Sensitivity and Robustness Analysis of Hybrid Forecasting Models for Mission-Critical Energy Systems under Noisy Operational Conditions

STYLIANOS PAPPAS[1], ALEXANDROS GAZIS[2], NIKOS MASTORAKIS[3]
[1]Department of Electrical Engineering and Computer Science
Hellenic Naval Academy
Terma Chatzikyriakou, 18539, Piraeus
GREECE

[2]Department of Electrical and Computer Engineering,
Democritus University of Thrace,
Xanthi, 67100,
GREECE

[3]Department of Electrical Engineering and Computer Science
Hellenic Naval Academy
Terma Chatzikyriakou, 18539, Piraeus
GREECE

Abstract: Reliable operation of modern power systems demands accurate forecasting. Errors can reduce resilience, especially in mission-critical military, defense, and strategic infrastructures. While hybrid forecasting models perform well under normal conditions, their robustness to noisy data still needs thorough evaluation. In this work, a sensitivity and robustness analysis of a novel adaptive hybrid forecasting framework is examined using two scenarios of correlated Gaussian noise. These scenarios were selected to represent measurement degradation, communication failures, and physical disturbances commonly encountered in resource-limited defense operational environments. The proposed hybrid method integrates the Multi Model Partitioning Filter, Nonlinear Autoregressive Exogenous models, and a Genetic Algorithm for Resource Allocation, with performance evaluated using real commercial data. The results highlight the successful performance of the proposed method and underline the importance of the sensitivity analysis as a validation process for forecasting methods intended for defense energy systems that need to remain highly reliable at all times.

Key-words: Defense Energy Systems, Hybrid Forecasting, Mission Critical Infrastructure, Sensitivity Analysis, Robustness Assessment, Power System reliability, Gaussian Noise, Defense Energy Systems.

Received: May 16, 2025. Revised: June 29, 2025. Accepted: September 9, 2025. Published: December 31, 2025.

1 Introduction

Accurate forecasting is a essential for the reliable and safe operation of modern power systems. It helps with energy generation planning, reserve allocation, and system stability under continually changing operating conditions. Such requirements become stringent in autonomous grids, islanded microgrids, naval installations, and military facilities, where forecasting errors lead to undesired operational consequences, [1], [2].

Lately, the increasing penetration of renewable energy sources and the variation of load profiles have significantly increased the complexity of power system dynamics. Therefore, machine learning and hybrid forecasting approaches have been widely adopted, outperforming traditional forecasting methods, such as ARIMA-based approaches, which struggle to capture the nonlinear data characteristics of modern power systems, [3], [4], [5].

1.1 Sensitivity Analysis

A widely established assessment framework for quantifying how input uncertainty affects a system's outputs is sensitivity analysis (SA). In forecasting applications, SA has been applied to robustness evaluation under weather uncertainty, [1], uncertainty attribution in renewable power forecasting, [6], parameter tuning in neural networks, [7], and many more. The importance of SA in understanding model behavior under uncertainty, particularly in nonlinear and data-driven systems, is also emphasized in broad methodological reviews, [8], [9]. SA has also been used in evaluating the forecast robustness under groundwater systems, [10], showing how noise and measurement errors significantly affect predictive performance. Someone may come across similar findings in renewable energy bidding and market participation studies, where forecast errors influence operational and economic outcomes, [2], [11], [12].

1.2 Contribution of this work

Based on the literature, SA focuses on parameter sensitivity (learning rates, network depth, etc.), input relevance ranking, weather driven uncertainty in renewable energy generation. Relatively few studies address how noise affects the forecasting performance, [8], [9].

This paper extends prior hybrid forecasting work by introducing sensitivity analysis based on noisy data scenarios. Instead of modifying the model structure, controlled stochastic noise is applied to the input data to emulate realistic operational disturbances, including sensor inaccuracies, communication degradation, and cyber-physical perturbations. Different correlated Gaussian noise levels are examined to reflect increasing uncertainty conditions commonly encountered in military applications.

2 Problem Reformulation

The electric load forecasting task is redefined using an extension of the MMPF proposed in earlier studies, [13], [14]. In its standard formulation, the MMPF, partitions the estimation space into multiple parallel subfiters or submodels, each representing different characteristics of the system being investigated. Every submodel generates an independent probabilistic output estimate, and the final forecast is computed as a weighted combination of these individual results.

In the proposed approach, the MMPF employs Nonlinear Autoregressive Exogenous (NARX) models rather than Kalman Filters (KF) or Extended Kalman Filters (EKF). Each submodel functions as a nonlinear predictor trained to estimate the electric load, while accounting for exogenous variables that represent external factors, such as time-related and weekly patterns. Additional variants of the MMPF can be found in [15], [16], [17], [18], [19], [20], [21].

2.1 MMPF-NARX-GARA

The MMPF structure based on EKF, shown in Figure 1, was introduced in [13]. Each MMPF is composed of a parallel set of EKFs, typically ten, although a larger number may be employed depending on the complexity of the application, all running independently at each iteration. Equal initial posterior probabilities are assigned to every model, and as the process evolves, the probability of the most accurate model tends to one, while the others approach zero. The overall estimate is then computed as a weighted combination of the outputs from all submodels. This configuration enables parallel execution, significantly reducing computational time in real-time or high-dimensional tasks.

The optimal Minimum Mean Square Error (MMSE) estimate is given by:

$$\hat{x}(\kappa|k) = \sum_{i=1}^{M}(k|k;\theta)\rho(\theta|k)\,d\theta \quad (1)$$

The probabilities are calculated online in a recursive manner as follows:

$$p(\theta|k) = \frac{L(k|k\,;\theta)}{\sum_{i=1}^{M} L(k|k\,;\theta)p(\theta|k-1)d\theta}\,p(\theta|k-1) \quad (2)$$

where,

$$L\,(k|k;\theta) = \,|P_z\,(k|k-1;\theta)|^{\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}\,\|\tilde{z}(k|k-1;\theta)\| \cdot P_z^{-1}\,(k|k-1;\theta)\right\} \quad (3)$$

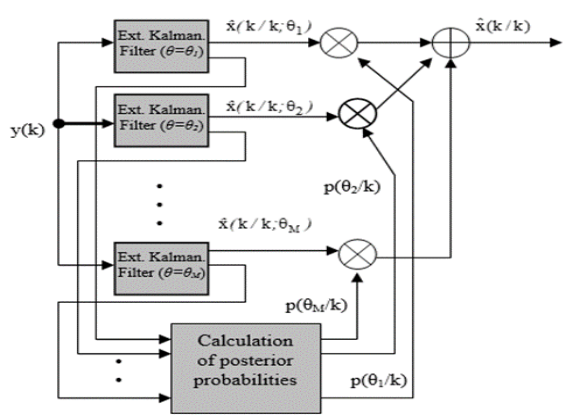

Fig. 1: Block diagram of MMPF+EKF. Every EKF implements an ARMA model.
Source: created by the authors.

In the proposed method, instead of using ARMA models in every EKF in Figure 1, Nonlinear Autoregressive Exogenous (NARX) models are implemented. Each submodel acts as a nonlinear predictor within the MMPF structure, as shown in Figure 2. The GARA module receives posterior probabilities from the MMPF and optimizes the weight allocation among the NARX submodels to minimize forecasting error.

The parallel training of the several NARX submodels aims to approximate the nonlinear relationship between the present and past values of the electric load, using both autoregressive and exogenous variables. The general form of each NARX submodel is expressed as:

$$\tilde{y}_i(k) = F_i\,(y(k-1), y(k-2),\ldots, y(k-p), u(k-1), \ldots\, u(k-q)) \quad (4)$$

where $F_i\,(\cdot)$ represents the nonlinear vector of the i^{th} submodel, p and q are the autoregressive and exogenous input vector orders, respectively. The predicted electric load at time instant k, by the MMPF-NARX, is denoted by $\tilde{y}_i(k)$.

In this case study, the exogenous input vector u(k) is defined as:

$$u(k) = \left[\sin\left(\frac{2\pi D_{hour}}{24}\right), \cos\left(\frac{2\pi D_{hour}}{24}\right), D_{week}(k), \sin\left(\frac{2\pi D_{month}}{24}\right), \cos\left(\frac{2\pi D_{month}}{24}\right), \overline{y}_{(k-6,k)}\right] \quad (5)$$

where:

- $D_{hour}\,(k)$ is the current hour of the day (1–24)
- $D_{month}\,(k)$ represents the month of the year (1-12)

both expressed in a unit circle. This encoding prevents discontinuities (i.e., from 23:00 to 0:00, or from the 12th month to the 1st). It effectively models the diurnal pattern of energy consumption.

- $D_{week}\,(k)$ is the day of the week (1–7)
- $\overline{y}_{(k-6,k)}$ is the moving average of the load, over the past six hours, defined as: $\frac{1}{6}\sum_{j=k-6}^{k-1} y(j)$

Each NARX submodel minimizes its own mean-squared error:

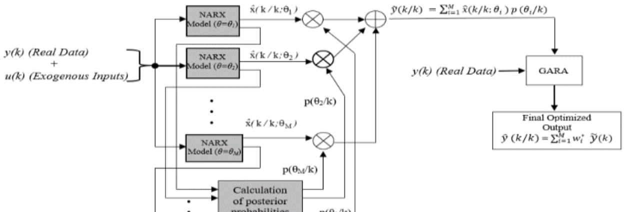

Fig. 2: The structure of the proposed MMPF-NARX-GARA method.
Source: created by the authors.

$$J_i = \frac{1}{M}\sum_{k=1}^{M}(y(k) - \tilde{y}_i(k))^2 \quad (6)$$

NARX has been successfully applied in load forecasting in several studies, [22], [23], [24], [25], [26].

2.2 Adaptive Genetic Resource Allocation (GARA)

Each chromosome represents a candidate weight vector:

$$W = [w_1\ , w_2\ ,\ldots\, w_M\,], \quad (7)$$

with $\sum_{i=1}^{M} w_i\ )= 1$, and $w_i\ \in [0,1]$

The first chromosome is initialized directly using the vector of posterior probabilities estimated by the MMPF: $\tilde{y}_k^{(i)} = y_k - h_i\left(\hat{x}_{k^1 k-x}^{(i)}\right)$with $\sum_{i=1}^{M} w_i\ )= 1$, and $w_i\ \in [0,1]$.

The fitness function measuring the forecasting accuracy is given by:

$$f(\boldsymbol{W}) = \frac{1}{N}\sum_{k=1}^{N}\left|y(k) - \sum_{i=1}^{M} w_i \tilde{y}_i(k)\right|^2 \quad (9)$$

The Genetic Algorithm evolves the population through its typical procedures of selection, crossover, and mutation until convergence. The mutation rate is defined as

$$\mu_t = \mu_0\left(1 - \frac{t}{t_{max}}\right) + \varepsilon \quad (10)$$

where,

- μ_t = is the mutation rate at generation t
- μ_0 = is the initial mutation rate
- t_{max} = maximum number of generations
- ε is the residual mutation term, ensuring minimal diversity near convergence

In this study, the value of μ_0 is 0.1. This leads to a moderate mutation (10%), preventing early convergence and keeping a balanced level of exploration. The weights assigned to each NARX submodel are adaptively optimized, so t_{max} was decided to be equal to 40. Finally, ε was set to 0.003 to introduce a small stochastic disturbance for maintaining diversity near convergence and preventing stagnation when fitness plateaus, without affecting stability.

The optimal weight vector is given by:

$$\boldsymbol{W}^* = argmin\, f(\boldsymbol{W}) \quad (11)$$

and the final optimized forecast is expressed by

$$\hat{y}\,(k/k) = \sum_{i=1}^{M} w_i^*\, \tilde{y}_i(k) \quad (12)$$

where, w_i^* represents the optimal weight assigned to the i[th] NARX submodel.

Genetic algorithms have also been widely used for optimizing forecasting models and improving prediction accuracy in energy forecasting problems, [27], [28], [29], [30], [31].

2.3 Mathematical Reformulation for the proposed MMPF-NARX-GARA

This section presents the theoretical formulation of the proposed method. The non-linear discrete-time state-space form is given by:

$$x_{k+1}^{(i)} = f_i\left(x_k^{(i)}, u_k\right) + w_k^{(i)}$$
(13)

$$y_k^{(i)} = h_i\left(x_k^{(i)}\right) + u_k^{(i)}$$
(14)

where:

- $x_k^{(i)}$ is the internal state vector of the i[th] submodel
- u_k represents the exogenous input vector
- $y_k^{(i)}$ is the predicted electric load
- $w_k^{(i)}$ is the zero mean process noise
- $u_k^{(i)}$ is the zero mean measurement noise

The non linear mapping $f_i(\cdot)$ corresponds to the NARX predictor, as indicated in Equation (4), therefore $F_i(\cdot)$ represents the non linear regression operator of each NARX submodel.

Due to the non-linearities of the system under investigation, the EKF is implemented. To avoid any misconception with the $F_i(\cdot)$, the Jacobian of the state transition function will be denoted as:

$$A_k^{(i)} = \frac{\partial f_i(x,u)}{\partial x}\Big|_{k-1|k-1}^{\hat{x}(i)}$$
(15)

And the measurement Jacobian as:

$$H_k^{(i)} = \frac{\partial h_i(x)}{\partial x}\Big|_{k|k-1}$$
(16)

The prediction step now is given by equations (17)-(23):

$$\hat{x}_{k|k-1}^{(i)} = f_i\left(\hat{x}_{k-1|k-1}^{(i)}, u_k\right)$$
(17)

$$P_{k|k-1}^{(i)} = A_k^{(i)} P_{k-1|k-1}^{(i)} \left(A_k^{(i)}\right)^T + Q_i$$
(18)

The update is described by the following equations:

Innovation: $\tilde{y}_k^{(i)} = y_k - h_i\left(\hat{x}_{k|k-1}^{(i)}\right)$
(19)

Innovation Covariance: $S_k^{(i)} = H_k^{(i)} P_{k|k-1}^{(i)} \left(H_k^{(i)}\right)^T + R_i$
(20)

Kalman Gain: $K_k^{(i)} = P_{k|k-1}^{(i)} \left(H_k^{(i)}\right)^T + \left(S_k^{(i)}\right)^{-1}$
(21)

State Update: $\hat{x}_{k|k}^{(i)} = \hat{x}_{k|k-1}^{(i)} + K_k^{(i)} \tilde{y}_k^{(i)}$
(22)

Covariance Update: $P_{k|k}^{(i)} = \left(I - K_k^{(i)} H_k^{(i)}\right) P_{k|k-1}^{(i)}$
(23)

This formulation ensures minimum mean square error estimation under noise assumptions.

The MMPF posterior probabilities are calculated using Bayes' rule:

$$p(\theta_i|Y_k) = \frac{p(y_k|\theta_i, Y_{k-1})p(\theta_i|Y_{k-1})}{\Sigma_{j=1}^{M} p(y_k|\theta_j, Y_{k-1})p(\theta_j|Y_{k-1})}$$
(24)

where

$$p(y_k|\theta_i, Y_{k-1}) = \mathcal{N}\left(0, S_k^{(i)}\right)$$
(25)

The overall forecast is given by:

$$\hat{y}_k = \sum_{i=1}^{M} p(\theta_i|Y_k)\hat{y}_k^{(i)}$$
(26)

In equations (24)-(26) $Y_k = \{y_1, y_2, \cdots, y_k\}$ denoting the set of observations available up to time instant *k*.

2.4 Stability of the MMPF-NARX architecture

Assuming that each NARX model is BIBO stable:

$$\left|y_k^{(i)}\right| \le C_i \qquad \forall k$$
(27)

The final forecast is a convex combination:

$$\hat{y}_k = \sum_{i=1}^{M} w_i(k) y_k^{(i)}$$
(28)

where

$$w_i(k) \ge 0, \sum_{i=1}^{M} w_i(k) = 1$$
(29)

Therefore,

$$|\hat{y}_k| \le \sum_{i=1}^{M} w_i(k) \left|y_k^{(i)}\right| \le \sum_{i=1}^{M} w_i(k)\, C_i \le \max_i C_i$$
(30)

So, the architecture remains BIBO stable provided that each submodel remains stable.

3 Data Preparation

The dataset covers the period from January 2020 to December 2025 and is used sequentially to preserve temporal dependencies. Approximately 60% of the data is employed for training the MMPF–NARX submodels, 20% for validation, and the remaining 20% for testing and prediction. Since the purpose of this work is to assess the robustness of the proposed method under measurement corruption, correlated Gaussian noise was added into the test segment (the 20% stated before), while the training and validation data sets were left noise-free. The prediction horizon is set to two months, January – February $\boldsymbol{S_1}$, April – May $\boldsymbol{S_2}$, August – September $\boldsymbol{S_3}$, corresponding to a medium-term load forecasting scenario and enabling, for the first time, an evaluation of the proposed hybrid method under such conditions. To assess robustness, 2 different noise scenarios were tested, namely:

- **Moderated Correlated Noise**: *correlation coefficient ρ = 0.95*, *noise ratio = 0.25*.
- **Heavy Correlated Noise**: *correlation coefficient ρ = 0.98*, *noise ratio = 0.50*.

emulating measurement uncertainty and data degradation effects.

Additionally, the proposed method will be tested using different numbers of NARX submodels. Specifically, different configurations implementing 7, 8, 9, and 10 NARX submodels will be considered for each noise scenario. This analysis will assess how the number of NARX submodels implemented affects the accuracy of prediction, the robustness, and the error characteristics under noise conditions.

Quantification of the model's forecasting performance across these scenarios provides valuable insight into robustness margins, enabling assessment of whether its architecture is suitable for deployment in strategic energy systems.

The electric load forecasting problem is reformulated based on the extension of the MMPF-NARX-GARA model as introduced in the previous section. The overall prediction is obtained as a weighted combination of these individual forecasts.

In the proposed method, the MMPF implements Nonlinear Autoregressive Exogenous (NARX) models, instead of KF or EKF. Each submodel acts as a nonlinear predictor trained to model the electric load, incorporating exogenous variables that.

4 Results

To assess the accuracy of the predictions, the Mean Absolute Percentage Error (MAPE) is calculated. The MAPE is defined as

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{A_i - P_i}{A_i}\right| \quad (31)$$

where A_i is the actual value, P_i is the predicted value, and n is the total number of samples.

Additionally, the Normalized Root Mean Square Error (NRMSE) is also calculated to quantify the deviation between the forecasted and the real values with increased sensitivity to large prediction errors. It is defined as:

$$\text{NRMSE} = \frac{1}{\bar{A}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - P_i)^2} \times 100\% \quad (32)$$

Where $\overline{A}$ is the average of the real values. The rest parameters are the same as for MAPE.

4.1 Moderated Correlated Noise: correlation coefficient ρ = 0.95, noise ratio = 0.25.

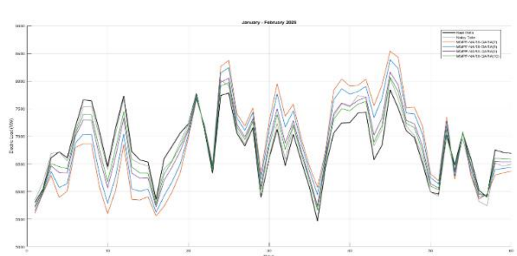

Fig. 3: January February 2025. Daily System Load for day ahead forecast, assessed by testing configurations with 7, 8, 9, and 10 NARX submodels.
Source: created by the authors.

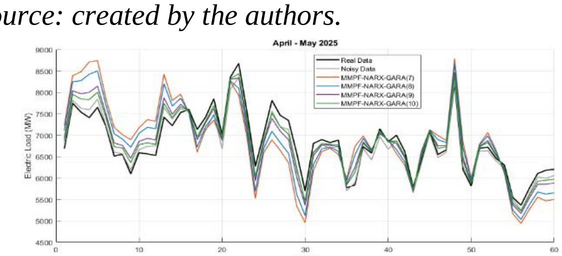

Fig. 4: April May 2025. Daily System Load for day-ahead forecast, assessed by testing configurations with 7, 8, 9, and 10 NARX submodels. Source: created by the authors.

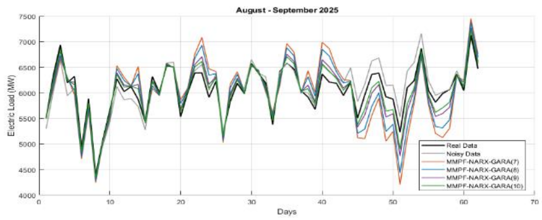

Fig. 5: August September 2025. Daily System Load for day-ahead forecast, assessed by testing configurations with 7, 8, 9, and 10 NARX submodels.
Source: created by the authors.

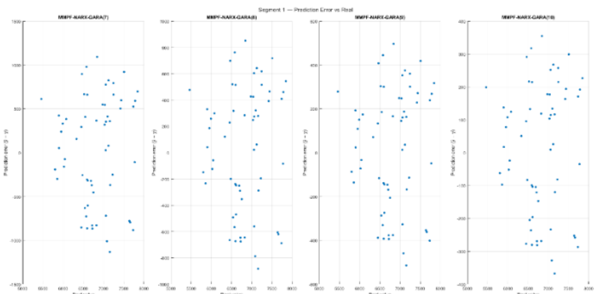

Fig. 6: Real and Forecasted Values for all testing configurations, January - February 2025.
Source: created by the authors.

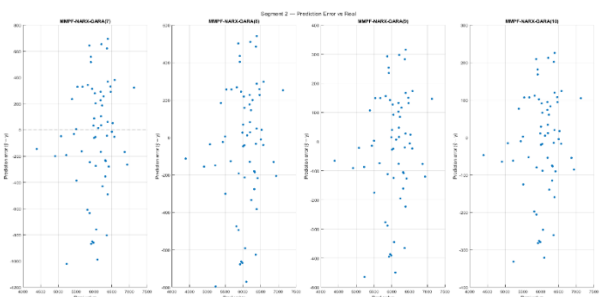

Fig. 7: Real and Forecasted Values for all testing configurations, April - May 2025.
Source: created by the authors.

Figure 3, Figure 4, and Figure 5 depict the performance of the proposed method with different numbers of implemented NARX models. The architecture implementing 10 submodels is closer to real data than the rest.

Figure 6, Figure 7, and Figure 8 depict that configurations using 8, 9, and 10 NARX submodels produce a narrower error distribution, with most values concentrated close to zero, while the configuration with 7 submodels presents the largest absolute prediction errors. However, the difference between the 9 and 10 submodel configurations is relatively small, meaning that the configuration with 9 NARX submodels can be considered optimum, providing a good balance between forecasting accuracy and computational cost.

The proposed method exhibits a mean-reversion behavior, meaning that it underestimates higher load values and overestimates lower ones. This behavior reflects the smoothing effect induced by ensemble filtering and the use of mean-squared error-based training objectives.

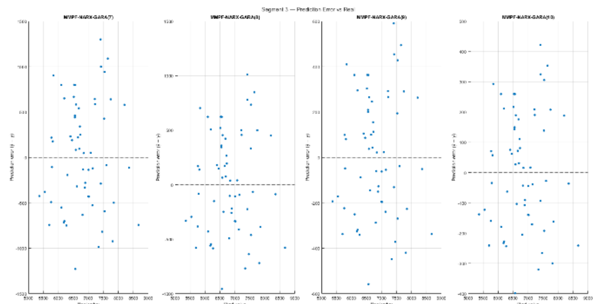

Fig. 8: Real and Forecasted Values for all testing configurations, August - September 2025. Source: created by the authors.

Table 1: Scenario 1. Accuracy Metrics per 2-month testing period

Number of NARX submodels implemented	MAPE % ***S_1 / S_2 / S_3***	Average MAPE %	NRMSE % ***S_1 / S_2 / S_3***	Average NRMSE%
7	5.6/5.5/5.5	5.50	6.4/6.6/6.5	6.50
8	3.9/3.6/3.7	3.73	5.6/5.5/5.7	5.60
9	2.7/2.9/2.8	2.80	3.3/3.2/3.2	3.23
10	2.1/2.6/2.5	2.67	3.0/3.1/3.0	3.03

Source: created by the authors.

Table 2: Scenario 1. Regression Coefficient per 2-month testing period

Number of NARX submodels implemented	***R^2*** *(Figure 7, Figure 8, and Figure 9)* ***S_1 / S_2 / S_3***	Average ***R^2***
7	0.63 / 0.66 / 0.64	0.64
8	0.74 / 0.72 / 0.0.73	0.73
9	0.76 / 0.78 / 0.77	0.77
10	0.81 / 0.83 / 0.82	0.82

Source: created by the authors.

Table 1 and Table 2 indicate that increasing the number of NARX submodels improves the performance across all test periods. This improvement becomes marginal beyond nine submodels. Therefore, the architecture implementing nine NARX submodels achieves a high level of robustness when compared to the one with ten submodels. Additionally, requires approximately 10% lower computational burden due to the implementation of one fewer model.

Additionally, the performance of the 8-submodel architecture is acceptable (MAPE ranging from 3.6-3.9%) and offers an overall 20% reduction in computational load compared to the ten-submodel configuration.

Consequently, the nine NARX submodels configuration may be seen as a favorable trade-off between model complexity and forecasting accuracy.

4.2 Heavy Correlated Noise: correlation coefficient ρ = 0.98, noise ratio = 0.50

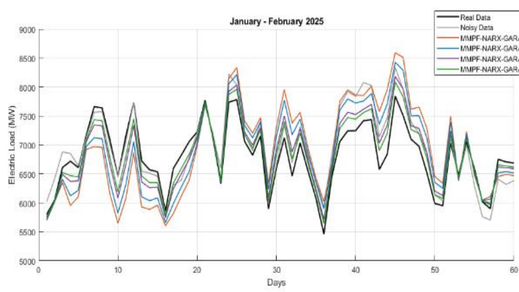

Fig. 9: January February 2025. Daily System Load for day-ahead forecast, assessed by testing configuration with 7, 8, 9 and 10 NARX submodels.
Source: created by the authors.

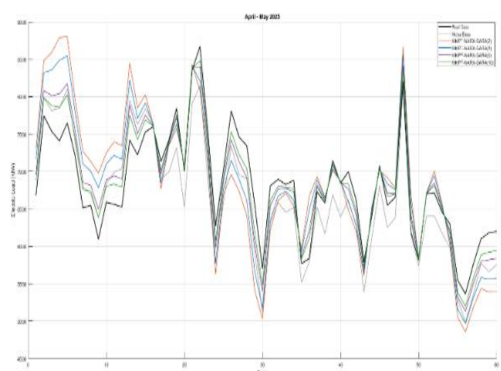

Fig. 10: April May 2025. Daily System Load for day-ahead forecast, assessed by testing configurations with 7, 8, 9, and 10 NARX submodels.
Source: created by the authors.

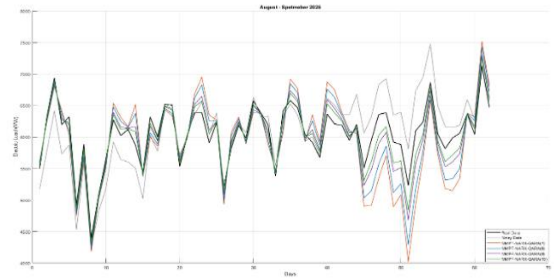

Fig. 11: August September 2025. Daily System Load for day-ahead forecast, assessed by testing configurations with 7, 8, 9, and 10 NARX submodels.
Source: created by the authors.

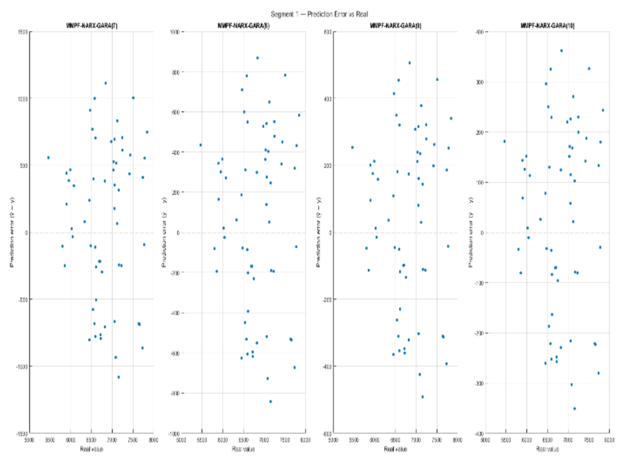

Fig. 12: Real and Forecasted Values for all testing configurations, January - February 2025.
Source: created by the authors.

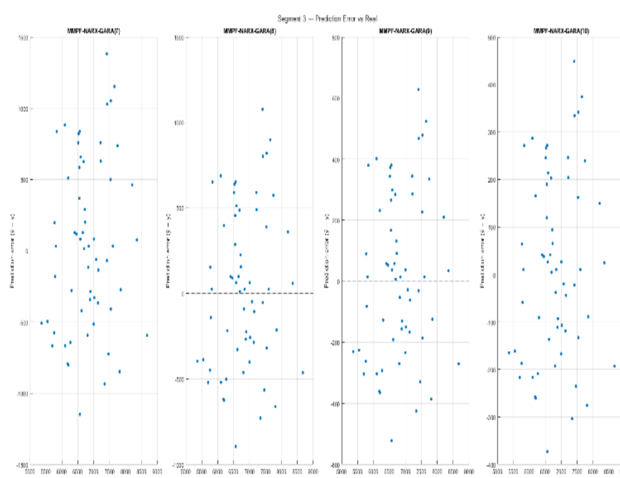

Fig. 13: Real and Forecasted Values for all testing configurations, April - May 2025.
Source: created by the authors.

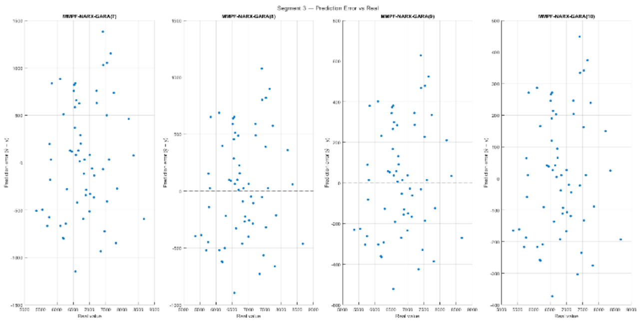

Fig. 14: Real and Forecasted Values for all testing configurations, August - September 2025.
Source: created by the authors.

Table 3: Scenario 2. Accuracy Metrics per 2-month testing period

Number of NARX submodels implemented	MAPE % S_1 / S_2 / S_3	Average MAPE %	NRMSE % S_1 / S_2 / S_3	Average NRMSE%
7	5.8/5.6/5.8	5.73	6.8/6.7/6.7	6.77
8	3.9/4.1/3.8	3.93	5.8/5.9/5.8	5.83
9	2.8/2.9/2.8	2.83	3.4/3.3/3.3	3.33
10	2.7/2.8/2.7	2.73	3.2/3.3/3.2	3.23

Source: created by the authors.

Table 4: Scenario 2. Regression Coefficient per 2-month testing period

Number of NARX submodels implemented	R^2 *(Figure 7, Figure 8, and Figure 9)* S_1 / S_2 / S_3	Average R^2
7	0.63 / 0.65 / 0.61	0.63
8	0.72 / 0.72 / 0.70	0.71
9	0.77 / 0.75 / 0.76	0.76
10	0.81 / 0.78 / 0.79	0.79

Source: created by the authors.

The conclusions obtained for the second scenario (Table 3, and Table 4) under heavy noise are similar to the ones obtained for the first. As it is proven, the proposed algorithm maintains robust performance for configurations with 8, 9, and 10 NARX submodels, leading to a reduction in the computational burden by 10% and 20% when 1 or 2 fewer NARX submodels are implemented. This feature is promising if the method is adapted to different applications demanding speed of convergence without a significant loss in accuracy. Similar results have been derived form Figure 9, Figure 10, Figure 11, Figure 12, Figure 13, and Figure 14.

Figure 15 depicts the performance improvement in terms of Average MAPE(%), Average NRMSE(%), and Average R^2 per number of NARX submodels used and for the different noise levels. It becomes obvious that the plateau is reached after using 9 submodels.

To assess the stability and the convergence of the GARA component, the Average Training MAPE (%) per iteration was calculated over 5 runs. Figure 16, Figure 17, Figure 18, Figure 19, Figure 20, and Figure 21 depict the convergence behavior for different numbers of NARX models used and for different noise levels. They also depict a very fast decrease in training error during the initial iterations. Minor oscillations can be observed as well as small fluctuations at the final iterations due to the stochastic nature of the procedure. These phenomena are confined to a narrow range, indicating the stability of the approach.

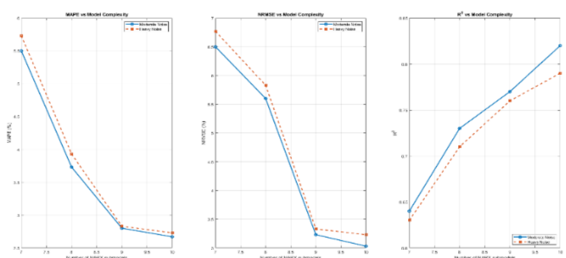

Fig. 15: Average MAPE(%), Average NRMSE(%), and Average R^2 against model complexity. The plateau between 9 and 10 NARX submodel architecture is obvious.
Source: created by the authors.

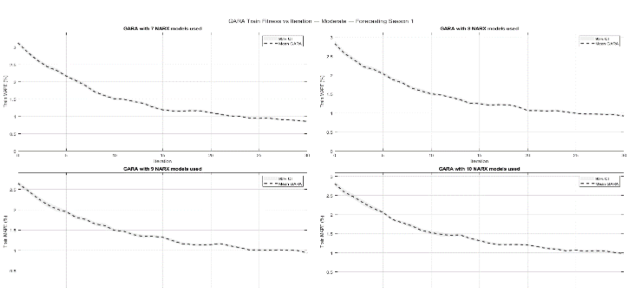

Fig. 16: Average Training MAPE per Iteration – Season 1 (Moderate Noise). Source: created by the authors.
Source: created by the authors.

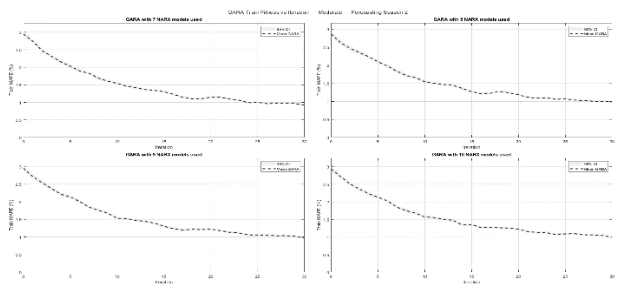

Fig. 17: Average Training MAPE per Iteration – Season 2 (Moderate Noise).
Source: created by the authors.

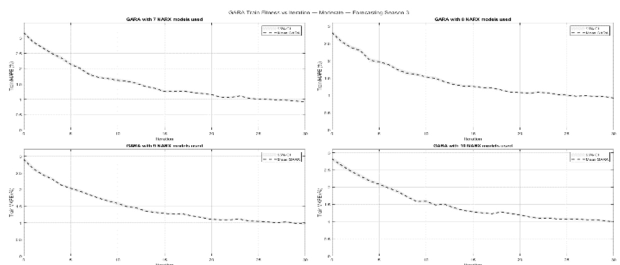

Fig. 18: Average Training MAPE per Iteration – Season 3 (Moderate Noise).
Source: created by the authors.

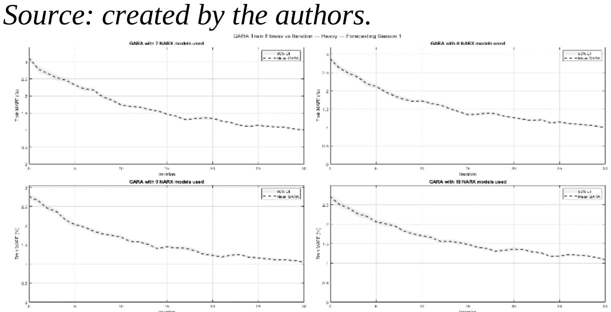

Fig. 19: Average Training MAPE per Iteration – Season 1 (Heavy Noise.
Source: created by the authors.

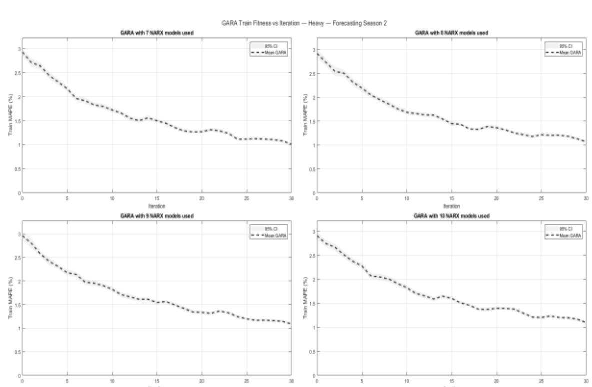

Fig. 20: Average Training MAPE per Iteration – Season 2 (Heavy Noise).
Source: created by the authors.

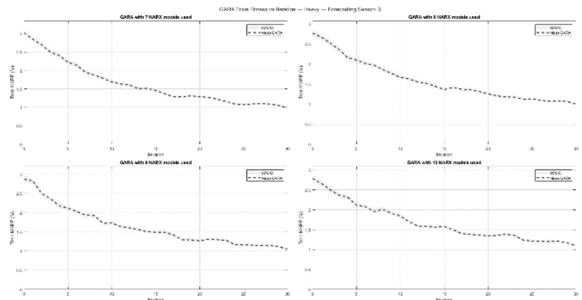

Fig. 21: Average Training MAPE per Iteration – Season 3 (Heavy Noise).
Source: created by the authors.

5 Conclusions

This paper investigates the robustness, sensitivity and stability of a hybrid MMPF–NARX–GARA forecasting method with the implementation of noise, which is representative in mission-critical energy systems. By introducing correlated Gaussian noise exclusively in the testing phase, the study provided a realistic sensitivity analysis without affecting the learning process.

The results demonstrate that the proposed architecture maintains high accuracy even under heavy measurement degradation. Increasing the number of NARX submodels improves, as expected, the performance; however, architectures with eight and nine submodels achieved a good balance between robustness and computational efficiency with minimal loss in accuracy.

Additionally, the improvement becomes marginal beyond nine submodels. These findings confirm the suitability of the proposed method for deployment in resource-constrained and defense-related energy applications where reliability is essential.

Further work was introduced to demonstrate the BIBO stability of the architecture, provided that every individual submodel remains stable. This is an important parameter in critical infrastructures where forecasting reliability is crucial for operational stability. The convergence plots presented in Figure 16, Figure 17, Figure 18, Figure 19, Figure 20, and Figure 21 prove the stability of the proposed method, showing fast reduction of the training error and limited oscillatory behavior due to the stochastic nature of the genetic algorithm.

Future work will focus on extending the sensitivity analysis to a wider range of noise scenarios, including different correlation structures and non-Gaussian disturbances, to further assess robustness margins and missing data scenarios. Additionally, the proposed framework will be applied to short-term and ultra-short-term forecasting horizons, enabling its evaluation in real-time operational settings and fast-response energy management systems.

References

[1] G. Chitalia, M. Pipattanasomporn, V. Garg, and S. Rahman, "Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks," Applied Energy, vol. 278, Art. no. 115410, 2020. https://doi.org/10.1016/j.apenergy.2020.115410

[2] J. Martinez-Rico, E. Zulueta, U. Fernandez-Gamiz, I. Ruiz de Argandoña, and M. Armendia, "Forecast error sensitivity analysis for bidding in electricity markets with a hybrid renewable plant," Sustainability, vol. 12, no. 9, Art. no. 3577, 2020. https://doi.org/10.3390/su12093577

[3] A. Mamun, M. Sohel, N. Mohammad, M. S. H. Sunny, D. R. Dipta, and E. Hossain, "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," IEEE Access, vol. 8, pp. 134911–134939, 2020. https://doi.org/10.1109/ACCESS.2020.3010702

[4] M. Hammad, B. Jereb, B. Rosi, and D. Dragan, "Methods and models for electric load forecasting: A comprehensive review," Logistics & Sustainable Transport, vol. 11, no. 1, pp. 51–76, 2020. https://doi.org/10.2478/jlst-2020-0004

[5] R. Jamil, "Hydroelectricity consumption forecast for Pakistan using ARIMA modeling," Renewable Energy, vol. 154, pp. 1-10, 2020. https://doi.org/10.1016/j.renene.2020.02.117

[6] M. Pierro, D. Gentili, F. Romano Liolli, C. Cornaro, D. Moser, A. Betti, M. Moschella, E. Collino, D. Ronzio, and D. van der Meer, "Progress in regional PV power forecasting: A sensitivity analysis on the Italian case study," Renewable Energy, vol. 189, pp. 983-996, 2022. https://doi.org/10.1016/j.renene.2022.03.041

[7] P. Mandal, A.K. Srivastava, M. Negnevitsky, and J.W. Park, "Sensitivity analysis of neural network parameters to improve electricity price forecasting," International Journal of Energy Research, vol. 33, no. 1, pp. 38-51, 2009. https://doi.org/10.1002/er.1469

[8] E. Borgonovo, and E. Plischke, "Sensitivity analysis: A review of recent advances," European Journal of Operational Research, vol. 248, no. 3, pp. 869-887, 2016. https://doi.org/10.1016/j.ejor.2015.06.032

[9] G. Qian, and A. Mahdi, "Sensitivity analysis methods in the biomedical sciences," Mathematical Biosciences, vol. 323, Art. no. 108306, 2020. ISSN: 0025-5564. https://doi.org/10.1016/j.mbs.2020.108306

[10] A. Amaranto, F. Pianosi, D. Solomatine, G. Corzo, and F. Munoz-Arriolla, "Sensitivity analysis of data-driven groundwater forecasts to hydroclimatic controls," Journal of Hydrology, vol. 587, Art. no. 124957, 2020. https://doi.org/10.1016/j.jhydrol.2020.124957

[11] Y. Song, Y. Yang, Z. He, C. Li, and L. Li, "A hybrid forecasting system based on multi-objective optimization for predicting short-term electricity load," Journal of Renewable and Sustainable Energy, vol. 11, no. 6, 2019. https://doi.org/10.1063/1.5109213

[12] F. Dakheel and M. Çevik, "Optimizing smart grid load forecasting via a hybrid long short-term memory–XGBoost framework," Energies, vol. 18, no. 11, pp. 2842, 2025. https://doi.org/10.3390/en18112842

[13] S. S. Pappas, "Application and comparison of evolutionary techniques for forecasting the Hellenic grid electricity load," International Journal of Power and Energy Research, vol. 1, no. 3, pp. 139-149, 2017. https://doi.org/10.22606/ijper.2017.13001

[14] S. S. Pappas, V. C. Moussas, and S. K. Katsikas, "Application of the multi-model partitioning theory for simultaneous order and parameter estimation of multivariate ARMA models," International Journal of Modelling, Identification and Control, vol. 4, no. 3, pp. 242-249, 2008. https://dx.doi.org/10.1504/IJMIC.2008.021161

[15] S. Katsikas, S. Likothanassis, G. Beligiannis, K. Berketis, and D. Fotakis, "Genetically determined variable structure multiple model estimation," IEEE Trans. Signal Processing, vol. 49, no. 10, pp. 2253-2261, 2001. https://doi.org/10.1109/78.950781

[16] G. Beligiannis, L. Skarlas, and S. Likothanassis, "A generic applied evolutionary hybrid technique," IEEE Signal Processing Magazine, vol. 21, no. 3, pp. 28-38, 2004. https://doi.org/10.1109/MSP.2004.1296540

[17] S. S. Pappas, "Adaptive Forecasting Techniques Applied to Short Time Wind Speed Forecasting," International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO), Athens, Greece, pp. 121-128, 2019. https://doi.org/10.1109/ICCAIRO47923.2019.00027

[18] V. C. Moussas, S. D. Likothanassis, S. K. Katsikas, A. K. Leros, "Adaptive online multiple source detection," Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. iv/1029-iv/1032, Philadelphia, PA, USA, 2005. https://doi.org/10.1109/ICASSP.2005.1416187

[19] D. G. Lainiotis, S. K. Katsikas and S. D. Likothanasis, "Adaptive deconvolution of seismic signals-performance, computational analysis, parallelism," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, no. 11, pp. 1715-1734, Nov. 1988, https://doi.org/10.1109/29.9009

[20] N. Nikitakos, A. Leros, S. Katsikas, S., "Towed array shape estimation using multimodel partitioning filters.", IEEE Journal of Oceanic Engineering, 2002, vol. 23, no. 4, pp. 380-384, https://doi.org/10.1109/48.725232

[21] A. Leros, V. Moussas, V. C. "Performance of the MMPF algorithm for DOA estimation under sensor failures," Proceedings of the IASTED International Conference of Intelligent Systems and Control, vol. 742, 2011 Computational Bioscience. https://doi.org/10.2316/P.2011.744-084

[22] J. Buitrago and S. Asfour, "Short-term forecasting of electric loads using nonlinear autoregressive artificial neural networks with exogenous vector inputs," Energies, vol. 10, no. 1, Art. no. 40, 2017. https://doi.org/10.3390/en10010040

[23] F. Abbas, D. Feng, S. Habib, U. Rahman, A. Rasool, and Z. Yan, "Short term residential load forecasting: An improved optimal nonlinear autoregressive (NARX) method with exponential weight decay function," Electronics, vol. 7, no. 12, Art. no. 432, 2018. https://doi.org/10.3390/electronics7120432

[24] K. Chreng, H. S. Lee, and S. Tuy, "A hybrid model for electricity demand forecasting using improved ensemble empirical mode decomposition and nonlinear autoregressive neural networks with exogenous inputs," Energies, vol. 15, no. 9, Art. No. 7434, 2022. https://doi.org/10.3390/en15197434

[25] N. Shirzadi, A. Nizami, M. Khazen, and M. Nik-Bakht, "Medium-term regional electricity load forecasting through machine learning and deep learning," Designs, vol. 5, no. 2, Art. No. 27, 2021. https://doi.org/10.3390/designs5020027

[26] N. Sultana, S. Hossain, S. H. Almuhaini, and D. Düştegör, "Bayesian optimization algorithm-based statistical and machine learning approaches for forecasting short-term electricity demand," Energies, vol. 15, no.9, Art. No. 3425, 2022. https://doi.org/10.3390/en15093425

[27] M. Jawad, S. M. Ali, B. Khan, C. A. Mehmood, U. Farid, Z. Ullah, S. Usman, A. Fayyaz, J. Jadoon, N. Tareen, A. Basit, M. A. Rustam, and I. Sami, "Genetic algorithm-based nonlinear autoregressive neural network for electrical load and wind speed forecasting," The Journal of Engineering, vol. 2018, no. 8, pp. 721-729, 2018. https://doi.org/10.1049/joe.2017.0873

[28] T. Akman, C. Yilmaz, and Y. Sönmez, "Short-term electric energy load forecasting of Ankara region using artificial intelligence methods," Journal of Polytechnic, vol. 26, np. 4, pp. 1517-1531, 2023. https://doi.org/10.2339/politeknik.911634

[29] N. Nauman, S. Akram, M. Rashid, A. Jaffar, S. Bhatti, and B. Fida, "Dual-model approach for concurrent forecasting of electricity prices and loads in smart grids: Comparison of sparse encoder NAR and GA-optimized LSTM," PLOS One, vol. 20, no. 9, pp. e0330024, 2025. https://doi.org/10.1371/journal.pone.0330024

[30] S. Ling, F. Leung, H. Lam, Y.-S. Lee, and P. Tam, "A novel genetic-algorithm-based neural network for short-term load forecasting," IEEE Transactions on Industrial Electronics, vol. 50, no. 4, pp. 793–799, 2003. https://doi.org/10.1109/TIE.2003.814869

[31] E. Uwimana and Y. Zhou, "A novel two-stage hybrid model optimization for accurate and stable short-term load forecasting," Technologies, vol. 12, no. 10, Art no. 194, 2024. https://doi.org/10.3390/technologies12100194

Contribution of individual authors to the creation of a scientific article (Ghostwriting Policy)

Stylianos Pappas developed the methodology, performed the simulations, data analysis, and manuscript preparation.

Alexandros Gazis reviewed and validated the results, contributed to the interpretation of the findings, and participated in the final editing of the manuscript.

Nikos Mastorakis supervised the research, contributed to the conceptual design of the study, reviewed the results, and participated in the final editing of the manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article