Data-Driven Estimation of Noisy Electrical Demand for Public Power Company in Ecuador

1st Kerly Mishell Vaca Vallejo
Independent
0000-0002-2650-6400

2nd Braulio Paúl Balseca Dahua
ESPOCH
0000-0002-7111-7949

3rd Alexis Bayardo Vaca Barahona
UNICAL
0000-0003-2114-2027

4th Byron Gabriel Vaca Vallejo
ESPOCH
0009-0001-2190-4346

***Abstract*—Real electrical demand data in distribution systems often contain noise, missing samples, and occasional outliers, which can degrade multi-step forecasting when a model is iterated over long horizons. This paper presents a data-driven for demand estimation and forecasting in Guayaquil and Los Ríos. First, the measurements are smoothed using a MAD-based robust filter. Then, the demand is separated into a slow long-term trend and a short-term residual component. The residual dynamics are embedded using Hankel matrices and modeled with a low-rank DMD operator. LMI-based Lyapunov stability certificate for the autonomous residual dynamics. Forecasts are produced by iterating the residual model forward and adding the estimated trend back to recover the demand scale. The approach is evaluated using real demand data from 2016 to 2023 provided by Public Company, and its performance is compared against an ARIMA/SMA baseline.**

***Index Terms*—load forecasting, estimation, Hankel matrix, dynamic mode decomposition (DMD), reduced-order modeling, Lyapunov stability, linear matrix inequalities (LMIs), median absolute deviation (MAD)**

I. Introduction

The increasing variability of electrical demand driven by renewable integration, economic growth, and climate-dependent consumption patterns has made reliable demand estimation and forecasting a central problem for power system planning and operation. In developing regions, such as Ecuador, these challenges are amplified by rapid load growth, heterogeneous consumption behaviors, and imperfect measurement infrastructures [1]–[3]. Accurate forecasting is therefore required not only to minimize prediction error, but also to ensure numerical reliability and robustness when models are used recursively over extended horizons.

Traditionally, load forecasting has been addressed using parametric time-series models and machine learning techniques such as neural networks and neuro-fuzzy systems [2]. Although these approaches often achieve a predictive accuracy, they generally lack explicit mechanisms for managing uncertainty, noise, and guarantees regarding the internal stability of the learned dynamics. Recent probabilistic and forecast-assisted approaches incorporate Bayesian and Gaussian-process-based elements to better represent uncertainty under noisy data [4], [5]. However, these approaches focus primarily on forecast accuracy and uncertainty, giving less attention to the dynamic properties of data-driven models.

In parallel, the control and systems community has developed data-driven and model-free methodologies that avoid explicit system identification and rely directly on measured (input – output) trajectories. Within this paradigm, behavioral and trajectory-based approaches have demonstrated that fundamental system properties can be inferred from data structured in Hankel matrices. Recent approaches have also extended these data-driven paradigms to set-theoretic frameworks for ensuring robust constraint satisfaction [6]–[8]. These ideas have led to data-driven predictive and economic control frameworks, where prediction and optimization are performed directly from historical trajectories without requiring a parametric model description [9]–[11]. Recent work has strengthened this data-driven line by using multimodel linear embeddings to represent nonlinear trajectories [X] and by developing set-theoretic receding horizon methods that enforce stability and constraint satisfaction directly from data [Y] [12]. Hankel-based representations also connect with operator-theoretic predictors such as Koopman approximations and Dynamic Mode Decomposition (DMD). Koopman lifting represents nonlinear dynamics through a linear operator in an augmented space, which supports scalable convex prediction and control [13]. DMD provides a practical way to extract dominant linear dynamics from time-shifted snapshot matrices [14]. These embedding ideas have been combined with set-theoretic predictive control to keep stability and constraints under uncertainty [15].

Despite their advantages, linear operators identified through data-driven approaches may exhibit unstable or ill-conditioned behavior when applied recursively in multi-step prediction tasks, particularly over medium and long horizons, where modeling errors tend to accumulate. To mitigate this issue, data-driven modeling is integrated with convex stability and analyses via Linear Matrix Inequalities (LMIs), enabling rigorous and computationally efficient Lyapunov stability certification $V(x)$ for discrete-time operators [16], [17]. These approaches have been extended to predictive control frameworks, systems with input saturation, and multi agent consensus problems, explicitly exploiting Hankel matrix structures and input–output information [11], [18], [19]. Nevertheless, the application of LMI certified data-driven operators to the estimation and forecasting of electrical demand in real distribution systems remains limited. This work proposes a data-driven framework for the estimation and forecasting of electrical demand in the cities of Guayaquil and Los Ríos, combining Hankel matrix embeddings, reduced order DMD modeling, and stability certification via LMIs.

The proposed approach separates the demand into a long-term trend and a short-term residual. The stability certificate is applied only to the autonomous residual model after detrending and scaling, and it does not imply that the real demand converges, since demand is driven by external factors such as climate, socioeconomic changes, and energy policies [1], [20]. Here, stability is used as a check of the numerical consistency of the learned residual predictor. The method is tested on real demand data from the Ecuadorian distribution system provided by Public Company [1], [21], [22], showing stable multi-step forecasts and an explicit Lyapunov-based stability certificate.

II. Data Preparation and Robust Treatment

This section introduces the data used in the study and the basic preparation done before modeling. The data are prepared to obtain a smoother and more consistent time series. After that, the signal is put in a form that is easier to use in the next identification step.

A. Dataset and temporal granularity

Historical measurements of aggregated active power demand (kW) are considered for two Public Company regional aggregates: Guayaquil and Los Ríos. After quality control and temporal alignment, each region is represented by a uniformly sampled time series

$$y_k \in \mathbb{R}, \qquad k = 1, 2, \dots, N, \tag{1}$$

with fixed sampling interval $\Delta t = 30$ minutes. This resolution preserves intra-day variability while keeping the identification and stability certification steps computationally tractable.

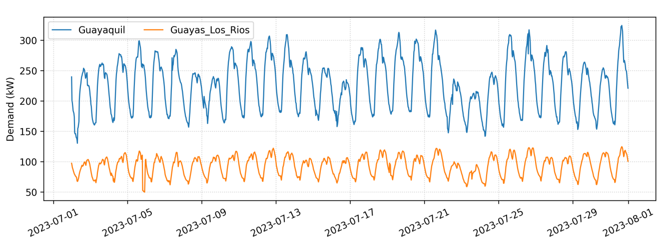


Figure 1: Regional demand patterns.

Figure 1 highlights the demand patterns for each region over a 30-day period. The data shows a daily cycle where consumption rises and falls at different scales. For instance, Guayaquil has the most significant changes between peaks and valleys, while Los Ríos follows a similar pattern but at a lower volume.

B. Noise and outlier handling via MAD

Practical demand measurements include stochastic fluctuations and occasional corrupted samples caused by telemetry dropouts, communication failures, or switching events. The observed series can be described as

$$y_k = \bar{y}_k + \eta_k, \tag{2}$$

where $\bar{y}_k$ represents the underlying demand evolution and η_k collects measurement noise and outliers. To reduce the impact of corrupted samples prior to identification, a anomaly detector based on the median absolute deviation (MAD) is applied:

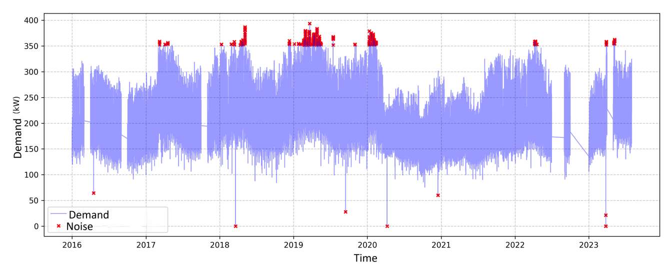
$$\mathrm{MAD} = \mathrm{median}\big(\,|y_k - \tilde{y}|\,\big), \tag{3}$$

where $\tilde{y}$ is the sample median. A modified Z-score is then computed as

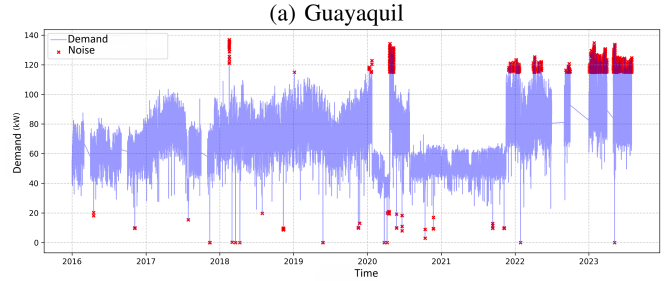
$$M_k = \frac{1.4826\,(y_k - \tilde{y})}{\mathrm{MAD}}, \qquad |M_k| > \tau, \tag{4}$$

is used as a robust anomaly score, where $\tilde{y}$ is the sample median and MAD. The factor 1.4826 makes the MAD comparable to the standard deviation under approximately Gaussian inlier behavior, so that M_k measures the deviation of each sample from the robust center in normalized robust units.

The threshold $\tau = 6$ is adopted to flag only very pronounced and isolated deviations, in order to reduce the risk of removing legitimate demand peaks.



(a) Guayaquil



(b) Los Ríos

Figure 2: Noise/anomaly visualization for the two provinces.

C. Trend–residual decomposition and normalization

After cleaning, the demand series is decomposed into a slow trend component t_k and a residual component

$$r_k = y_k - t_k. \tag{5}$$

The trend is calculated and saved for each region so it can be added back later when converting residual forecasts into actual demand values. Before starting the model identification, the data is standardized. This normalization step improves numerical conditioning by placing the data within a comparable scale, which reduces sensitivity in the SVD-based reduction and avoids poorly scaled variables in the subsequent LMI optimization.

$$\hat{y}_k = \frac{y_k - \mu}{\sigma}, \tag{6}$$

where μ and σ are computed from the corresponding regional series. Standardizing the series this way is helpful because it balances the different demand levels across regions. It also ensures that the math remains reliable and stable when building the Hankel matrix and estimating the final operator.

D. From residual series to Hankel trajectories

To identify the dominant short-term dynamics of the detrended signal, the residual sequence introduced previously is lifted into a trajectory space through a Hankel embedding of depth L. Let r_k, $k = 1, \dots, N$, denote the residual series after preprocessing. The associated Hankel data matrix is constructed as

$$H_r = \begin{bmatrix} r_1 & r_2 & \cdots & r_m \\ r_2 & r_3 & \cdots & r_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ r_L & r_{L+1} & \cdots & r_{L+m-1} \end{bmatrix} \in \mathbb{R}^{L \times m}, \tag{7}$$

where $m = N-L+1$. Each column of H_r represents a length-L delay window of the residual signal, so that the original scalar series is recast as a sequence of overlapping trajectory vectors in an L-dimensional embedding space.

For the subsequent DMD identification, the Hankel matrix is partitioned into two time-shifted snapshot blocks,

$$X_1 = H_r(:, 1 : m-1), \qquad X_2 = H_r(:, 2 : m), \tag{8}$$

so that consecutive columns in X_1 and X_2 represent one-step-shifted delay trajectories. In this form, the identification problem is posed directly in the embedded trajectory space, where the residual evolution is approximated by a linear mapping between successive Hankel snapshots.

E. DMD model

The residual evolution can be approximated by a linear mapping

$$X_2 \approx A X_1, \tag{9}$$

with $A \in \mathbb{R}^{L \times L}$ [23]. Instead of identifying a order operator, by the computing of the reduced-order model through a singular value decomposition (SVD) of X_1,

$$X_1 = U \Sigma V^\top, \qquad U_r \in \mathbb{R}^{L \times r},\ \Sigma_r \in \mathbb{R}^{r \times r},\ V_r \in \mathbb{R}^{(m-1) \times r}, \tag{10}$$

where $r \ll L$ is the retained rank. Using this decomposition, the reduced operator is obtained as

$$A_r = U_r^\top X_2 V_r \Sigma_r^{-1} \in \mathbb{R}^{r \times r}. \tag{11}$$

This predictor reduced coordinates $z_k = U_r^\top x_k$,

$$z_{k+1} = A_r z_k, \qquad \hat{x}_k = U_r z_k. \tag{12}$$

To keep the identification focused on the oscillatory residual component, the trend contribution is removed before forming X_1 and X_2.

F. Problem Formulation

Figure 3 provides a graphical summary of the proposed workflow. The proposed problem consists of: i) residual dynamics identification in reduced coordinates, ii) reconstruction of the original demand signal, and iii) stability certification for iterative prediction.

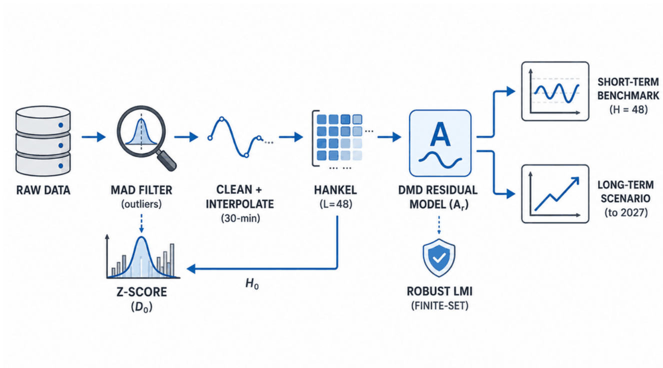


Figure 3: Overview of the work proposal

III. LMI Certification of Residual Dynamics

A recurrent limitation of purely data-driven multi-step forecasting is the accumulation of modeling errors under recursive propagation. Here,the short-horizon predictor is obtained from a reduced residual operator A_r identified via Hankel–DMD, and stability is assessed with a Lyapunov linear matrix inequality (LMI) certificate.

A. Finite-set LMI certification

After identifying the reduced residual operator A_r, stability is certified with a discrete-time Lyapunov LMI under a *finite-set* uncertainty model. The uncertainty set is built by sampling N_{vert} perturbed operators around A_r with a Frobenius-norm bound:

$$A^{(0)} = A_r, \qquad A^{(j)} = A_r + \Delta \frac{E^{(j)}}{\|E^{(j)}\|_F}, \quad j = 1, \dots, N_{\text{vert}}, \tag{13}$$

where each $E^{(j)}$ is a random matrix and Δ is the uncertainty radius. The resulting finite set is denoted by

$$\mathcal{A}_\Delta = \{A^{(j)}\}_{j=0}^{N_{\text{vert}}}. \tag{14}$$

A common Lyapunov matrix $P \succ 0$ is computed such that the inequality holds for every vertex in $\mathcal{A}_\Delta$:

$$(A^{(j)})^\top P A^{(j)} - P + \varepsilon I \preceq 0, \qquad \forall A^{(j)} \in \mathcal{A}_\Delta, \tag{15}$$

together with a strict positivity constraint

$$P - \delta I \succeq 0, \tag{16}$$

where $\varepsilon > 0$ and $\delta > 0$ are small numerical margins.

A conservative worst-case margin is computed as:

$$m_{\text{wc}} = \min_{A^{(j)} \in \mathcal{A}_\Delta} \lambda_{\min}\left(P - (A^{(j)})^\top P A^{(j)}\right). \tag{17}$$

IV. Statistical Forecasting and Uncertainty Analysis

This section uses a classical ARIMA/SMA baseline as a reference for comparison with the proposed Hankel-DMD predictor. The baseline is reported only to contextualize the error levels and to provide an auxiliary residual check.

Figure 4 reports the ACF and PACF of the ARIMA residuals with 95% bounds. Most values stay inside the limits, so there is little autocorrelation left. For this reason, ARIMA is used as the statistical baseline against the Hankel-LMI approach.

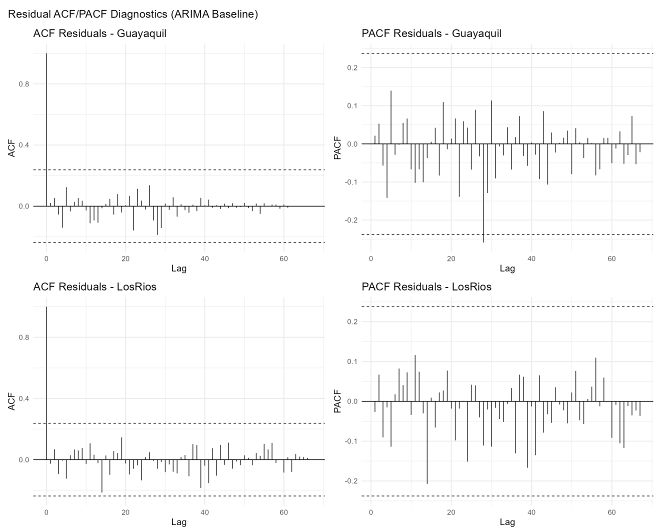


Figure 4: ACF and PACF of the ARIMA model residuals.

The uncerteinty was evaluted with laps of 30 minutes by ARIMA prediction intervals. In Fig. 5, the mean forecast stays relatively stable across regions, but the bands gradually expand as the horizon increases.

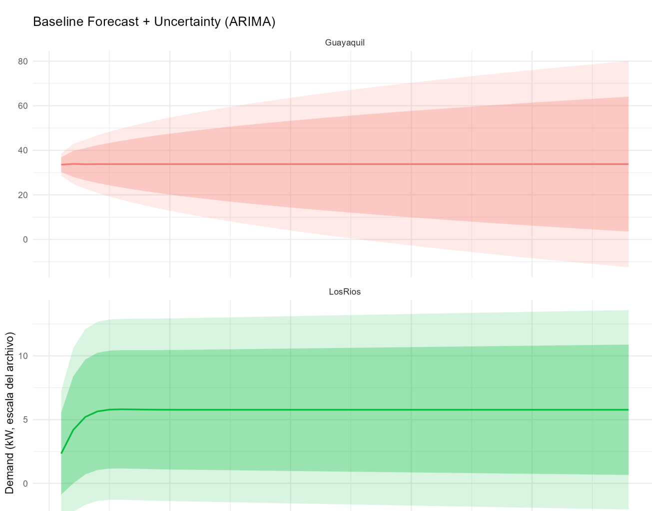


Figure 5: Baseline forecast with uncertainty bands (ARIMA)

A. Benchmark comparison

Figure 6 compares the proposed DMD-based model against the ARIMA/SMA baseline from four complementary viewpoints. In absolute terms, the proposed approach reduces the MAE from 29.49 kW to 6.84 kW and the RMSE from 38.17 kW to 8.46 kW. In normalized terms, these values correspond to only 23.2% and 22.2% of the baseline errors, respectively. This is equivalent to relative reductions of 76.8% in MAE and 77.8% in RMSE, or improvement factors of about 4.3 and 4.5.

V. Experiments

In this part, we evaluate the pipeline using 30-minute demand data from 2016 to 2023, specifically looking at the Guayaquil and Los Ríos regions. The 24-hour-ahead accuracy for late 2023 (with $H = 48$) and scenario projections through 2027, calculated by applying the residual operator step by step.

A. Experimental setup and evaluation metrics

The regional forecasting performance of the proposed model is summarized in Table I, where the errors are reported in the original demand scale (kW). The results show that Guayaquil presents larger MAE and RMSE values than Los Ríos, which is consistent with its higher demand magnitude and stronger temporal variability. In contrast, the lower error values obtained in Los Ríos suggest that the proposed reduced-order predictor can represent its dominant short-term demand fluctuations more accurately. Overall, these results indicate that the identified residual dynamics capture the main repetitive patterns of the demand series and provide a reasonable basis for short-term forecasting in both regions.

Table I: Statistical performance of the proposed model (kW).

Region	MAE (kW)	RMSE (kW)
Guayaquil	25.95	34.46
Los Ríos	8.78	10.88

The residual dynamics can be checked for bounded recursive propagation by Lyapunov-LMI feasibility. In contrast to many purely statistical or black-box predictors, the present framework provides an explicit sufficient condition to verify that the identified residual operator continues to be compatible with bounded multi-step evolution over the uncertainty set used in the certification stage.

B. Stability-certified residual dynamics and scenario projections

Forecasting is next linked to the stability analysis. Figures 7 report residual reconstructions under a feasible Lyapunov certificate. In each region, the one-step residual reconstruction remains consistent over the short horizon, and recursive propagation does not exhibit divergent behavior when the LMI is feasible. This helps because the consider long-horizon projections are produced by repeated application of the identified operator; without a stability check, small modeling errors may accumulate and lead to non-physical growth.

Figure 8 presents two complementary views of the proposed forecasting framework. Panel (A) shows the short-term

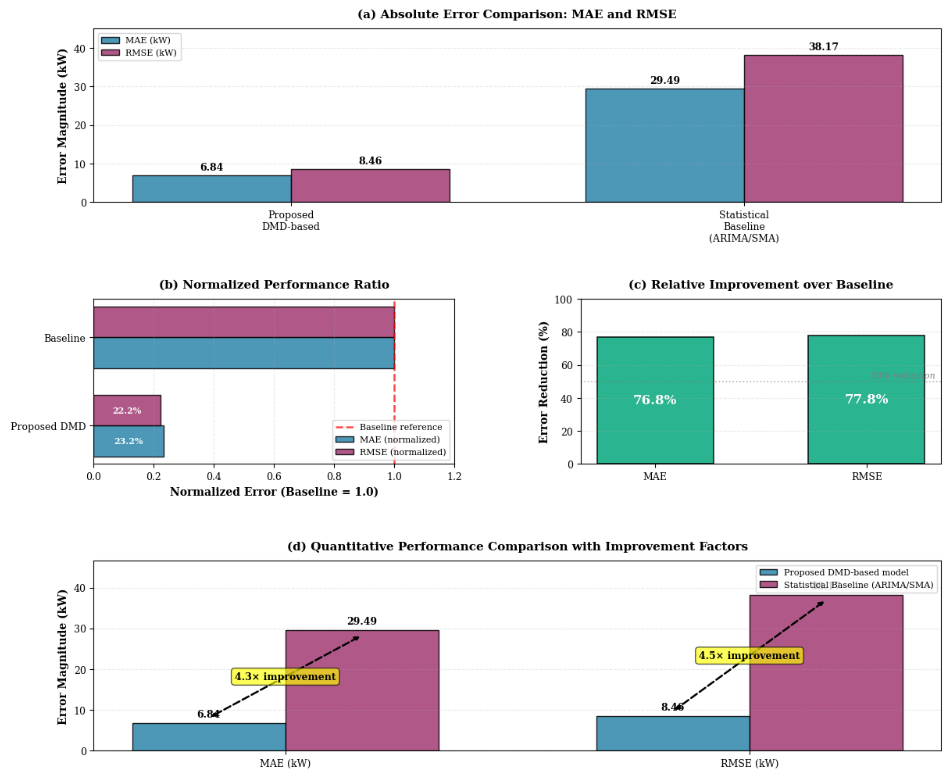


Figure 6: Quantitative comparison between the proposed DMD-based model and the ARIMA/SMA baseline. (a) Absolute MAE and RMSE values. (b) Normalized error ratios with respect to the baseline. (c) Relative error reduction achieved by the proposed approach. (d) Improvement factors in MAE and RMSE over the baseline.

certified projection, where the identified residual dynamics preserve the local oscillatory structure of the load over the prediction window. Panel (B) shows the long-term scenario after re-trending, expressed in terms of daily average demand up to 2027. Since the lower panel represents a trend-level projection rather than a pointwise operational forecast, the short-term oscillations are intentionally not displayed there. The projected trajectory should therefore be interpreted as a smooth long-horizon demand scenario induced by the estimated trend component, not as an exact extrapolation of the daily peaks and valleys.

VI. Conclusions

This work uses 30-minute electricity demand data from 2016 to 2023 for tworegions (Guayaquil and Los Ríos).

The analysis also provides an interpretable view of demand evolution in the two studied regions. The results indicate that the load series can be decomposed into a recurrent short-term component and a slower long-term variation, which supports the use of a residual dynamic model together with a re-trended scenario representation. In this sense, the proposed framework suggests that medium- and long-horizon demand evolution is better interpreted through smooth regional demand scenarios, whereas the intraday oscillatory structure remains a short-term forecasting feature. For the data analyzed here, Guayaquil shows a more variable and demanding profile than Los Ríos, which explains its higher prediction errors.

References

[1] D. Moya, D. Copara, A. Borja, C. Pérez, P. Kaparaju, Ángel Pérez-Navarro, S. Giarola, and A. Hawkes, “Geospatial and temporal estimation of climatic, end-use demands, and socioeconomic drivers of energy consumption in the residential sector in ecuador,” *Energy Conversion and Management*, vol. 261, p. 115629, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0196890422004253

[2] P. Arévalo, A. Cano, and F. Jurado, “Large-scale integration of renewable energies by 2050 through demand prediction with anfis, ecuador case study,” *Energy*, vol. 286, p. 129446, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544223028402

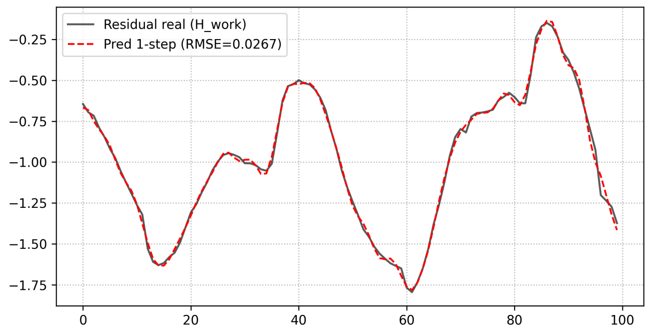
[3] A. Alarcón and J. Quirós-Tortos, “Unlocking the grid: How to ensure reliable and sustainable energy in latin americaand the caribbean,” 2025. [Online]. Available: https://doi.org/10.18235/0013798

[4] X. Ji, Z. Yin, Y. Zhang, M. Wang, X. Zhang, C. Zhang, and D. Wang, “Real-time robust forecasting-aided state estimation of power system based on data-driven models,” *International Journal of Electrical Power & Energy Systems*, vol. 125, p. 106412, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061520302702

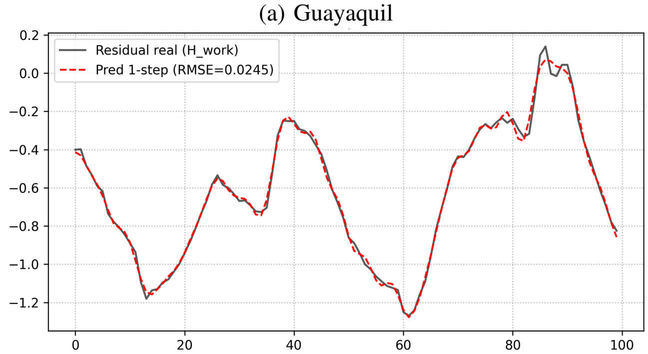
[5] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, “Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 896–913, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207016000133

[6] I. ”Markovsky and P. Rapisarda, “”data-driven simulation and control”,” *”Int. J. Control”*, vol. 81, no. 12, pp. ”1946–1959”, dec 2008.

[7] J. Coulson, J. Lygeros, and F. Dörfler, “Data-enabled predictive control: In the shallows of the deepc,” in *2019 18th European Control Conference (ECC)*, 2019, pp. 307–312.



(a) Guayaquil



(b) Los Ríos

Figure 7: Certified residual reconstruction and short-horizon forecast for the two provinces.

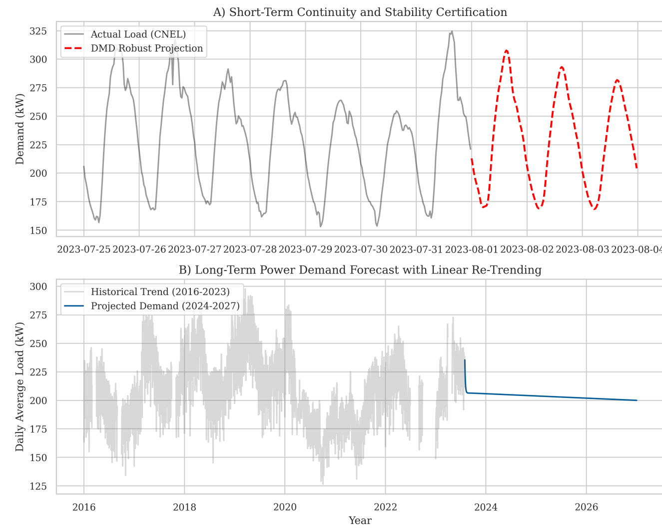


Figure 8: Short-term certified continuity and long-term re-trended demand scenario. Panel (A) shows the short-horizon projection in the original demand scale, while Panel (B) shows the long-horizon projection expressed as daily average demand.

[8] F. Giannini and D. Famularo, "A data-driven approach to set-theoretic model predictive control for nonlinear systems," *Information*, vol. 15, no. 7, p. 369, jun 2024.

[9] F. Ostovar, L. Urbas, and A. A. Safavi, "Data-driven economic mpc with asymptotic stability and strong duality verification using hankel matrix," *Journal of Process Control*, vol. 139, p. 103230, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959152424000702

[10] F. Ostovar, A. A. Safavi, and L. Urbas, "Data-driven distributed empc for economic optimization of interconnected systems: A hankel matrix approach," *ISA Transactions*, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019057826000455

[11] H. H. Nguyen, M. Friedel, and R. Findeisen, "Lmi-based data-driven robust model predictive control," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 4783–4788, 2023, 22nd IFAC World Congress. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896323016476

[12] F. Giannini, G. Franzè, F. Pupo, and G. Fortino, "Set-theoretic receding horizon control for nonlinear systems: a data-driven approach," in *IEEE EUROCON 2023 - 20th International Conference on Smart Technologies*, 2023, pp. 579–584.

[13] M. Korda and I. Mezić, "Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control," *Automatica*, vol. 93, pp. 149–160, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000510981830133X

[14] P. J. "Schmid, ""dynamic mode decomposition of numerical and experimental data"," *"J. Fluid Mech."*, vol. 656, pp. "5–28", aug 2010.

[15] G. Franzè, F. Giannini, V. Puig, and G. Fortino, "Embedding the state trajectories of nonlinear systems via multimodel linear descriptions: A data-driven-based algorithm," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 11, pp. 7143–7155, 2024.

[16] D. Henrion, D. Arzelier, and D. Peaucelle, "Positive polynomial matrices and improved lmi robustness conditions," *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 169–174, 2002, 15th IFAC World Congress. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474667015387838

[17] ——, "Positive polynomial matrices and improved lmi robustness conditions," *Automatica*, vol. 39, no. 8, pp. 1479–1485, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109803001298

[18] Y. Li, Z. Liu, and M. Sader, "Input-output data-driven consensus control of multiagent systems under switching communication network topologies," *Journal of the Franklin Institute*, vol. 363, no. 2, p. 108344, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0016003225008361

[19] F. Porcari, V. Breschi, L. Zaccarian, and S. Formentin, "Data-driven control of input saturated systems: a lmi-based approach," *IFAC-PapersOnLine*, vol. 58, no. 15, pp. 205–210, 2024, 20th IFAC Symposium on System Identification SYSID 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896324013090

[20] J. Torres-Quezada, A. Isalgue-Buxeda, A. Pages-Ramon, J. Lopez-Besora, and A. Delgado-Bohorquez, "Does higher embodied energy lead to lower operational energy? insights from residential buildings in ecuador's costa region," *Journal of Building Engineering*, vol. 113, p. 113870, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352710225021072

[21] S. Pulla-Galindo, D. Borge-Diez, and D. Icaza, "Chapter 4 - energy sector and public lighting," in *Sustainable Energy Planning in Smart Grids*, D. Borge-Diez and E. Rosales-Asensio, Eds. Elsevier, 2024, pp. 53–69. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780443141546000053

[22] S. P. Galindo, D. Borge-Diez, and D. Icaza, "Energy sector in ecuador for public lighting: Current status," *Energy Policy*, vol. 160, p. 112684, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421521005498

[23] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2022.