Vol.4, No.9, Page 2428- 2440, 2026 https://doi.org/10.59628/jast.v4i9.3303 ISSN 2958-9568 JAST

مجلة جامعة صنعاء للعلوم التطبيقية والتكنولوجيا
Sana'a University Journal of Applied Sciences and Technology
https://journals.su.edu.ye/index.php/jast/

NFGA-LINEAR: An Explainable Neuro-Fuzzy Genetic Framework for Forecasting Anomaly-Affected Data-Scarce Time Series

Muneef Abdulkareem Farea Ahmed * and Ahmed Al-Shalabi

Department of Computer Science, Faculty of Computer and Information Technology, Sana'a University, Sana'a, Yemen

*Corresponding author: muneef.ahmed@su.edu.ye

Abstract

Reliable forecasting becomes particularly challenging when a time series is short, nonlinear, and contaminated by occasional, abnormal observations. We develop NFGA-LINEAR, an interpretable Takagi–Sugeno–Kang neuro-fuzzy framework in which a genetic algorithm jointly determines the number of rules and optimizes Gaussian antecedents and rule-specific linear consequents. A robust residual-based procedure was incorporated for anomaly detection. The framework was assessed using weekly data from the Cholera, ILINet, and household electricity series using a leakage-controlled, one-step-ahead walk-forward design. XGBoost, LSTM, NFGA-Core, and NFGA-LINEAR were evaluated using ten matched random seeds, whereas deterministic methods were summarized through single-run descriptive comparisons. Performance was dataset-dependent: Persistence produced the lowest RMSE for Cholera, ARIMA ranked first for ILINet, and Prophet and XGBoost differed by only approximately 1.1×10^{-6} RMSE on Electricity. In the prespecified matched ablation, replacing the constant consequents of NFGA-Core with local linear consequents lowered the mean RMSE by 69.0%, 40.2%, and 4.5% for Cholera, ILINet, and Electricity, respectively. Exact Wilcoxon signed-rank tests, followed by Holm correction across the three primary comparisons, gave adjusted p-values of 0.0059, 0.0059, and 0.0371. Under synthetic anomaly-injection rates of 5% and 10%, NFGA-LINEAR achieved the highest macro-F_1 for Cholera and ILINet, whereas Prophet performed best on Electricity. Fixed-model NFGA-LINEAR inference required approximately 0.081–0.110 ms per observation, with an approximate stored parameter representation of 1.9–2.8 KB. These findings position NFGA-LINEAR as a compact and transparent accuracy–complexity compromise while confirming that neither its forecasting nor detection performance is uniformly superior across datasets.

Article Info

Keywords:
Neuro-Fuzzy Systems, Genetic Algorithm, Time Series Forecasting, Data Scarcity, Anomaly Detection.

Article History:
Received: 09-June-2026,
Revised: 20-June-2026,
Accepted: 05-August-2026,
Published: 28 September 2026.

1. INTRODUCTION

Reliable predictions from temporally ordered observations are important in epidemiological surveillance, electricity management, environmental assessment, and many other operational domains. The task becomes substantially more difficult when historical records are limited. Under such conditions, highly parameterized models may exhibit unstable estimates or require more observations than are available, whereas low-complexity alternatives may be unable to capture nonlinear and locally varying relationships. Occasional anomalous observations further complicate the problem because outbreaks, reporting inconsistencies, sensor malfunctions, and rare operational events can distort the model fitting and alter the distribution of forecast residuals. Several forecasting families address different aspects of this problem. Classical autoregressive methods remain widely used because of their compact structure and direct statistical interpretation [1]. Prophet represents a time series through an additive trend and seasonal components and was designed for scalable forecasting applications [2]. Nonlinear learners, such as XGBoost [3] and Long Short-Term Memory networks [4] can model complex temporal patterns, although their learned representations are generally less transparent and may be difficult to estimate reliably from short records. Neuro-fuzzy systems provide a different compromise by expressing nonlinear mappings through explicit fuzzy rules with inspectable antecedent and consequent components [5, 6]. In this study, we developed and reevaluated a genetic neuro-fuzzy framework for one-step-ahead forecasting and residual-based anomaly detection. Its proposed variant, NFGA-LINEAR, extends NFGA-Core by replacing the zero-order rule consequents with local linear Takagi–Sugeno–Kang (TSK) functions. The genetic search jointly optimizes the Gaussian antecedent parameters, consequent coefficients, and bounded rule count. Complexity penalties are included to discourage unnecessarily large rule bases and excessive consequent coefficients from being generated. The central question is not whether a neuro-fuzzy architecture can outperform every competing method on every dataset. Rather, the study examines whether local linear consequents yield a reproducible improvement over an otherwise matched zero-order architecture, whether that improvement remains supported after correction for multiple statistical comparisons, and whether the resulting model provides a defensible compromise among forecast accuracy, structural interpretability, anomaly detector behavior, and computational cost. The main contributions of this study are as follows:

- a variable-rule TSK neuro-fuzzy forecaster whose Gaussian antecedents and local linear consequents are jointly optimized by a genetic algorithm;
- a corrected full-horizon walk-forward protocol using training-only preprocessing, ten matched random seeds for stochastic models, and separate descriptive reporting for deterministic baselines;
- an anomaly-detection benchmark that distinguishes forecasting models from independent detectors and compares Robust-Z, Isolation Forest, Local Outlier Factor, One-Class SVM, and forecast-residual methods;
- a prespecified matched ablation supported by exact Wilcoxon and Friedman analyses, together with one-factor-at-a-time sensitivity analysis, residual-threshold sensitivity, and empirical measurements of training time, inference latency, and approximate model-storage cost.

2. RELATED WORK AND POSITIONING

Neuro-fuzzy forecasting represents temporal relationships through fuzzy rules while estimating model parameters from observed data. Existing approaches include conventional neuro-fuzzy predictors, hybrid learning schemes, and distributed rule-based architectures for time-series prediction [5–8]. Subsequent research has combined ANFIS-type models with metaheuristic optimization in applications such as wind-power prediction, carbon-price forecasting, software-reliability estimation, and financial time-series analysis [9–14]. Attention mechanisms and decomposition-based representations have also been integrated with neuro-fuzzy inference to improve nonlinear feature modeling and temporal prediction [15, 16]. The presence of fuzzy rules does not ensure that a model remains practically interpretable. Transparency may deteriorate when the rule base becomes large, the antecedents span many dimensions, or the optimization procedure produces structures that are difficult to inspect. Previous studies have therefore examined the balance between predictive accuracy and explainability in data-driven forecasting systems [17, 18]. NFGA-LINEAR addresses this issue through an explicit rule-based representation in which every rule contains a Gaussian antecedent and local affine consequent. The selected rule count and associated antecedent and consequent parameters remain available for direct inspection after training. Anomalies in time-series data can be identified from individual observations, multivariate or lagged windows, or deviations between forecasts and realized values. Robust median-based statistics reduce the influence of extreme observations during threshold estimation [19], whereas Isolation Forest identifies unusual samples through recursive random partitioning without requiring a specified parametric distribution [20]. Local Outlier Factor identifies observations whose local density is substantially lower than that of their surrounding neighbors [21], while One-Class SVM estimates the support of the normal-data distribution and flags observations that fall outside the learned boundary [22]. In the present experimental design, the Isolation Forest, Local Outlier Factor, and One-Class SVM were evaluated exclusively as independent anomaly detectors. Therefore, they are excluded from the forecasting-performance ranking. Forecasting accuracy is assessed only for methods that generate one-step-ahead numerical predictions, whereas detector performance is reported separately using classification-oriented measures. Table 1 presents the positioning of the evaluated framework relative to the model families included in this study. NFGA-LINEAR is distinguished by the combination of a genetically optimized rule structure, local linear consequents, complete walk-forward evaluation, residual and independent detector comparisons, matched ablation, sensitivity analysis, and measured computational cost.

3. METHODOLOGY

3.1. Forecasting formulation

Let $Y = \{y_1, y_2, \ldots, y_T\}$ be a univariate time series. For a window length w, the input at time t is

$$\mathbf{x}_t = [y_{t-w+1}, y_{t-w+2}, \ldots, y_t]^\top, \tag{1}$$

Table 1. Positioning of the model families as evaluated in the present study.

Approach	Explicit rules	Local linear consequents	Structural optimization	Full walk-forward	Detector evaluation	Cost/sensitivity audit
ARIMA/Prophet	No	No	No	Yes	Residual-based	Partial
XGBoost/LSTM	No	No	Hyperparameter tuning	Yes	Residual-based	Measured here
Optimized ANFIS literature	Yes	Often	Study-dependent	Study-dependent	Usually separate	Study-dependent
NFGA-Core	Yes	No	GA with variable K	Yes	Residual-based	Yes
NFGA-LINEAR	Yes	Yes	GA with variable K	Yes	Residual and independent detectors	Yes

The target is y_{t+1}. Standardization uses only the currently available training history.

$$z_t = \frac{y_t - \mu_{\text{train}}}{\sigma_{\text{train}}}. \tag{2}$$

The fitted transformation was applied to later observations without using future test information.

3.2. Gaussian antecedents and normalized firing

For rule $i \in \{1, \ldots, K\}$, let $\mathbf{c}_i \in \mathbb{R}^w$ denote its center and $\sigma_i > 0$ denote its spread. The Gaussian firing strength is

$$\mu_i(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_i\|_2^2}{2\sigma_i^2}\right), \tag{3}$$

and the normalized strength is

$$\bar{\mu}_i(\mathbf{x}) = \frac{\mu_i(\mathbf{x})}{\sum_{j=1}^{K} \mu_j(\mathbf{x}) + \varepsilon}, \tag{4}$$

where a small ε prevents the numerical division by zero.

3.3. NFGA-Core and NFGA-LINEAR

NFGA-Core uses a zero-order consequent p_i for each rule as follows:

$$\widehat{y}_{\text{Core}}(\mathbf{x}) = \sum_{i=1}^{K} \bar{\mu}_i(\mathbf{x}) p_i. \tag{5}$$

NFGA-LINEAR replaces p_i with a local affine function,

$$f_i(\mathbf{x}) = a_{i0} + \sum_{m=1}^{w} a_{im} x_m, \tag{6}$$

so that

$$\widehat{y}_{\text{Linear}}(\mathbf{x}) = \sum_{i=1}^{K} \bar{\mu}_i(\mathbf{x}) f_i(\mathbf{x}). \tag{7}$$

Therefore, each prediction can be decomposed into normalized rule activations and rule-local linear contributions.

3.4. Genetic representation and optimization

A chromosome contains the rule count and all active rule parameters.

$$\mathcal{C} = \left[K, \{\mathbf{c}_i, \sigma_i, a_{i0}, a_{i1}, \ldots, a_{iw}\}_{i=1}^{K}\right]. \tag{8}$$

The minimized objective is

$$\mathcal{J} = \text{RMSE}_{\text{train}} + \lambda_K K + \lambda_a \sum_{i=1}^{K} \|\mathbf{a}_i\|_2^2, \tag{9}$$

where λ_K penalizes rule count and λ_a weakly regularizes the local linear coefficients. Tournament selection, one-point crossover, Gaussian parameter mutation, and rule insertion/deletion were used. Table 2 reports the prespecified configuration used in the main experiment.

Table 2. Main NFGA configuration.

Parameter	Value
Rule-count range $[K_{\min}, K_{\max}]$	$[2, 8]$
Population size	25
Maximum generations	40
Crossover probability	0.90
Center mutation probability	0.25
Spread mutation probability	0.25
Consequent mutation probability	0.30
Rule-count penalty λ_K	0.005
Coefficient penalty λ_a	10^{-6}
Random seeds	11, 22, ..., 110

3.5. Residual anomaly detector

Absolute forecast residuals are defined by

$$e_t = |y_t - \widehat{y}_t|. \tag{10}$$

Using calibration residuals only, the median absolute deviation and its normal-consistency scaling are

$$\text{MAD} = \text{median}_j \left|e_j - \text{median}(e)\right|, \tag{11}$$

$$s_{\text{MAD}} = 1.4826\,\text{MAD}. \tag{12}$$

The threshold is

$$\tau = \text{median}(e) + k s_{\text{MAD}}, \qquad k = 3.5, \tag{13}$$

and a test point is flagged when $e_t > \tau$. The value $k = 3.5$ was fixed before the detector comparison; the sensitivity analysis did not post-hoc replace it.

3.6. Complexity

For K rules and window dimension w, one fixed-model NFGA-LINEAR prediction evaluates K distances and K affine consequents, giving $O(Kw)$ time. The active centers and linear coefficients also require $O(Kw)$ of storage. If a population of size P is evaluated for G generations over N training windows, the dominant genetic search cost is approximately $O(GPNKw)$. The expression describes the implementation-level scaling and does not imply equal constants across the model families.

4. EXPERIMENTAL DESIGN

4.1. Datasets and chronological splits

Three weekly univariate datasets were utilized. Cholera contained 417 observations, ILINet contained 260 observations, and electricity contained 207 observations. Chronological 80%/20% splits produced complete test horizons of 84, 52, and 42 one-step-ahead forecasts. The window lengths were fixed at 52, 24, and 16. Table 3 summarizes the study design.

Table 3. Dataset and walk-forward configuration.

Dataset	Total	Train	Test	Window
Cholera	417	333	84	52
ILINet	260	208	52	24
Electricity	207	165	42	16

4.2. Forecasting baselines

The forecast comparison included Persistence, Seasonal Naive, ARIMA, Prophet, XGBoost, LSTM, NFGA-Core, and NFGA-LINEAR. Prophet was re-implemented as a genuine forecasting baseline rather than an anomaly model. Five candidate configurations were selected based on the lowest RMSE in a chronological validation segment that was contained within the training partition. The selected additive configurations were then refitted at each test step in a walk-forward evaluation. XGBoost, LSTM, NFGA-Core, and NFGA-LINEAR were repeated with ten matched seeds, and the deterministic methods were run once and interpreted in a descriptive manner.

4.3. Anomaly scenarios and detector baselines

Synthetic anomalies were injected at prespecified 5% and 10% rates in the test horizon, producing few positives by design: 4/8 for cholera, 3/5 for ILINet, and 2/4 for electricity. The same scenario labels were used for each detector. The comparison includes the NFGA-LINEAR and Prophet residual detectors, Robust-Z, Isolation Forest, Local Outlier Factor, and One-Class SVM. Independent window detectors were trained without access to the test labels. Robust-Z labels computed from the original series were retained only as an exploratory proxy sanity check and were excluded from the main detector ranking because they were not verified ground truths; evaluating Robust-Z against labels generated by the same rule would be circular. Precision, recall, F1, specificity, false-positive rate (FPR), and balanced accuracy were calculated when a scenario contained positive reference points. Precision, recall, and F1 are reported as N/A when the reference set contains no positives, rather than being assigned an artificial zero.

4.4. Evaluation and statistical analysis

Forecasting uses a strict one-step-ahead walk-forward evaluation. At each step, the model forecasts the next observation using only available history. The observation was then revealed and appended. The RMSE is the prespecified primary forecast endpoint, with MAE, sMAPE, and R^2 as secondary summaries. For the four stochastic models, a Friedman test was used to assess the overall model effect separately for each dataset and metric. Exact two-sided Wilcoxon signed-rank tests were used to compare NFGA-LINEAR with NFGA-Core, XGBoost, and LSTM using matched seeds. Pairwise forecast comparisons used the Holm correction within each dataset and metric. The primary ablation family is defined separately as the three NFGA-LINEAR versus NFGA-Core RMSE tests, one per dataset, and receives a Holm correction across the three tests. Paired benefits were accompanied by 95% bootstrap confidence intervals based on 20,000 resamples and rank-biserial effects. Deterministic single-run baselines were not included in the seed-level tests. The earlier five-seed analysis had a minimum attainable two-sided exact Wilcoxon value of 0.0625. The corrected ten-seed design removes this discrete limitation; statistical conclusions are based on the corrected tests rather than the original five-seed values.

4.5. Sensitivity and resource measurement

One-factor-at-a-time (OFAT) sensitivity varied $K_{\max}$, window length, rule penalty, mutation rate, crossover proba-

Table 4. Corrected forecasting performance under the full test walk-forward protocol.

Dataset	Model	RMSE	MAE	sMAPE (%)	R^2
Cholera	Persistence	96.034	55.167	18.706	0.9522
Cholera	ARIMA	103.009	60.959	21.458	0.9450
Cholera	NFGA-LINEAR	136.072 ± 41.789	110.960 ± 48.963	59.170 ± 29.725	0.8960 ± 0.0690
Cholera	LSTM	193.804 ± 31.841	140.938 ± 31.611	81.101 ± 30.173	0.8008 ± 0.0649
Cholera	XGBoost	391.631 ± 24.707	295.072 ± 17.132	68.509 ± 3.787	0.2029 ± 0.1025
Cholera	NFGA-Core	459.729 ± 118.382	376.704 ± 80.946	98.554 ± 34.808	-0.1598 ± 0.5985
Cholera	Seasonal Naive	618.738	488.988	126.525	-0.9826
Cholera	Prophet	1158.884	860.190	163.163	-5.9550
ILINet	ARIMA	0.3951	0.2236	6.125	0.9578
ILINet	NFGA-LINEAR	0.4422 ± 0.0143	0.2615 ± 0.0099	7.899 ± 0.976	0.9471 ± 0.0034
ILINet	Persistence	0.5017	0.3203	8.935	0.9319
ILINet	XGBoost	0.7009 ± 0.0169	0.3750 ± 0.0087	10.120 ± 0.239	0.8671 ± 0.0064
ILINet	NFGA-Core	0.7414 ± 0.0449	0.4943 ± 0.0617	16.852 ± 3.241	0.8508 ± 0.0183
ILINet	LSTM	0.7539 ± 0.0701	0.4545 ± 0.0558	14.097 ± 1.711	0.8451 ± 0.0299
ILINet	Seasonal Naive	1.0248	0.6474	19.211	0.7159
ILINet	Prophet	1.5270	1.3393	45.409	0.3693
Electricity	Prophet	0.130018	0.100413	10.601	0.6872
Electricity	XGBoost	0.130019 ± 0.001560	0.105078 ± 0.001966	11.802 ± 0.249	0.6872 ± 0.0075
Electricity	LSTM	0.1384 ± 0.0046	0.1061 ± 0.0019	11.442 ± 0.216	0.6452 ± 0.0237
Electricity	NFGA-LINEAR	0.1479 ± 0.0031	0.1134 ± 0.0036	12.079 ± 0.301	0.5953 ± 0.0171
Electricity	ARIMA	0.1487	0.1183	12.527	0.5910
Electricity	Seasonal Naive	0.1526	0.1200	13.170	0.5690
Electricity	NFGA-Core	0.1553 ± 0.0106	0.1138 ± 0.0075	12.387 ± 0.848	0.5518 ± 0.0609
Electricity	Persistence	0.1606	0.1245	13.190	0.5226

Deterministic methods are single runs. XGBoost, LSTM, NFGA-Core, and NFGA-LINEAR are mean $\pm$ standard deviation over ten matched seeds. Prophet and XGBoost differ by approximately 1.1×10^{-6} RMSE on Electricity; no significant claim is made from that descriptive difference.

bility, and genetic budget while holding the other factors at the baseline. This design measures marginal sensitivity but not the interactions. In a separate detector analysis, k was varied from 2.0 to 5.0 while retaining $k = 3.5$ for the main comparison. The training time, fixed-model inference latency, per-step refit time (where applicable), and approximate stored representation size were measured in the same experimental environment. Size definitions differ by model and are interpreted as approximate footprints rather than directly equivalent memory measurements.

5. RESULTS

5.1. Forecasting performance

Table 4 and Fig. Figure 1 shows the corrected full horizon results. No model was the best for all the datasets. Persistence gives the lowest Cholera RMSE (96.034), ARIMA leads ILINet (0.3951), and Prophet and XGBoost are numerically tied on electricity at four decimal places (0.1300). NFGA-LINEAR ranked third, second, and fourth for Cholera, ILINet, and Electricity, respectively. It is consistently better than NFGA-Core and better than the evaluated stochastic black-box baselines on Cholera and ILINet.

Table 5. Matched ten-seed primary RMSE ablation of local linear consequents.

Dataset	Core	Linear	Gain	Holm p
Cholera	459.729	136.072	69.0%	0.0059
ILINet	0.7414	0.4422	40.2%	0.0059
Electricity	0.1553	0.1479	4.5%	0.0371

Holm correction is across the three prespecified Core-versus-LINEAR RMSE comparisons.

5.2. Matched ablation and statistical tests

Primary ablation isolates the change from zero-order to local linear consequences under matched seeds and otherwise aligned settings. Table 5 and Fig. 2 show reductions of 69.0%, 40.2%, and 4.5% in the mean RMSE. All three primary Holm-adjusted values were below 0.05. The effect is large and consistent on Cholera and ILINet; on Electricity, NFGA-LINEAR is better in eight of ten seeds, and the effect is smaller.

The Friedman test was significant for every dataset and metric among the four stochastic models (RMSE p values: 1.4×10^{-5} for Cholera, 8.9×10^{-5} for ILINet, and 6.0×10^{-6} for Electricity). Table 6 presents the pairwise RMSE results. NFGA-LINEAR was significantly better than NFGA-Core, XGBoost, and LSTM for Cholera and ILINet. For Electricity, it remained significantly better than NFGA-Core but significantly worse than XGBoost and LSTM. This dataset-dependent reversal prevents a universal superiority from being claimed.

5.3. Synthetic anomaly detection

Table 7 and Fig. Figure 3 aggregates the 5% and 10% injected scenarios. NFGA-LINEAR has the highest macro-F1 on Cholera (0.568) and ILINet (0.467), whereas Prophet is the highest on Electricity (1.000). Across the six scenarios, Prophet won three, NFGA-LINEAR won two, and Robust-Z won one. The perfect electricity result is based on only two and four injected positives and must not be generalized to real operational anomalies.

Independent window detectors often produce high FPR, especially for cholera, indicating sensitivity to the distribution shift between training windows and injected test scenarios. The residual detector of NFGA-LINEAR

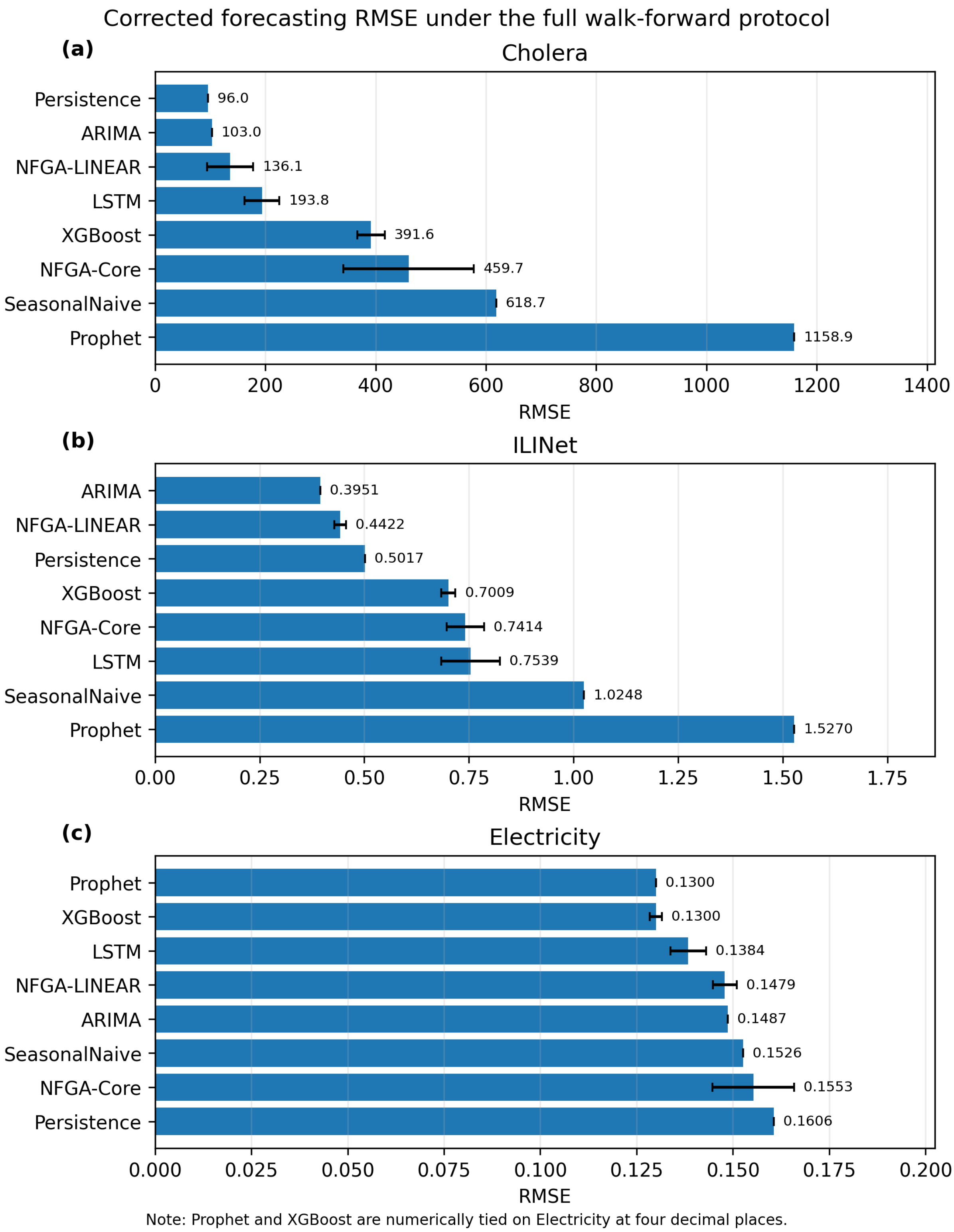


Figure 1. Forecast RMSE under the corrected full-horizon walk-forward protocol.

Table 6. Exact matched-seed Wilcoxon tests for RMSE. Positive rank-biserial values favor NFGA-LINEAR.

Dataset	Comparator	Mean benefit	95% CI	Rank-biserial	Holm p
Cholera	NFGA-Core	323.657	[257.152, 392.898]	1.000	0.0059
Cholera	XGBoost	255.559	[219.905, 286.697]	1.000	0.0059
Cholera	LSTM	57.731	[19.474, 91.832]	0.818	0.0195
ILINet	NFGA-Core	0.2992	[0.2747, 0.3263]	1.000	0.0059
ILINet	XGBoost	0.2587	[0.2427, 0.2727]	1.000	0.0059
ILINet	LSTM	0.3117	[0.2743, 0.3560]	1.000	0.0059
Electricity	NFGA-Core	0.00745	[0.00214, 0.01294]	0.745	0.0371
Electricity	XGBoost	-0.01784	[-0.01983, -0.01573]	-1.000	0.0059
Electricity	LSTM	-0.00945	[-0.01203, -0.00690]	-1.000	0.0059

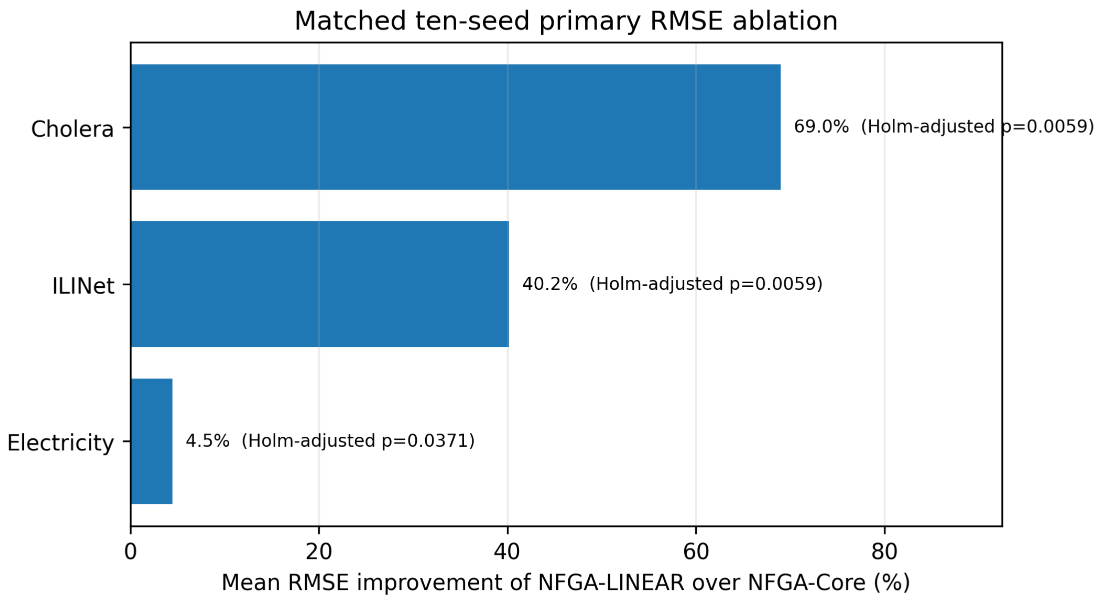


Figure 2. Matched NFGA-Core versus NFGA-LINEAR primary RMSE ablation.

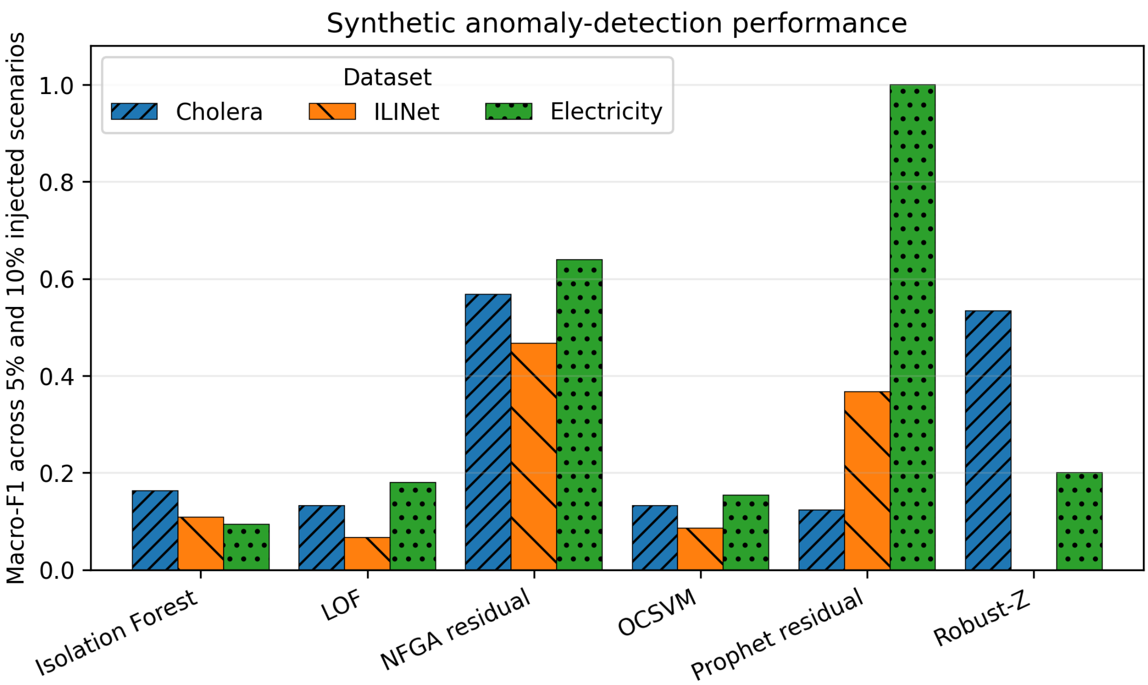


Figure 3. Macro-F1 for the main synthetic anomaly-detector comparison.

Table 7. Macro detector performance across 5% and 10% synthetic injection scenarios.

Dataset	Detector	Precision	Recall	F1	FPR
Cholera	NFGA-LINEAR residual	0.489	0.750	0.568	0.055
Cholera	Robust-Z	1.000	0.375	0.533	0.000
Cholera	Isolation Forest	0.092	0.894	0.163	0.710
Cholera	Local Outlier Factor	0.071	1.000	0.132	1.000
Cholera	One-Class SVM	0.071	1.000	0.132	1.000
Cholera	Prophet residual	0.068	0.688	0.123	0.609
ILINet	NFGA-LINEAR residual	0.313	0.990	0.467	0.186
ILINet	Prophet residual	0.286	0.533	0.367	0.104
ILINet	Isolation Forest	0.071	0.267	0.109	0.284
ILINet	One-Class SVM	0.048	0.433	0.086	0.604
ILINet	Local Outlier Factor	0.037	0.333	0.067	0.425
ILINet	Robust-Z	0.000	0.000	0.000	0.197
Electricity	Prophet residual	1.000	1.000	1.000	0.000
Electricity	NFGA-LINEAR residual	0.471	1.000	0.640	0.089
Electricity	Robust-Z	0.500	0.125	0.200	0.000
Electricity	Local Outlier Factor	0.099	1.000	0.180	0.684
Electricity	One-Class SVM	0.084	1.000	0.154	0.823
Electricity	Isolation Forest	0.054	0.375	0.094	0.481

Proxy robust-z labels are excluded from the main ranking because they are not verified ground truth and are circular for Robust-Z itself.

achieved high recall but incurred nonzero false positives; therefore, its usefulness depends on the operational cost assigned to missed events versus false alarms.

5.4. Sensitivity analysis

The baseline configuration was reproduced for all 30 dataset-seed rows within an absolute RMSE tolerance of 10^{-10}. None of the 33 OFAT comparisons remained significant after Holm correction within the dataset. Some alternatives yielded lower descriptive means, but the evidence did not justify replacing the prespecified baseline. For example, a 1.5-times electricity window reduced the mean RMSE from 0.147863 to 0.142423, but its raw $p = 0.0371$ became Holm $p = 0.3711$. Threshold sensitivity also showed that F1 could change with k. Because the same injected scenarios were used to summarize those changes, selecting the best k would require post-hoc optimization. Therefore, the main detector comparison retained $k = 3.5$. Figures 4 and 5 provide complete visual summaries.

5.5. Computational cost

Table 8 and Fig. 6 show the measured cost. NFGA-LINEAR fixed-model inference required 0.081 ms on Cholera, 0.103 ms on ILINet, and 0.110 ms on Electricity. In this environment, it was approximately 6–13 times faster than XGBoost and 89–143 times faster than LSTM for online inference. Its approximate parameter array footprint ranges from 1.91 to 2.81 KB. ARIMA and Prophet are not directly comparable in the online column because they were refitted at each test step.

6. DISCUSSION

6.1. Forecast accuracy, transparency, and computational cost

The empirical results indicate that forecasting performance is strongly dependent on the dataset. Low-complexity and conventional methods remain competitive on the evaluated short series: persistence achieves the lowest Cholera RMSE, while ARIMA performs best on ILINet. For Electricity, Prophet and XGBoost produced nearly identical RMSE values at the reported numerical precision, whereas NFGA-LINEAR occupied the fourth position in the descriptive ranking. Consequently, the evidence does not support the use of NFGA-LINEAR as a uniformly superior forecasting method. Its main advantage is relative to the matched neuro-fuzzy architecture from which it was derived. Replacing zero-order consequents with rule-specific linear functions yields statistically supported RMSE reductions for all three datasets. The improvements were particularly substantial for Cholera and ILINet and modest for Electricity.

NFGA-LINEAR also compares favorably with the evaluated stochastic black-box models for Cholera and ILINet,

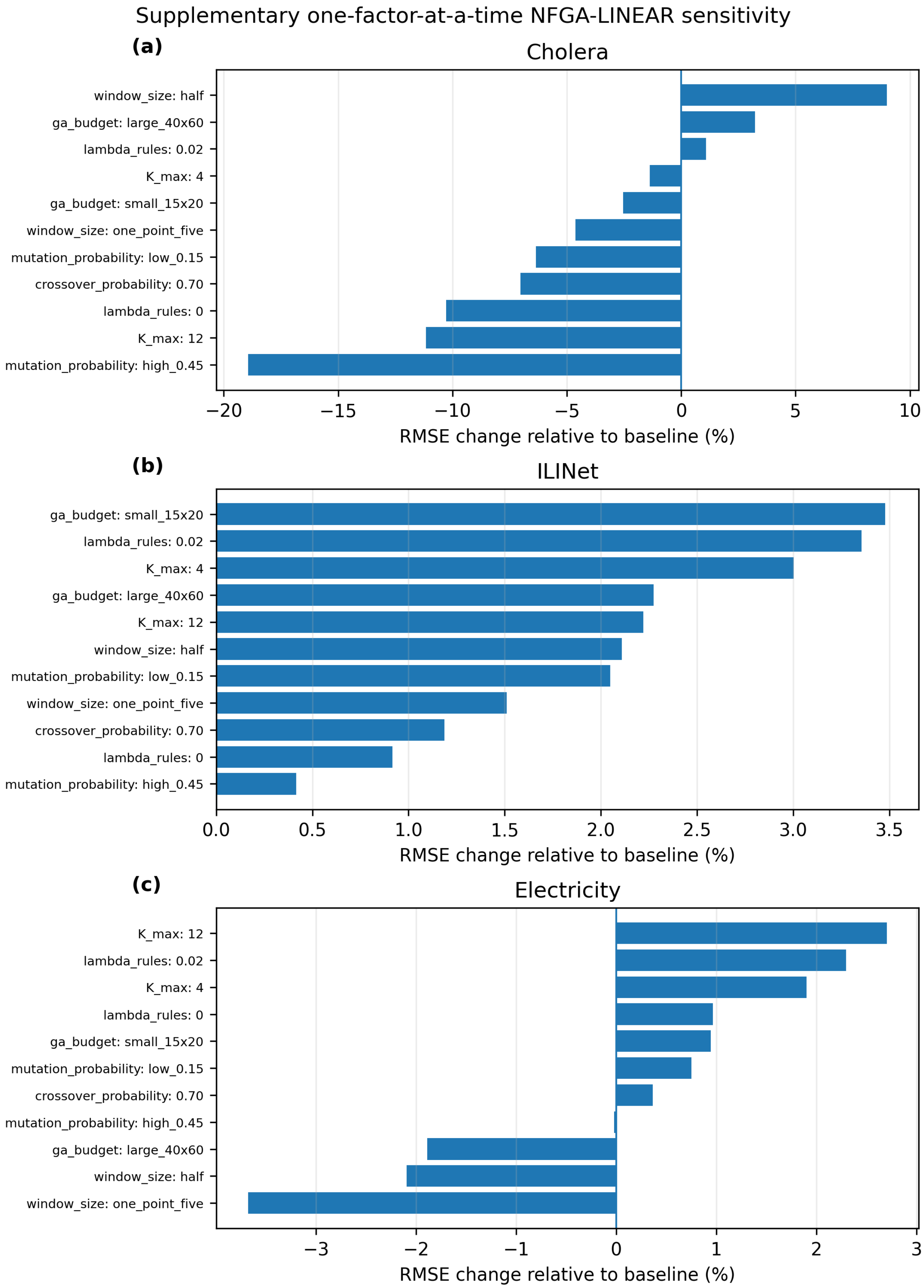


Figure 4. OFAT sensitivity of NFGA-LINEAR. Alternative configurations are descriptive; none is Holm-significant against the baseline.

Table 8. Measured computational costs in the experimental environment.

Dataset	Model	Training (s)	Online step (ms)	Approx. stored size (KB)
Cholera	ARIMA	N/A	16.505	N/A
Cholera	Prophet	N/A	222.272	44.51
Cholera	XGBoost	2.038	1.074	560.82
Cholera	LSTM	3.531	11.553	19.13
Cholera	NFGA-Core	1.095	0.076	3.12
Cholera	NFGA-LINEAR	0.798	0.081	2.07
ILINet	ARIMA	N/A	76.896	N/A
ILINet	Prophet	N/A	93.403	31.17
ILINet	XGBoost	0.123	0.650	175.31
ILINet	LSTM	2.591	10.961	19.13
ILINet	NFGA-Core	0.901	0.094	1.46
ILINet	NFGA-LINEAR	0.918	0.103	2.81
Electricity	ARIMA	N/A	31.638	N/A
Electricity	Prophet	N/A	62.603	26.65
Electricity	XGBoost	0.330	0.700	296.76
Electricity	LSTM	5.219	9.828	19.13
Electricity	NFGA-Core	0.773	0.077	1.12
Electricity	NFGA-LINEAR	0.940	0.110	1.91

ARIMA and Prophet online costs include per-step refitting. The NFGA sizes are parameter arrays, the LSTM sizes are weight arrays, the XGBoost sizes are serialized boosters, and the Prophet sizes are serialized fitted objects; these values are approximate and not strictly commensurate.

while retaining an explicit representation based on normalized rule activations and local affine consequents. Together with its low fixed-model inference latency and small approximate parameter footprint, these properties define its practical value as an accuracy–interpretability–cost compromise.

6.2. Dataset-dependent behavior of Prophet

Prophet was configured through chronological validation restricted to the training partition and was refitted at each forward-walk test step. Under this protocol, the performance varied markedly across the datasets. The selected additive trend and annual seasonality configuration modeled the electricity series effectively, producing an RMSE that differed from XGBoost by only approximately 1.1×10^{-6}. In contrast, it was less successful for Cholera and ILINet. One plausible interpretation is that the additive specification represented the smoother structure of the electricity series more adequately than the abrupt changes observed in Cholera or the particular temporal trajectory of ILINet. This explanation should be treated as an empirical interpretation rather than a general conclusion about the Prophet. Because Prophet was evaluated as a deterministic single-run baseline, its near tie with XGBoost on Electricity is descriptive and cannot be converted into a matched-seed significance claim.

6.3. Interpretation of the anomaly-detection results

The anomaly detection performance also depends on both the dataset and injected scenario. NFGA-LINEAR achieves the highest macro-F_1 on Cholera and ILINet, whereas Prophet leads Electricity and obtains the best results in three of the six individual injection scenarios. The perfect Prophet score on electricity must be interpreted cautiously because the corresponding 5% and 10% scenarios contain only two and four injected positives, respectively. Several independent window-based

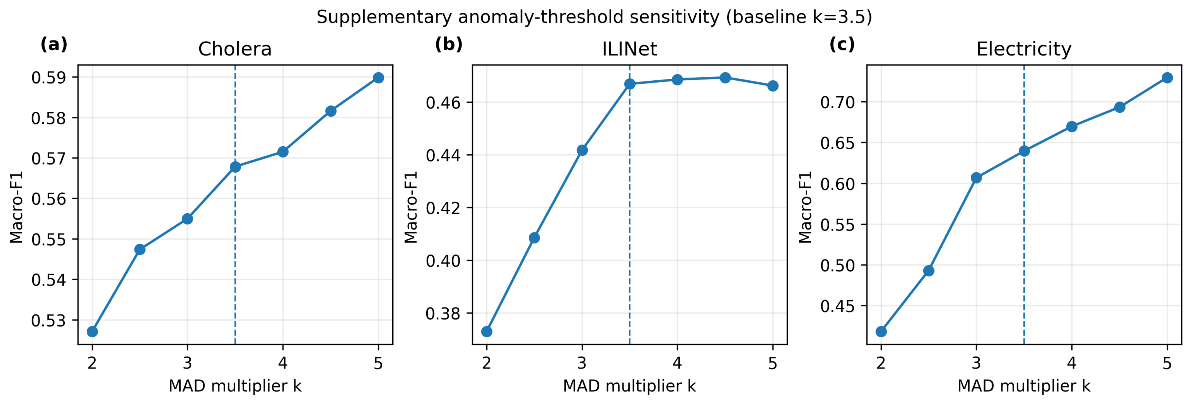
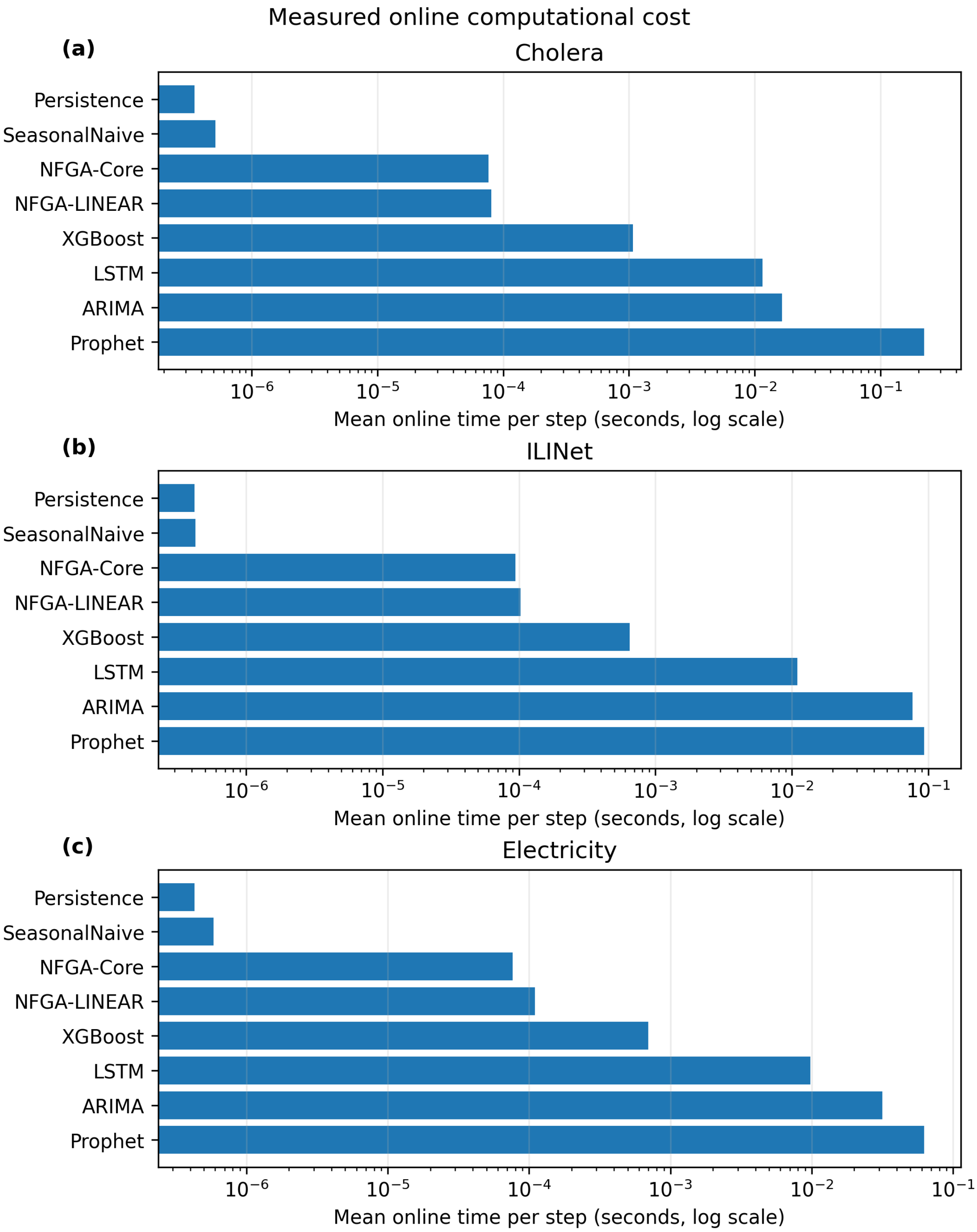


Figure 5. Residual-threshold sensitivity. The main analysis retains the prespecified $k = 3.5$.

detectors achieve high recall but also generate substantial false-positive rates. This behavior demonstrates that recall alone is insufficient for judging operational usefulness. A detector that flags nearly every injected event may still be impractical if it produces many false alarms. Reporting F_1 together with the false-positive rate provides a more balanced account of the detector behavior. The findings were limited to controlled synthetic perturbations. They did not establish that any evaluated detector would perform similarly in naturally occurring outbreaks, report irregularities, or detect sensor faults or other operational anomalies.



Note: ARIMA and Prophet include per-step refitting; the remaining models use fixed fitted predictors.

Figure 6. Online-step latency. ARIMA and Prophet include walk-forward refitting; the remaining models are fixed at inference time.

6.4. Limitations

Several limitations constrain the scope of these conclusions. First, the evaluation covered only three univariate weekly datasets and one-step-ahead forecasting. The findings may not directly transfer to multivariate, higher-frequency, irregularly sampled, or multi-horizon settings. Second, the synthetic anomaly experiments contained a small number of positive cases and represented only the specific injection mechanisms used in this study. They do not reproduce the full diversity of real epidemiological, reporting, or sensor-related anomalies. The Robust-Z proxy labels are exploratory and are not expert-verified ground truths; comparisons based on these labels are therefore excluded from the main detector ranking. Third, deterministic forecasting methods were evaluated using single walk-forward runs. They can be compared descriptively with stochastic models but cannot participate in matched random-seed statistical inference. Fourth, the sensitivity study varied one factor at a time. This design describes marginal responses around the prespecified configuration but does not estimate the interactions among the rule count, window length, regularization, mutation, crossover, and genetic budget. Fifth, the measured training times, inference latencies, and stored representation sizes depend on the software implementation, hardware environment, and measurement procedure. The storage values are also not strictly commensurate across model families because they represent parameter arrays, weight arrays, serialized boosters, or serialized fitted objects. Finally, the interpretability claim is structural, rather than user-validated. NFGA-LINEAR exposes rule activations, Gaussian antecedents, and local linear coefficients; however, direct inspection does not guarantee that domain specialists will find the resulting rules understandable or actionable. A formal user study is required to evaluate the practical interpretability.

7. CONCLUSION

This study introduced NFGA-LINEAR, a genetic neuro-fuzzy forecasting framework that replaces the constant consequences of NFGA-Core with local linear Takagi–Sugeno–Kang functions and combines the resulting forecaster with robust residual-based anomaly detection. The framework was evaluated using a leakage-controlled, full-horizon, one-step-ahead walk-forward protocol on Cholera, ILINet, and household electricity series. The results did not identify a single forecasting method that was the best for every dataset. Persistence leads to Cholera, ARIMA leads to ILINet, and Prophet and XGBoost are numerically almost identical to Electricity. The matched ten-seed ablation nevertheless provides consistent evidence that local linear consequents improve the evaluated architecture of the NFGA. Relative to NFGA-Core, NFGA-LINEAR reduced the mean RMSE by 69.0%, 40.2%, and 4.5% for Cholera, ILINet, and Electricity, respectively, with all three prespecified comparisons remaining significant after Holm correction. In the synthetic anomaly experiments, NFGA-LINEAR recorded the highest macro-F_1 on Cholera and ILINet, whereas Prophet performed best on Electricity. NFGA-LINEAR also offers a low fixed-model inference latency and compact approximate parameter representation. These findings support it as an interpretable and computationally economical forecasting alternative rather than a universally dominant forecaster or anomaly detector. Future research should examine multivariate and multi-horizon forecasting, expert-verified real-world anomaly labels, sensitivity designs that estimate hyperparameter interactions, adaptive online rule evolution, and formal assessments of whether domain experts can understand and use learned fuzzy rules.

REFERENCES

[1] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2015.

[2] S. J. Taylor and B. Letham, "Forecasting at scale," *The Am. Stat.*, vol. 72, no. 1, pp. 37–45, 2018. DOI: 10.1080/00031305.2017.1380080

[3] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. DOI: 10.1145/2939672.2939785

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: 10.1162/neco.1997.9.8.1735

[5] W. Peng et al., "Time series forecasting with a neuro-fuzzy modeling scheme," *Appl. Soft Comput.*, vol. 27, pp. 27–39, 2015.

[6] A. Jallal et al., "A hybrid neuro-fuzzy inference system-based algorithm for time series forecasting," *Expert Syst. with Appl.*, vol. 139, p. 112 828, 2020.

[7] S. Samanta et al., "Nfis-dn for time series forecasting," *Neural Comput. Appl.*, vol. 31, pp. 1237–1252, 2019.

[8] E. Rodrigues and A. Serra, "Distributed neuro-fuzzy network for time series prediction," *IEEE Trans. on Fuzzy Syst.*, vol. 28, pp. 1520–1532, 2020.

[9] M. Al-qaness et al., "Boosted anfis using marine predator algorithm for wind power forecasting," *Energy*, vol. 238, p. 121 832, 2022.

[10] H. Alrassas et al., "Optimized anfis model using aquila optimizer," *Appl. Soft Comput.*, vol. 115, p. 108 283, 2021.

[11] E. Kaya, A. Kaya, and C. Bastemur Kaya, "Adaptive network-based fuzzy inference system training using nine different metaheuristic optimization algorithms for time-series analysis of brent oil price and detailed performance analysis," *Symmetry*, vol. 17, no. 5, p. 786, 2025. DOI: 10.3390/sym17050786

[12] M. Abd Elfattah, A. A. Ewees, G. I. Sayed, A. Darwish, and A. E. Hassanien, "Carbon price time series forecasting utilizing an optimized anfis model," *Evol. Intell.*, vol. 17, no. 5–6, 2024. DOI: 10.1007/s12065-024-00955-2

[13] A. K. Behera, M. Panda, and S. Dehuri, "A recurrent anfis tuned by modified differential evolution for efficient prediction of software reliability," *Evol. Intell.*, vol. 17, no. 5–6, 2024. DOI: 10.1007/s12065-024-00940-9

[14] C. Bastemur Kaya, E. Kaya, and E. Siramkaya, “Hybrid anfis–mpa and ffnn–mpa models for bitcoin price forecasting,” *Biomimetics*, vol. 10, no. 12, p. 827, 2025. DOI: 10.3390/biomimetics10120827

[15] L. N. T. My, V. Nguyen, and T. Vo, “Learnable decomposition meets the neuro-fuzzy learning for robust rich spatial–temporal feature representation learning and better stock price forecasting,” *Soft Comput.*, vol. 29, no. 19-20, 2025. DOI: 10.1007/s00500-025-10898-0

[16] J. Huang, Q. Shi, and L. Kang, “A self attention neuro fuzzy inference system for time series prediction,” in *Frontiers in Artificial Intelligence and Applications*, vol. 417, 2026. DOI: 10.3233/FAIA251639

[17] J. Querales et al., “Explainability of machine learning models for hydrological time series forecasting,” *Environ. Model. & Softw.*, vol. 165, p. 105 363, 2023.

[18] K. Averkin and V. Yarushev, “Rules extraction from neural networks using neuro-fuzzy approaches,” *Neurocomputing*, vol. 470, pp. 188–201, 2022.

[19] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. Hoboken, NJ, USA: John Wiley & Sons, 1987, ISBN: 9780471852339.

[20] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation forest,” in *2008 Eighth IEEE International Conference on Data Mining*, 2008, pp. 413–422. DOI: 10.1109/ICDM.2008.17

[21] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF: Identifying density-based local outliers,” in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, ACM, 2000, pp. 93–104. DOI: 10.1145/342009.335388 [Online]. Available: https://doi.org/10.1145/342009.335388

[22] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the support of a high-dimensional distribution,” *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001. DOI: 10.1162/089976601750264965 [Online]. Available: https://doi.org/10.1162/089976601750264965