Working paper · Internal review | v1.0 | 2026-11-20

Forecasting hourly load on the Dominican grid one day ahead

A fixed-model test on all 8,784 hours of 2024

Adrian Dunkley | info@caribbeanaiassociation.com

Abstract

We tested whether a linear regression could forecast hourly active power on the Dominican Republic's interconnected grid more closely than copying an earlier reading. We fitted the model to 26,087 eligible hours before 2024 and scored each method against the same 8,784 hours in 2024. The regression reduced mean absolute error from 123.42 MW for the previous-day benchmark to 105.39 MW, a 14.6% reduction. It underpredicted demand by 30.91 MW on average. The improvement in average error therefore came with a larger directional bias. [1, 2]

Research question

Does a model that combines past demand and the calendar beat yesterday's reading for the same hour when it forecasts load 24 hours ahead?

Data and methods

We used version 5 of Rafael Uceta's SENI dataset and took the target directly from active_power_mw. The file supplies 35,039 hourly observations from 2 January 2021 at 01:00 to 31 December 2024 at 23:00. We checked each timestamp against the month and hour columns, rejected duplicate timestamps and rebuilt all lags by clock time. The first 168 observations lacked a prior week in the supplied target series. [1, 2]

For target hour t, the regression used demand at t−24, t−48 and t−168 hours, an intercept and indicators for hour, weekday and month. At forecast issue time t−24, the latest reading must already be available. We used ordinary least squares on all eligible hours through 2023 and held the coefficients fixed throughout 2024. The two comparison rules copied demand at t−24 or t−168. All three received the same historical observations at each forecast origin.

We calculated mean absolute error (MAE) in megawatts and signed error as forecast minus observation. To estimate uncertainty in the difference between the previous-day rule and the regression, we averaged their paired absolute-error differences by day. We drew 5,000 circular block samples of seven consecutive days until each sample contained 366 days, with seed 20261120. We repeated the calculation with 14-day blocks.

Results

The previous-day rule had a mean absolute error of 123.42 MW. The week-earlier rule had 162.97 MW. The regression had 105.39 MW and improved on the previous-day rule in all 12 calendar months. July had the largest regression error at 130.38 MW. [1]

The mean gain over the previous-day rule was 18.04 MW. The 95% percentile interval was 11.40 to 24.31 MW with seven-day blocks and 12.20 to 24.12 MW with 14-day blocks. The regression's mean signed error was −30.91 MW, compared with −0.39 MW for the previous-day rule. Root mean squared errors were 137.42 and 170.29 MW, respectively. [1]

Average size of the hourly forecast error in 2024

MW



Lower values indicate smaller errors. Each method uses the same 8,784 held-out hourly readings. Source [1], author calculations.

Discussion

A grid planner who cares about typical hourly misses would prefer the regression on this test year. A planner who needs an unbiased estimate would also examine its shortfall. The 30.91 MW average shortfall calls for a separate test of peak-hour errors and of any correction to the forecast. A correction estimated from these same 2024 errors would need fresh observations for evaluation.

Limitations

This retrospective test uses one completed year of supplier-prepared data. It assumes prompt publication of load readings and does not recreate historical release delays or revisions. We excluded recorded weather because weather measured at the target hour would be unavailable at issue time. We did not assess unserved demand, reserve requirements, plant scheduling costs or the performance of the system operator. Block intervals describe variation within the test year conditional on the fitted model.

Reproduce the analysis

Download datos_modelo.csv from dataset version 5 [1]. Interpret dates as day/month/year and check the month and hour columns.

Join the hourly load series to itself at 24, 48 and 168 hours. Fit the specified regression before 2024, using the first category of each calendar variable as the reference.

For 2024, take the absolute forecast error for each hour and average across all hours. To combine monthly figures, weight by the number of hours in each month.

Data table

2024 month	Hours	Previous day MAE (MW)	Previous week MAE (MW)	Regression MAE (MW)
1	744	102.297	125.613	92.504
2	696	126.808	178.018	102.88
3	744	108.949	158.052	95.925
4	720	130.14	207.691	112.416
5	744	132.782	230.034	116.888
6	720	120.838	172.054	98.426
7	744	163.895	180.52	130.382
8	744	126.179	117.589	104.585
9	720	112.595	164.216	103.157
10	744	117.224	122.403	107.985
11	720	120.623	163.38	105.494
12	744	118.673	138.836	93.773

References

[1] Uceta, R. (2025). Short-Term Forecasting Dataset SENI Dominican Rep, version 5. Kaggle. File: datos_modelo.csv.

[2] Uceta-Acosta et al. (2025). A dataset of exogenous variables and historical electricity demand for short-term load forecasting of the national interconnected electric system (SENI) in the Dominican Republic from 2021 to 2024. Data in Brief 63, 112057.

Data credits

Uceta, SENI v5. Original hourly data: CC BY-NC-SA 4.0. CAIA evaluation results.