.4	0.31	−45.6	0.46	−18.8
p_{50} [kW]	0.22	0.25	+14.6	0.28	+29.7	0.30	+37.6
p_{90} [kW]	0.63	0.82	+28.7	0.67	+6.1	0.79	+24.5
p_{95} [kW]	1.17	1.16	−0.8	0.96	−18.6	1.17	−0.3
p_{99} [kW]	2.69	2.54	−5.8	1.73	−35.7	2.32	−13.7
RMSE [kW]	–	–	–	0.54	–	0.64	–
KS statistic	–	–	–	0.25	–	0.26	–

Note: absolute error in kW (synthetic median of $K = 50$ realizations minus measured) for the p_{50}–p_{99} rows: AR(1) +0.03, +0.18, −0.01, −0.16; ON/OFF +0.06, +0.04, −0.22, −0.96; NLS +0.08, +0.15, −0.00, −0.37 (kW, in row order). Normalized by the measured standard deviation ($\sigma = 0.57$ kW), the upper-tail errors are AR(1) p_{90} $+0.32\sigma$, p_{95} -0.02σ, p_{99} -0.27σ; ON/OFF p_{90} $+0.07\sigma$, p_{95} -0.38σ, p_{99} -1.69σ; NLS p_{90} $+0.27\sigma$, p_{95} -0.01σ, p_{99} -0.65σ. Dashes (–) indicate statistics that are not applicable to the corresponding model.

The event-based formulation produces a qualitatively different tail behavior. The ON/OFF model reduces the p_{90} error to only +6.1%. This result demonstrates that the event superposition mechanism better captures the transition zone between base consumption and moderate peaks. However, the physical constraints of the ON/OFF model (nocturnal suppression and amplitude caps) severely compress the variance (−45.6%). As a consequence, the ON/OFF model underestimates p_{99} by 35.7%. This variance penalty is the central trade-off of the constrained ON/OFF approach for highly skewed loads. The restrictions that eliminate unrealistic events also remove the extreme events that generate most of the variance.

The NLS-calibrated model closes part of this variance gap and achieves near-exact p_{95} reproduction (−0.3%), with a p_{99} error of −13.7%. By explicitly including these percentiles in its objective function (Equation (6)), the optimizer adjusts the event amplitude and duration parameters to match the upper tail of the distribution. The cost is a systematic overestimation of the median (+37.6%) and mean (+25.8%). This bias is shared with all event-based models because the superposition of rectangular events on the base profile inflates the central region of the distribution. The standard deviation error (−18.8%) is substantially better than that of the ON/OFF model (−45.6%). This improvement confirms that the NLS calibration partially compensates the variance compression. Figure 5 compares the measured and synthetic residential distributions.

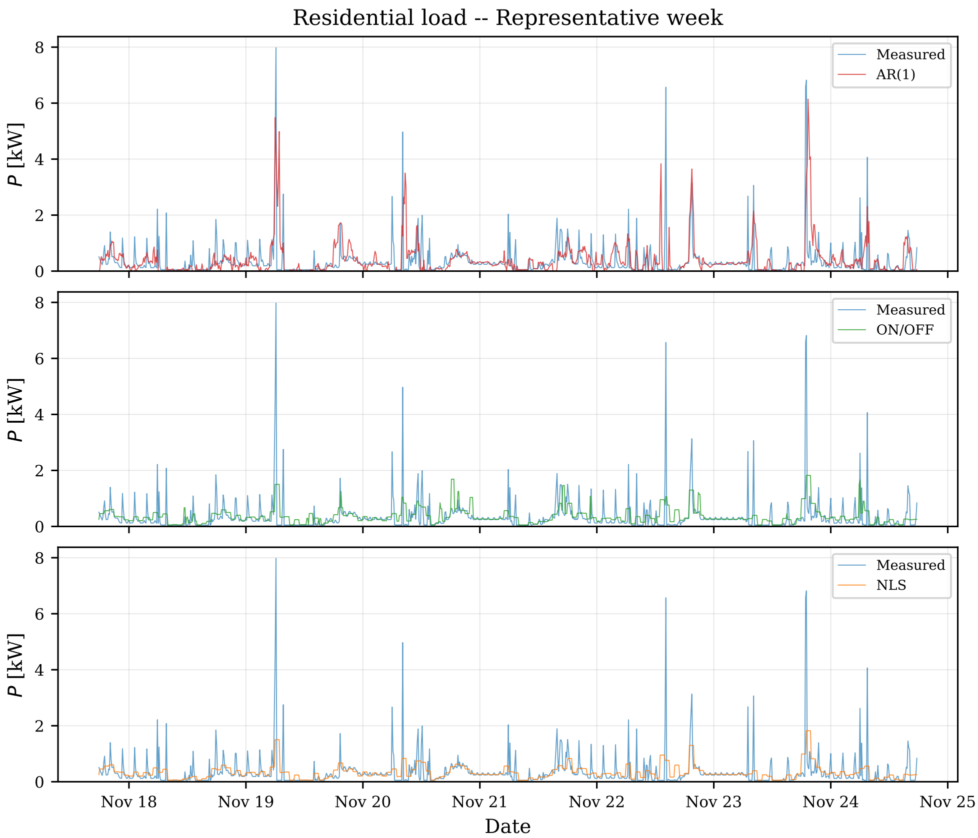

Figure 4. Residential load: representative time series showing measured P_{total} and synthetic series from AR(1), ON/OFF, and NLS models.

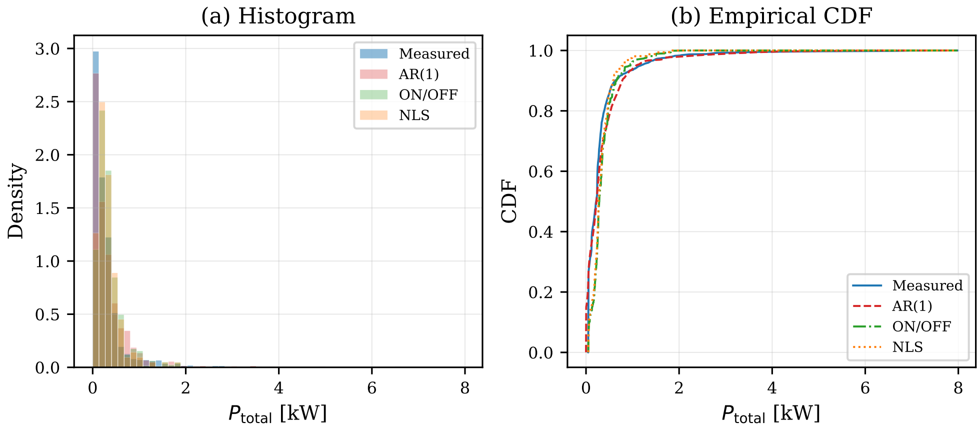

Figure 5. Residential load: histogram and empirical CDF comparing measured and synthetic distributions. Panel (**a**) shows the probability density histogram and panel (**b**) the empirical cumulative distribution function, in both cases for the measured series and for the AR(1), ON/OFF and NLS models.

Because all reported figures are ensemble medians, the dispersion of the ensemble itself quantifies how reproducible they are. Across the 50 realizations of each model the coefficient of variation of the reported statistics ranges from 0.4% to 15.3%. For the event-based models it is small: the median over their 24 model–statistic cells is 2.4%, and only one exceeds 10% (NLS residential p_{99}, 13.3%). The AR(1) ensemble is markedly more dispersed on the residential load, where three of its six cells exceed that threshold (Std: 12.2%; p_{95}: 10.0%; p_{99}: 15.3%), whereas all of its commercial cells remain below 3%. For the ON/OFF and NLS models this dispersion measures generation stochasticity given the calibrated parameter vector, which is estimated once per typology and shared by all realizations; it does not include re-calibration uncertainty.

The hourly variance profile reveals a further diagnostic difference. The ratio of synthetic to real hourly variance falls between 0.7 and 1.0 for the AR(1) model across all hours. By contrast, this ratio drops to 0.05–0.25 for the ON/OFF model during the evening peak hours (17:00–21:00), when high-amplitude residential events dominate. The NLS model partially recovers the variance during these hours (ratios of 0.4–0.7) because its optimization explicitly targets hourly variances as part of the 54-statistic objective vector.

The residential load, with its extreme skewness and heavy tails, stresses every model; the commercial load provides a contrasting test case in which the demand distribution is far more regular.

3.2. Commercial Load

Table 5 presents the same comparison for the commercial load.

Table 5. Comparison of distributional statistics: commercial load. All model columns report the median of $K = 50$ Monte Carlo realizations (seeds 42–91).

	Real	**AR(1)**	**Δ%**	**ON/OFF**	**Δ%**	**NLS**	**Δ%**
Mean [kW]	2.46	2.47	+0.2	2.54	+3.3	2.72	+10.6
Std [kW]	2.65	2.64	−0.6	2.58	−2.9	2.62	−1.2
p_{50} [kW]	1.37	1.45	+6.4	1.61	+17.6	1.92	+40.2
p_{90} [kW]	5.00	4.98	−0.4	4.93	−1.3	4.99	−0.0
p_{95} [kW]	6.23	6.24	+0.1	6.14	−1.5	6.31	+1.2
p_{99} [kW]	15.63	15.70	+0.4	15.29	−2.2	15.86	+1.4
RMSE [kW]	–	–	–	0.82	–	1.02	–
KS statistic	–	–	–	0.15	–	0.18	–

Note: absolute error in kW (synthetic median of $K = 50$ realizations minus measured) for the p_{50}–p_{99} rows: AR(1) +0.09, −0.02, +0.01, +0.06; ON/OFF +0.24, −0.07, −0.09, −0.34; NLS +0.55, −0.00, +0.08, +0.22 (kW, in row order). Normalized by the measured standard deviation ($\sigma = 2.65$ kW), the upper-tail errors are AR(1) p_{90} -0.01σ, p_{95} $+0.00\sigma$, p_{99} $+0.02\sigma$; ON/OFF p_{90} -0.02σ, p_{95} -0.03σ, p_{99} -0.13σ; NLS p_{90} -0.00σ, p_{95} $+0.03\sigma$, p_{99} $+0.08\sigma$. Dashes (–) indicate statistics that are not applicable to the corresponding model.

The results for the commercial load contrast sharply with the residential case. The AR(1) model reproduces all statistics with remarkable accuracy. The mean error is only +0.2%, the standard deviation error is −0.6%, and all percentiles fall within 7%. The lower skewness (3.47) and kurtosis (13.84) of the commercial distribution bring it closer to Gaussian, so the AR(1)'s innovation assumption is largely satisfied. Figure 6 shows a representative day of measured and synthetic commercial series.

The ON/OFF model achieves excellent tail reproduction (p_{90} to p_{99} errors below 2.5%) with a standard deviation error of only −2.9%, dramatically better than the −45.6% observed for the residential load. The physical constraints that compress variance in the residential case have minimal impact on the commercial load. The commercial variance arises from the regular day–night cycle rather than from isolated extreme events, and the constraints do not interfere with this mechanism. The KS statistic of 0.15 is substantially better than the residential value of 0.25. This result confirms a tighter overall distributional fit.

The NLS model maintains its advantage in upper-tail accuracy (p_{99}: +1.4%) and achieves a good standard deviation match (−1.2%). However, the NLS model also produces the largest median bias (+40.2%). The RMSE of 1.02 kW is higher than the ON/OFF value of 0.82 kW. This difference reflects the NLS model's prioritization of distributional statistics over point-wise accuracy. Figure 7 compares the measured and synthetic commercial distributions.

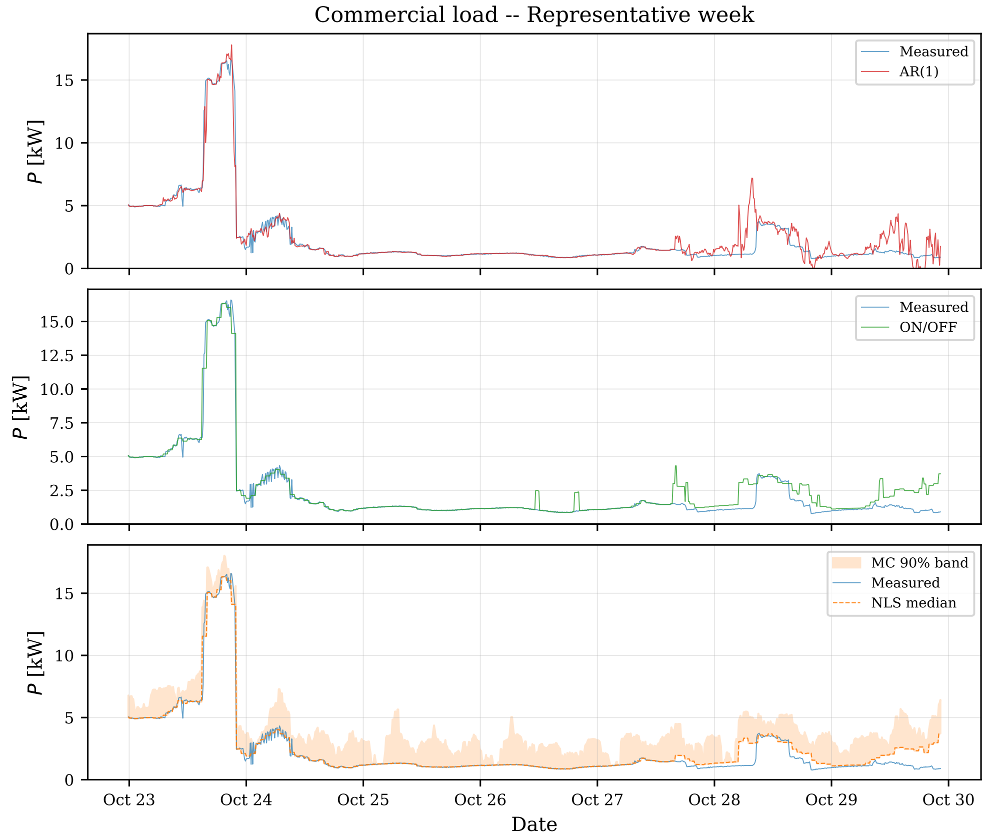

Figure 6. Commercial load: representative time series showing measured P_{total} and synthetic series. The AR(1) model captures the daytime plateau; event-based models reproduce the step transitions at business hours.

The hourly variance ratios for the commercial load range from 0.57 to 1.14 for the NLS model and from 0.70 to 1.50 for the ON/OFF model. These ranges are substantially wider than the 0.05–0.25 ratios observed for the ON/OFF model in the residential case. This confirms that the commercial variance structure is substantially easier to reproduce across all models.

With both typologies characterized individually, the question becomes which structural properties explain the divergent model performance.

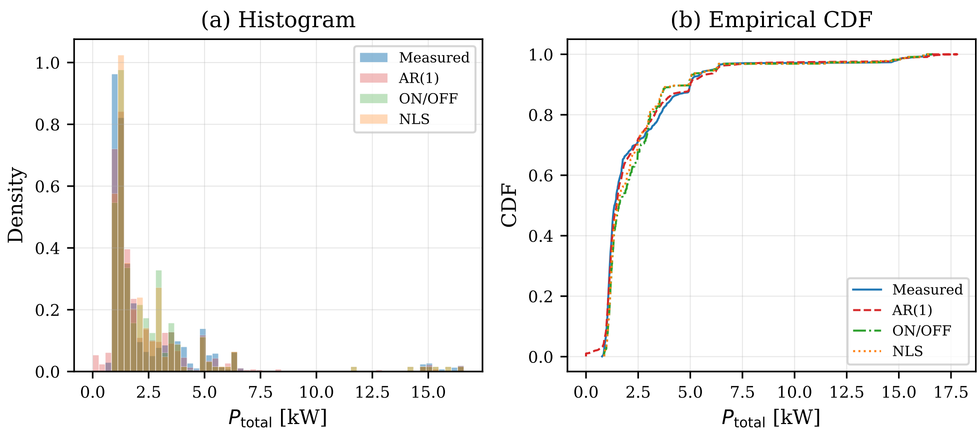

Figure 7. Commercial load: distribution comparison of measured and synthetic profiles. Panel (**a**) shows the probability density histogram and panel (**b**) the empirical cumulative distribution function, in both cases for the measured series and for the AR(1), ON/OFF and NLS models.

3.3. Cross-Typology Comparison

Table 6 summarizes the structural differences between the two load typologies that drive this divergence.

The residential load is characterized by many short, high-amplitude events that produce extreme skewness and very heavy tails. The commercial load exhibits fewer but longer events with a more regular amplitude distribution. As a result, the commercial distribution is more symmetric and lighter-tailed.

Table 6. Structural differences between load typologies.

Property	Residential	Commercial
Skewness	6.56	3.47
Excess kurtosis	61.41	13.84
p_{99}/p_{50}	12.4×	11.4×
Hourly pattern	Double hump	Daytime plateau
Events detected	55	14
Duration median	20 min	50 min
Max event amplitude	5.30 kW	2.28 kW

The temporal signature of these structural differences is captured by the hourly mean profiles of Figure 8. The residential double-hump pattern (morning and evening peaks) and the commercial daytime plateau are both reproduced by the ON/OFF model. However, the variance bars are wider in the residential case due to the sporadic high-power events.

Table 7 consolidates the absolute percentage errors across both typologies. The table highlights the interaction between model choice and load type.

Table 7. Absolute percentage errors [%] by model and typology. Best value per row in **bold.**

	Residential			Commercial		
Statistic	**AR(1)**	**ON/OFF**	**NLS**	**AR(1)**	**ON/OFF**	**NLS**
Mean	12.9	**10.0**	25.8	**0.2**	3.3	10.6
Std	**12.4**	45.6	18.8	**0.6**	2.9	1.2
p_{50}	**14.6**	29.7	37.6	**6.4**	17.6	40.2
p_{90}	28.7	**6.1**	24.5	0.4	1.3	**0.0**
p_{95}	0.8	18.6	**0.3**	**0.1**	1.5	1.2
p_{99}	**5.8**	35.7	13.7	**0.4**	2.2	1.4

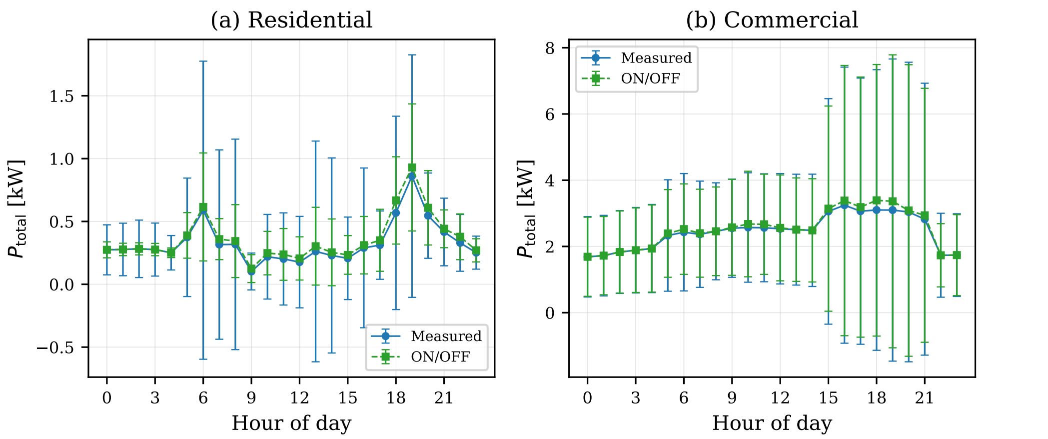

Figure 8. Hourly mean profiles with ±1 standard deviation bands: measured versus ON/OFF for (**a**) residential and (**b**) commercial loads.

Table 7 reveals a clear asymmetry: *every* model performs substantially better on the commercial load than on the residential load across nearly all statistics. The maximum absolute error for the commercial load is 40.2% (NLS, p_{50}), a shared median-bias artifact. The remaining commercial errors are below 18%. For the residential load, by contrast, seven cells exceed 20%. Load typology is therefore a stronger determinant of fit quality than model sophistication.

At the same time, no single model dominates across all statistics. Under the uniform ensemble treatment, best-model status is distributed across all three formulations for the residential load: the ON/OFF captures the mean and the transition zone (p_{90}), the AR(1) the dispersion, the median and p_{99}, and the NLS the p_{95}. On the commercial load the AR(1) is best in five of the six statistics. Two comparisons are decided by margins comparable to the ensemble noise and should be read as ties rather than rankings: residential p_{95} (NLS 0.3% versus AR(1) 0.8%, with per-realization coefficients of variation of 6.1% and 10.0%) and commercial p_{90} (NLS 0.0% versus AR(1) 0.4%). The residential p_{99} comparison is likewise decided by less than eight percentage points between AR(1) and NLS—both underestimate the tail—but the two estimates differ markedly in stability, so the choice there is discussed as a trade-off in Section 4.2 rather than as a ranking.

These percentile-level errors have a direct visual counterpart in the joint hour–power density. Figure 9 contrasts the measured distribution with the NLS-calibrated model for both typologies. The residential NLS reproduces the concentrated low-power band but underestimates the sporadic high-power excursions. By contrast, the commercial NLS closely matches the daytime plateau structure of the measured data.

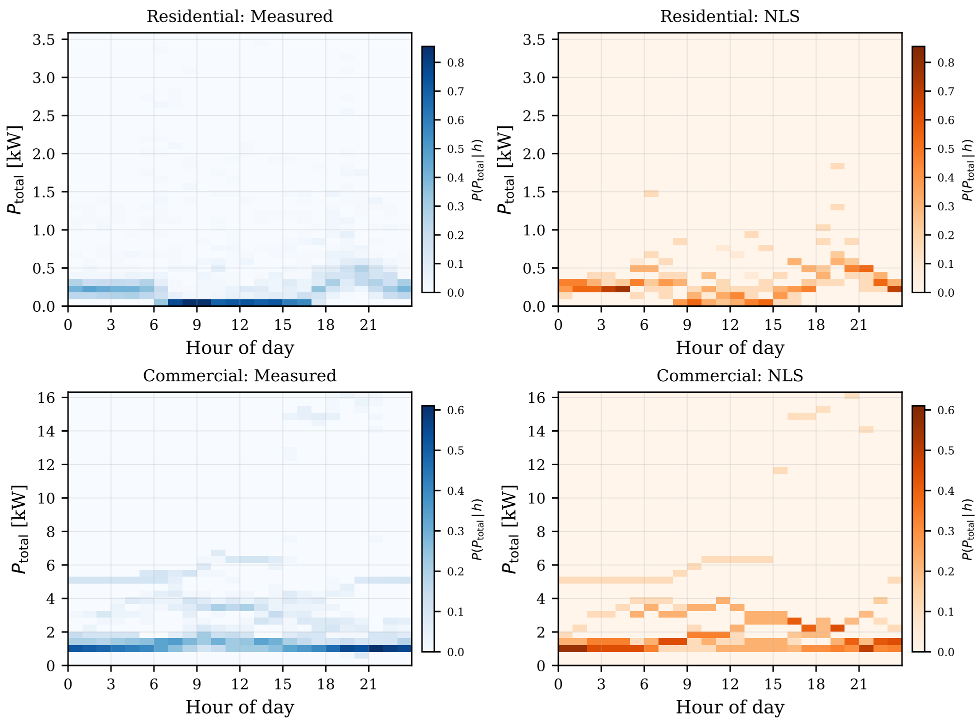

Figure 9. Hour–power density heatmaps: measured (**left**) versus NLS-calibrated (**right**) for residential (**top row**) and commercial (**bottom row**) loads. Each column is normalized so that the hour-conditional distribution $P(P_{\text{total}} \mid h)$ sums to one; cell values represent the probability mass of each power bin given the hour.

3.4. Comparison Against a Random Baseline

A natural question is what a physically motivated generator adds over a naive random draw, given that the marginal distribution of P_{total} can be reproduced almost by construction. To address this directly, the two random reference generators introduced in Section 2

(Baseline A: i.i.d. resampling of the empirical marginal; Baseline B: clipped Gaussian white noise with the measured mean and standard deviation) were run with the same interface, seeds (42–91) and $K = 50$ Monte Carlo realizations as the three models, on both typologies. Because neither baseline is time-synchronized with the measured series, point-wise metrics are uninformative by construction, so the comparison instead targets statistics that are sensitive to temporal structure rather than to the marginal alone.

Both baselines reproduce the marginal distribution essentially exactly: residential p_{99} is matched to within 0.9% (Baseline A) and commercial p_{99} to within 0.4% (Baseline A), comparable to or better than the three calibrated models (Tables 4 and 5). Yet the same baselines inflate the hourly variance ratio up to 12.4× for the residential load and 4.9× for the commercial load, whereas the three calibrated models remain within $[0.04, 1.7]$ across both typologies (Table 8). The effect is equally stark for the ramp distribution: for the commercial load, the measured standard deviation of $\Delta P = P(t) - P(t-1)$ is 0.41 kW; the three models reproduce it within 0.43–0.58 kW, while the two baselines inflate it to 3.2–3.7 kW, roughly 9× the measured value.

The daily-peak result is diagnostic in the opposite direction: because Baseline A resamples the true empirical tail, it is competitive with, or better than, the three calibrated models on this single statistic (Table 8), even though it has no notion of time of day whatsoever. This is precisely the concern that motivates this comparison: a marginal-only statistic does not discriminate between a structured generator and a memoryless random draw. The discrimination instead comes from the temporal-structure metrics above, on which both baselines fail by close to an order of magnitude while the three models stay close to the measured reference. An autocorrelation-based comparison was also computed but is intentionally not reported in Table 8: deseasonalizing a baseline with the real hour-of-day/day-of-week profile—which the baseline does not itself model—reintroduces the profile's own autocorrelation into the residual, so the resulting statistic conflates genuine dynamical memory with the leftover deterministic seasonal signal and is not comparable between baselines and models on the same basis.

Table 8. Structural (temporal) metrics for the three calibrated models and the two random baselines, by typology, against the measured reference. Hourly variance ratio range is the minimum–maximum of the synthetic-to-real variance ratio over the 24 h of day; Ramp σ is the standard deviation of $\Delta P = P(t) - P(t-1)$; Ramp KS and Excursion-duration KS are Kolmogorov–Smirnov statistics against the measured series (excursion durations above the same p_{90} residual threshold used for event detection in Section 2.4); the daily-peak p_{99} error is the difference between the 99th percentile of per-day maxima. All model and baseline values are medians over $K = 50$ Monte Carlo realizations (seeds 42–91). Baseline A: i.i.d. resampling; Baseline B: clipped Gaussian.

Generator	Var. Ratio	Ramp σ [kW]	Ramp KS	Exc. KS	Peak p_{99} Err. [kW]
Residential					
AR(1)	0.43–0.86	0.31	0.12	0.54	−2.65
ON/OFF	0.04–0.61	0.18	0.42	0.26	−5.13
NLS	0.18–1.69	0.51	0.34	0.17	−2.29
Baseline A	0.16–12.38	0.82	0.24	0.34	+0.00
Baseline B	0.14–10.64	0.62	0.29	0.48	−5.61
Real (ref.)	1.00–1.00	0.64	0.00	0.00	0.00
Commercial					
AR(1)	0.86–1.02	0.58	0.13	0.21	+0.57
ON/OFF	0.57–0.98	0.43	0.43	0.38	−0.17
NLS	0.82–1.12	0.53	0.41	0.21	+0.45
Baseline A	0.34–4.94	3.75	0.37	0.25	+0.99
Baseline B	0.23–3.60	3.19	0.41	0.20	−4.47
Real (ref.)	1.00–1.00	0.41	0.00	0.00	0.00

Figure 10 shows the hourly variance ratio and the ramp distribution for the measured reference and the five generators, visually confirming the divergence quantified in Table 8.

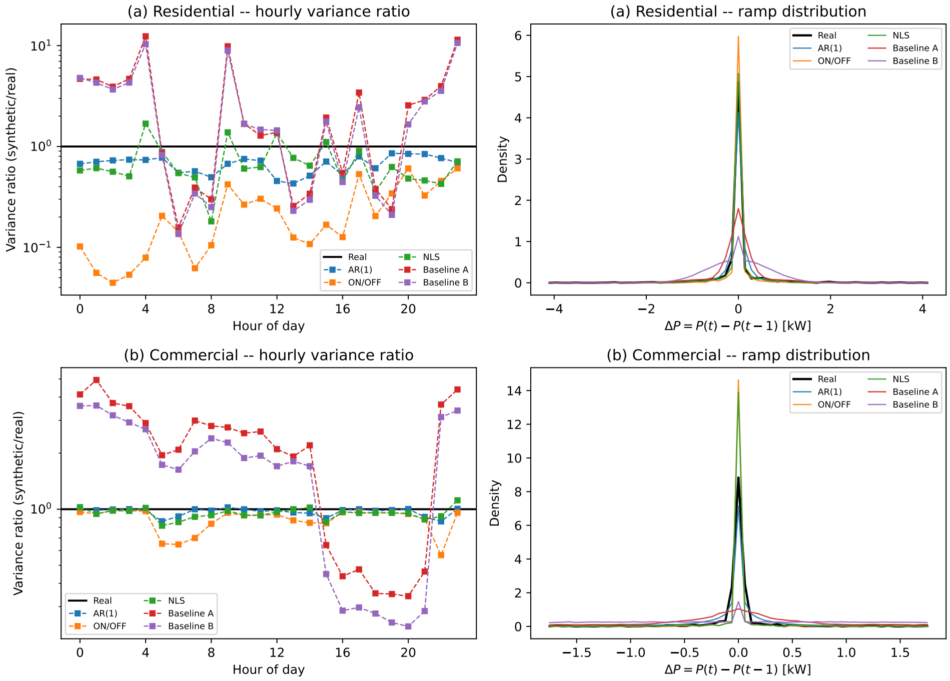

Figure 10. Hourly variance ratio (synthetic-to-real, left column) and ramp distribution $\Delta P = P(t) - P(t-1)$ (right column) for the measured reference and the five generators (AR(1), ON/OFF, NLS, Baseline A, Baseline B): (**a**) residential and (**b**) commercial loads. The labels (**a**,**b**) identify the rows of the figure, so each label appears above both panels of its row: (**a**) the two residential panels and (**b**) the two commercial panels. Within each row, the left panel shows the hourly variance ratio and the right panel the ramp distribution.

4. Discussion

The results above expose a consistent asymmetry: all three models reproduce the commercial load far more accurately than the residential one. Understanding why this asymmetry arises, what trade-offs each model entails, and how these translate into engineering recommendations is the purpose of this section.

4.1. Role of Load Typology

The most salient finding of this study is that the statistical nature of the load, not the complexity of the model, is the primary determinant of synthetic profile quality. The commercial load, composed of many office and laboratory devices operating simultaneously within a fixed institutional schedule, produces a demand signal with skewness 3.47 and excess kurtosis 13.84. These values, while still indicating a non-Gaussian distribution, are an order of magnitude lower than the residential skewness (6.56) and kurtosis (61.41). The aggregation of many independent loads at the transformer level attenuates individual switching transients and produces a distribution that approaches Gaussian. This behavior is a direct consequence of the central limit theorem.

This near-Gaussian character explains why the AR(1) model, whose innovations are exactly Gaussian, reproduces all commercial percentiles within 7%. In the residential case, the same model overestimates p_{90} by 29% because Gaussian tails decay exponentially. The observed residential distribution, by contrast, has a power-law-like heavy tail generated by the sporadic activation of high-wattage appliances. This result has a practical implication. For distribution networks serving predominantly commercial or institutional customers, simple profile-based generators with Gaussian noise are sufficient for most planning studies, and the additional complexity of event-based models is not justified. Conversely, for feeders with a significant residential component, the AR(1) approach will systematically misrepresent the upper tail. This misrepresentation leads to under- or over-sizing of protections depending on which percentile is overestimated or underestimated.

Given that load typology sets the difficulty floor, the relevant question for practitioners is which model offers the best trade-off for a given application.

4.2. Model Trade-Offs

Each of the three models occupies a distinct position in the accuracy–complexity–interpretability space. Table 9 summarizes these positions.

The **AR(1) model** is the fastest to calibrate because no optimization is required and the parameters are computed directly from sample statistics. The AR(1) reproduces the hourly mean and standard deviation by construction. Its limitation is structural. Gaussian innovations cannot generate isolated peaks of 3–8 kW that last only 1–2 samples and are separated by long periods of near-zero residual. Replacing Gaussian innovations with heavy-tailed distributions (e.g., Student-t or gamma) would address this limitation without abandoning the autoregressive framework [18,20]. The ensemble treatment reveals a second property that a single realization hides. Because the hourly marginals are copied from the data, the AR(1) ensemble median is the least biased of the three models on the residential tail (p_{99}: -5.8%); but because a high autocorrelation reduces the effective sample size of one trace, its residential statistics are also the most dispersed (Std CV $= 12.2\%$, p_{99} CV $= 15.3\%$, against 7.7% and 13.3% for ON/OFF and NLS). A near-unbiased tail therefore coexists with the lowest per-realization reliability, which matters whenever a single synthetic trace—rather than an ensemble—is used as the input of a downstream study.

Table 9. Comparative summary of the three stochastic models.

Aspect	AR(1)	ON/OFF	NLS
Parameters	168 (hourly μ, σ) plus ρ	Rates λ_h, amplitude and duration distributions, seven constraints	24 (6 rates + 6 mixing prob. + 6 amplitude + 6 duration)
Calibration cost	Negligible (sample statistics)	Low (empirical detection)	High (K=15 Monte Carlo × 18–25 eval.; ∼10 min)
Interpretability	Low (statistical profile)	High (per-event identity)	Medium (optimized events)
Best statistic (residential)	Std (-12.4%), p_{50} ($+14.6\%$), p_{99} (-5.8%)	Mean ($+10.0\%$), p_{90} ($+6.1\%$)	p_{95} (-0.3%)
Main limitation	Realization dispersion (p_{99} CV: 15.3%)	Variance compression (Std: -45.6%)	Median bias ($+38$–40%)

The **ON/OFF model** addresses the residential tail from a different direction. The ON/OFF model provides physical interpretability because each synthetic event corresponds to a plausible appliance activation with a classified amplitude and lognormal duration. This property makes the ON/OFF model suitable for sensitivity analyses (e.g., evaluating the impact of adding electric vehicle charging as a new event category) and for bottom-up demand-response studies where individual load components must be identifiable. Its central trade-off is the variance penalty. The physical constraints (nocturnal suppression, amplitude caps, minimum duration) that eliminate unrealistic events also remove the extreme events that generate most of the variance in the residential case (standard deviation error: -45.6%). This penalty virtually disappears for the commercial load (-2.9%) because commercial variance arises from the regular day–night cycle rather than from isolated extreme events. This asymmetry implies that physical constraints should be calibrated per typology, not applied generically.

This variance penalty motivates formal optimization. The **NLS model** achieves the best tail accuracy by explicitly including p_{95} and p_{99} in the optimization objective. Its 24-parameter, 59-statistic formulation balances multiple dimensions of distributional fidelity in a way that heuristic tuning cannot match. This accuracy comes at three costs. First, the NLS model requires substantially higher computational expense ($K = 15$ Monte Carlo runs per function evaluation; 18–25 evaluations sufficed here, within a budget of 300). Second, the non-convexity of the objective landscape makes the optimal solution dependent on the initial point. Third, the NLS model exhibits a persistent median bias (+38–40%) shared with the uncalibrated ON/OFF model. This bias indicates that the rectangular-pulse superposition structure inherently inflates the center of the distribution. Future work could explore non-rectangular pulse shapes or amplitude distributions conditioned on the base-load level to mitigate this bias.

Because no single model dominates, the practical question is which model to deploy for which engineering task.

Table 10 quantifies the computational cost of the three models on the same machine (AMD Ryzen 7 4800H, 16 GB RAM, Python 3.11, single-threaded). The AR(1) and ON/OFF models are calibrated in under a second and 20 s respectively, whereas the NLS calibration takes 9.5 min (commercial) and 11.5 min (residential) and accounts for more than 98% of its total cost of 9.7 min and 11.7 min respectively. The optimizer converged by its `xtol` criterion after 18 and 25 function evaluations, well below the 300-evaluation budget; the cost per evaluation is not constant, because parameter vectors far from the optimum generate many more events per simulation than the calibrated vector. Once the parameters are fixed, generating realizations is inexpensive for all three models: the full 50-realization ensemble takes 41 ms (AR(1)), 20 s (ON/OFF, including diagnostics) and 7.7 s (NLS).

Table 10. Computational cost by model (AMD Ryzen 7 4800H, 16 GB RAM, Python 3.11, single-threaded).

Model	Stage	Residential	Commercial
AR(1)	Calibration (lookup table)	0.74 s	0.51 s
	50-realization ensemble	41 ms	30 ms
ON/OFF	Full pipeline (calib. + 50 real.)	20.1 s	17.0 s
NLS	Calibration (18–25 eval.)	∼11.5 min	9.5 min
	50-realization ensemble	∼7.7 s	7.7 s
	Total	**11.7 min**	**9.7 min**

Note: entries in bold report the total calibration-plus-generation time of the NLS model for each typology.

4.3. Practical Recommendations

Table 11 maps each model to the engineering applications for which it is best suited, based on the results of this study.

Two caveats qualify the protection-sizing entry. First, all three models underestimate the residential p_{99} (AR(1): -5.8%; NLS: -13.7%; ON/OFF: -35.7%), so a design margin is warranted whenever the tail of a skewed load is taken from synthetic profiles rather than from measurements. Second, the AR(1) ensemble median is the least biased at p_{99}, but it is also the most dispersed across realizations (CV = 15.3% against 13.3% for the NLS), so the

NLS is preferable when few realizations are propagated downstream, whereas an AR(1) ensemble is competitive when the full ensemble is available.

Table 11. Recommended model by engineering application.

Application	Model	Rationale
Average consumption	AR(1)	Most accurate mean/std
Rapid screening/MC	AR(1)	Fast, no optimization
Protection sizing	NLS	Best p_{95}; tail explicitly calibrated
Reliability/conf. bands	NLS + MC	Embedded Monte Carlo
Physical understanding	ON/OFF	Classifies events
Demand-response	ON/OFF	Identifiable components

The recommendation matrix has direct implications for the design of solar energy systems. For rooftop PV hosting capacity studies, probabilistic assessments rely on Monte Carlo demand and generation scenarios to quantify the risk of voltage, current and transformer constraint violations across realistic feeders [29]; recent frameworks couple source- and load-side uncertainty explicitly, showing that deterministic operating points underestimate the range of limiting network conditions, which frequently involve the coincidence of high PV output with low demand [30]. This shifts the requirement away from upper-tail accuracy: a generator that inflates the low-consumption regime understates net injection and therefore overstates the admissible PV capacity. The median bias of the event-based models acts in that direction (+37.6% residential and +40.2% commercial for the NLS), whereas the AR(1) reproduces the median to within 14.6% and 6.4% respectively. Because the evaluation reported here spans the p_{50}–p_{99} range, it constrains only the upper half of the distribution; quantifying the low-demand regime that governs PV hosting limits would require extending the metric set below the median, which we identify as future work. For BESS sizing in solar microgrids, the combination of the AR(1) model (base profile) with the NLS model (extreme events) enables Monte Carlo construction of net generation-minus-demand scenarios. For microgrid operation studies involving nocturnal disconnection, the ON/OFF model identifies the residential loads that can be targeted for active demand management during low-irradiance hours.

For voltage-dependent studies such as conservation voltage reduction (CVR) or power-flow analysis, the deterministic ZIP model remains the standard format required by network simulators. However, the ZIP model cannot capture temporal variability. The ZIP model produces an essentially constant output when the voltage range is narrow (3.5–4.3% in both measurement sites). Therefore, the ZIP model should be complemented with one of the stochastic generators presented here to produce the demand scenarios over which power flows are executed.

The overarching implication is *typology-specific model selection*. Rather than seeking a single universal generator, planners should segment their loads by typology and apply the simplest model that meets the accuracy requirements of the target application. The recommendation matrix of Table 11 provides the entry point for this segmentation.

These recommendations rest on two specific measurement campaigns; the following limitations bound their generalizability.

4.4. Limitations

Both measurement campaigns cover short periods (9–13 days) within a single season (October–November). These durations are insufficient to capture seasonal effects such as air-conditioning loads in summer or heating in winter. The 10 min aggregation interval attenuates short-duration events (microwave ∼2 min, toaster ∼3 min). This attenuation leads to a potential underestimation of peak amplitudes and event counts. A 1 min resolution would reveal a richer event structure at the cost of 10$\times$ more data. Each typology is represented by a single measurement point. The calibrated parameters therefore reflect a specific household and a specific building rather than a population. Generalization to other residential or commercial sites requires recalibration. A hierarchical model that learns parameter distributions across a panel of sites would be a natural extension. Finally, the narrow voltage range at both sites (3.5–4.3% of nominal) prevents reliable identification of voltage-dependent (ZIP) load components. This limitation could be addressed through controlled voltage variation experiments in the range of 105–130 V.

5. Conclusions

Load typology dominates over model sophistication as the determinant of synthetic profile quality for the two sites examined. Of the 18 model–statistic cells in Table 7, only one exceeds 20% absolute error for the commercial load (the NLS median bias of 40.2%, a structural artifact), whereas seven cells exceed that threshold for the residential load. The p_{99} error of the ON/OFF model alone varies from 2.2% (commercial) to 35.7% (residential), and the NLS model shows the same contrast (1.4% versus 13.7%). This asymmetry indicates that the statistical structure of the load imposes a performance ceiling that no increase in model complexity can overcome without changing the modelling paradigm.

The first-order autoregressive model is sufficient for near-Gaussian commercial loads but becomes unreliable on heavy-tailed residential demand. The commercial distribution (skewness 3.47, kurtosis 13.84) satisfies the Gaussian innovation assumption, and the AR(1) reproduces all percentiles within 7%. The residential distribution (skewness 6.56, kurtosis 61.41) violates this assumption. The AR(1) overestimates p_{90} by 28.7% because Gaussian tails decay exponentially, whereas the observed residential tail follows a power-law-like decay generated by sporadic high-wattage appliance activations. Its ensemble median reproduces p_{99} to within 5.8%, the smallest bias of the three models, but the spread across realizations is also the largest (CV = 15.3%): the AR(1) is nearly unbiased on average yet unreliable in any single trace. Model choice on skewed loads is therefore a bias–stability trade-off rather than a ranking, and it favours the AR(1) only when an ensemble, not a single realization, is propagated downstream.

The physical constraints of the ON/OFF model must be calibrated per typology rather than applied generically. The standard deviation error is -45.6% for the residential load versus only -2.9% for the commercial load. This asymmetry arises because the constraints (nocturnal suppression, amplitude caps, minimum duration) remove the extreme events that generate most of the residential variance, whereas the commercial variance arises from the regular day–night cycle and is not affected by these constraints. The ON/OFF model remains the only formulation that provides per-component interpretability, making it suitable for demand-response and sensitivity analyses.

The NLS calibration achieves the most accurate p_{95} reproduction at the cost of a structural median bias inherent to rectangular-pulse superposition. The p_{95} errors are -0.3% (residential) and $+1.2\%$ (commercial), and the commercial p_{99} error is $+1.4\%$. However, the median bias reaches $+37.6\%$ (residential) and $+40.2\%$ (commercial) because the superposition of rectangular events on the base profile systematically inflates the center of the distribution. Mitigating this bias requires reformulating the pulse geometry or conditioning amplitudes on the base-load level, not additional optimizer iterations.

A typology-specific model selection matrix is more useful for distribution network planning than a single universal generator. The recommendation matrix of Table 11 maps each model to the engineering applications for which it is best suited: the AR(1) for average consumption estimation and rapid screening, the ON/OFF for physical understanding and demand-response studies, and the NLS for protection sizing and reliability analyses. Coupling these stochastic generators with the deterministic ZIP format produces a "stochastic $\rightarrow$ ZIP $\rightarrow$ power flow" design pattern that enables end-to-end probabilistic planning for low-voltage networks.

Several limitations bound the generalizability of these conclusions and define the scope of future work. The models are calibrated on passive loads with a maximum detected amplitude of 5.30 kW, whereas a Level 2 electric vehicle (EV) charger introduces sustained events of 7–11 kW. Extending the ON/OFF framework to active loads requires re-estimating arrival rates and duration distributions from measurement campaigns that include these new components. Open-access tools that jointly generate residential demand, PV output, EV consumption, and vehicle-availability profiles from field datasets [36] provide a concrete template for this extension: the event-based structure of the ON/OFF and NLS models presented here is naturally compatible with such joint generators, since EV charging sessions can be incorporated as an additional event category with its own amplitude and duration distributions. Future work should extend the measurement base to multiple seasons and a panel of households to capture inter-site variability, adopt sub-minute resolution to reveal short-duration events, and explore heavy-tailed innovations (e.g., Student-t or gamma marginals) for the AR(1) framework to improve residential tail reproduction. A natural extension is to couple the synthetic load profiles with stochastic PV generation models and BESS efficiency curves to enable integrated analysis of solar microgrids.

Beyond the comparative results, the dynamic characterization of demand obtained with these three models may constitute a useful input for studies concerned with the integration of photovoltaic generation at the distribution level. The contrasts observed between the residential and commercial typologies in this study, expressed through their distributional and event-level features, could carry different implications for the sizing of PV systems and BESS units, although the present work does not attempt to quantify such implications. These observations are offered as an open avenue for further investigation rather than as findings of the comparative analysis reported above.

Author Contributions: Conceptualization, J.J. and J.R.-G.; methodology, J.J.; software, J.J.; validation, J.J. and R.I.-R.; formal analysis, J.J.; investigation, J.J.; data curation, J.J.; writing—original draft preparation, J.J.; writing—review and editing, R.I.-R. and J.R.-G.; visualization, J.J.; supervision, J.R.-G.; project administration, J.R.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Project Microgrids: energía flexible y eficiente para Cundinamarca (BPIN: 2021000100523), Convenio Especial de Cooperación No. SCTEI-CDCCO-123-2022, signed between the Departamento de Cundinamarca—Secretaría de Ciencia, Tecnología e Innovación, the Universidad Nacional de Colombia, and the Universidad de Cundinamarca. This research was also supported by the Electrical Machines and Drives (EM&D) Research Group from Universidad Nacional de Colombia, through the Energy Solutions Cooperation Network for Communities (code: 59384). This research was also supported by the Red Iberoamericana de Promoción de Comunidades Energéticas Locales para la Transición Justa y Sostenible, RIPCEL code 726RT0203, funded by Programa CYTED.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The measured datasets analysed in this study are not publicly available owing to confidentiality agreements: the residential series originates from a private dwelling and the commercial series from an institutional site governed by the data-use terms of the project under which the campaigns were conducted. The data are available from the corresponding author upon reasonable request, subject to these agreements. The Python source code implementing the three generators is available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the LABE laboratory at Universidad Nacional de Colombia for providing the power-quality measurement infrastructure used in the commercial campaign.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AR(1)	First-order autoregressive process
BESS	Battery energy storage system
CDF	Cumulative distribution function
CVR	Conservation voltage reduction
EV	Electric vehicle
KS	Kolmogorov–Smirnov
MC	Monte Carlo
NLS	Nonlinear least squares
PV	Photovoltaic
RMSE	Root mean square error
ZIP	Constant impedance–current–power load model

References

1. International Energy Agency (IEA). *Accelerating a Clean Energy Transition in Southeast Asia: The Role of Governments*; Technical report; IEA: Paris, France, 2019.
2. Priesmann, J.; Nolting, L.; Praktiknjo, A. Uncertainty Quantification in Load Profiles with Rising EV and PV Adoption. *Appl. Energy* **2022**, *317*, 119148.
3. Capitanescu, F. Risk vs Restriction—An Investigation of Capacity-Limitation Based Congestion Management in Distribution Networks. *Electr. Power Syst. Res.* **2023**, *215*, 108986.
4. Ji, H.; Wang, C.; Li, P.; Zhao, J.; Song, G.; Ding, F.; Wu, J. An Enhanced SOCP-Based Method for Feeder Load Balancing Using the Multi-Terminal Soft Open Point in Active Distribution Networks. *Appl. Energy* **2017**, *208*, 986–995. [CrossRef]
5. Budde, J. Residential Load Profile Tools—A Review and Comparison of Applied Tools. Master's Thesis, Münster University of Applied Sciences, Münster, Germany, 2022. [CrossRef]
6. XM S.A. E.S.P. Informes Mensuales de Análisis del Mercado. Portal Sinergox. 2025. Available online: https://www.xm.com.co/nuestra-empresa/informes/informes-de-la-operacion-y-el-mercado/informes-mensuales-de-analisis-del-mercado (accessed on 30 August 2026).
7. Chen, Y.; Wang, Y.; Kirschen, D.; Zhang, B. Model-Free Renewable Scenario Generation Using Generative Adversarial Networks. *IEEE Trans. Power Syst.* **2018**, *33*, 3265–3275. [CrossRef]
8. Castangia, M.; Giorgi, B.; Quer, S.; Bottaccioli, L.; Patti, E. Residential Load Modeling with Generative Adversarial Networks. *IEEE Trans. Sustain. Comput.* **2025**, *10*, 1216–1226. [CrossRef]
9. Kaneva, T.; Valova, I.; Gabrovska-Evstatieva, K.; Evstatiev, B. A Data-Driven Approach for Generating Synthetic Load Profiles with GANs. *Appl. Sci.* **2025**, *15*, 7835. [CrossRef]
10. Brash, A.; Lu, J.; Stephen, B.; Brown, R.; Atkinson, R.; Michie, C.; MacIntyre, F.; Tachtatzis, C. Coherent Load Profile Synthesis with Conditional Diffusion for LV Distribution Network Scenario Generation. *arXiv* **2025**, arXiv:2510.12832v2.
11. Valova, I.; Gabrovska-Evstatieva, K.G.; Kaneva, T.; Evstatiev, B.I. Generation of Realistic Synthetic Load Profile Based on the Markov Chains Theory: Methodology and Case Studies. *Algorithms* **2025**, *18*, 287. [CrossRef]
12. Arabzadeh, V.; Frank, R. Stochastic Markov-Based Modelling of Residential Lighting Demand in Luxembourg: Integrating Occupant Behavior and Energy Efficiency. *Energies* **2025**, *18*, 5133. [CrossRef]
13. Capasso, A.; Grattieri, W.; Lamedica, R.; Prudenzi, A. A Bottom-Up Approach to Residential Load Modeling. *IEEE Trans. Power Syst.* **1994**, *9*, 957–964. [CrossRef]
14. Osman, M.; Ouf, M.; Athienitis, A. Stochastic Bottom-Up Load Profile Generator for Canadian Households' Electricity Demand. *Build. Environ.* **2023**, *241*, 110490. [CrossRef]
15. Thorve, S.; Baek, Y.Y.; Swarup, S.; Mortveit, H.; Marathe, A.; Vullikanti, A.; Marathe, M. High Resolution Synthetic Residential Energy Use Profiles for the United States. *Sci. Data* **2023**, *10*, 76. [CrossRef]

16. Williams, B.L.M.; Hooper, R.J.; Gnoth, D.; Chase, J.G. Residential Electricity Demand Modelling: Validation of a Behavioural Agent-Based Approach. *Energies* **2025**, *18*, 1314. [CrossRef]
17. Giudice, R.; Amico, S.; Piscitelli, M.S.; Capozzoli, A. Generation of Synthetic Load Profiles for Different Typologies of Residential Users Through Metadata-Driven Generative AI Models. In Proceedings of the 19th IBPSA Conference, Brisbane, Australia, 24–27 August 2025. [CrossRef]
18. Dalla Maria, E.; Secchi, M.; Macii, D. A Flexible Top-Down Data-Driven Stochastic Model for Synthetic Load Profiles Generation. *Energies* **2022**, *15*, 269. [CrossRef]
19. Yılmaz, S.; Eren, T. A Novel Load Profile Generation Method Based on the Estimation of Regional Usage Habit Parameters with Genetic Algorithm. *Electr. Power Syst. Res.* **2023**, *217*, 109165. [CrossRef]
20. Bakouch, H.S.; Mohammadi, M.; Nadarajah, S. A First-Order Autoregressive Process with Size-Biased Lindley Marginals. *Mathematics* **2025**, *13*, 1787. [CrossRef]
21. Sundaram, M.S.B.; Chor, W.T.; Selvaraj, J.; Abu Bakar, A.H.; Tan, C. Categorization of Residential Appliances Using ZIP Load Modeling and Conservation Voltage Reduction Analysis. *Energies* **2025**, *18*, 1999. [CrossRef]
22. Zhu, Y.; Milanović, J.V. Automatic Identification of Power System Load Models Based on Field Measurements. *IEEE Trans. Power Syst.* **2018**, *33*, 3162–3171. [CrossRef]
23. Wang, H.; Chen, Q.; Zhang, L.; Yin, X.; Cui, H.; Zhang, Z.; Wei, H.; Chen, X. A Generalized Load Model Considering the Fault Ride-Through Capability of Distributed PV Generation System. *Energies* **2024**, *17*, 3595. [CrossRef]
24. Li, Z.; Liu, J.; Zhang, H. Uncertainty Quantification and Load Forecasting Based on CPO-VMD. *Energy Rep.* **2024**, *12*, 1050–1063.
25. Muqtadir, A.; Li, B.; Ying, Z.; Songsong, C.; Kazmi, S.N. Nowcasting the Next Hour of Residential Load Using Boosting Ensemble Machines. *Sci. Rep.* **2025**, *15*, 7157. [CrossRef]
26. Misiurek, K.; Olkuski, T.; Zysk, J. Review of Methods and Models for Forecasting Electricity Consumption. *Energies* **2025**, *18*, 4032. [CrossRef]
27. Turowski, M.; Heidrich, B.; Weingärtner, L.; Springer, L.; Phipps, K.; Schäfer, B.; Mikut, R.; Hagenmeyer, V. Generating Synthetic Energy Time Series: A Review. *Renew. Sustain. Energy Rev.* **2024**, *206*, 114842. [CrossRef]
28. Köhler, S.; Rongstock, R.; Hein, M.; Eicker, U. Similarity Measures and Comparison Methods for Residential Electricity Load Profiles. *Energy Build.* **2022**, *271*, 112327. [CrossRef]
29. Qammar, N.; Arshad, A.; Millar, R.J.; Mahmoud, K.; Lehtonen, M. Probabilistic Hosting Capacity Assessment Towards Efficient PV-Rich Low-Voltage Distribution Networks. *Electr. Power Syst. Res.* **2024**, *226*, 109940. [CrossRef]
30. Chen, C.; Peng, W.; Xie, C.; Dong, S.; Hua, Y. Photovoltaic Hosting Capacity Assessment of Distribution Networks Considering Source–Load Uncertainty. *Energies* **2025**, *18*, 2134. [CrossRef]
31. *IEC 61000-4-30:2015*; Electromagnetic Compatibility (EMC)—Part 4-30: Testing and Measurement Techniques—Power Quality Measurement Methods. International Electrotechnical Commission: Geneva, Switzerland, 2015.
32. Coleman, T.F.; Li, Y. An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM J. Optim.* **1996**, *6*, 418–445. [CrossRef]
33. Gouriéroux, C.; Monfort, A.; Renault, E. Indirect Inference. *J. Appl. Econom.* **1993**, *8*, S85–S118. [CrossRef]
34. McFadden, D. A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica* **1989**, *57*, 995–1026. [CrossRef]
35. Pakes, A.; Pollard, D. Simulation and the Asymptotics of Optimization Estimators. *Econometrica* **1989**, *57*, 1027–1057. [CrossRef]
36. Charbonnier, F.; Morstyn, T.; McCulloch, M. Home Electricity Data Generator (HEDGE): An Open-Access Tool for the Generation of Electric Vehicle, Residential Demand, and PV Generation Profiles. *MethodsX* **2024**, *12*, 102618. [CrossRef]