IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 November 2025, accepted 4 December 2025, date of publication 8 December 2025, date of current version 12 December 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3641712

RESEARCH ARTICLE

Optimizing Neural Network Ensemble for Short-Term Load Forecasting: A Case Study of a Brazilian Electrical Distribution System

FRANCO B. ROCHA[1], PAULO ROTELLA JUNIOR[2], PEDRO PAULO BALESTRASSI[3], FARID MELGANI[4], (Fellow, IEEE), AND ANTONIO C. ZAMBRONI DE SOUZA[5], (Senior Member, IEEE)
[1]Institute of Science, Federal University of Alfenas, Alfenas 37130-001, Brazil
[2]Department of Production Engineering, Federal University of Paraíba, João Pessoa 58051-900, Brazil
[3]Institute of Industrial Engineering and Management, Federal University of Itajubá, Itajubá 37500-903, Brazil
[4]Department of Information Engineering and Computer Science, University of Trento, 38122 Trento, Italy
[5]Institute of Electrical Systems and Energy, Federal University of Itajubá, Itajubá 37500-903, Brazil

Corresponding author: Pedro Paulo Balestrassi (pedro@unifei.edu.br)

This work was supported in part by the Research Support Foundation of the State of Minas Gerais (FAPEMIG), in part by the National Council for Scientific and Technological Development (CNPq), and in part by the Coordination for the Improvement of Higher Education Personnel (CAPES).

ABSTRACT This paper proposes a multivariate optimization framework to construct a neural-network ensemble for short-term load forecasting (STLF) at the distribution-substation level. A linear network, multilayer perceptron, and radial basis function network are combined via a {3,5} simplex-lattice mixture design on the probability simplex of ensemble weights. Correlated forecast-error metrics are reduced through factor analysis with principal-component extraction, and the resulting orthogonal factors are jointly optimized using the normal boundary intersection (NBI) method to generate a Pareto frontier of ensemble configurations. To select a single operating point, we adopt an entropy-based decision rule that maximizes the ratio between Shannon entropy of the mixing weights and a global percentage error (GPE) index, thereby balancing forecast accuracy and weight diversification. Tested on hourly load data from four Brazilian distribution substations over 24-, 48-, and 72-step horizons, the proposed ensemble reduces mean absolute percentage error (MAPE) by up to 20% relative to the best individual network, with consistent improvements also observed for RMSE and MAE. These gains are statistically supported by significance tests. The main contributions are: 1) casting ensemble-weight definition as a structured mixture-design problem; 2) integrating factor-analysis/PCA with NBI to handle multiple correlated error metrics in a principled way; and 3) introducing an entropy/GPE-based scalarization that yields accurate and robust ensembles suitable for implementation in operational smart-grid STLF workflows.

INDEX TERMS Neural network (NN), ensemble, short-term load forecasting (STLF), multiobjective optimization, distribution substations.

I. INTRODUCTION

Accurate short-term electricity load forecasting (STLF) at the distribution level is essential for guaranteeing the secure and economic operation of modern power systems. Reliable 24- to 72-hour predictions support feeder switching, energy-storage scheduling, and demand-response programs. They also help utilities balance demand and supply in real time and reduce technical losses [1], [2].

Over the last decade, neural networks (NNs) have become one of the most popular data-driven approaches to STLF thanks to their universal approximation capability and

The associate editor coordinating the review of this manuscript and approving it for publication was Ali Raza.

TABLE 1. List of abbreviations used in the paper.

Abbreviation	*Description*	*Abbreviation*	*Description*
ANN	Artificial Neural Network	PCA	Principal Component Analysis
EMS	Energy Management System	RBF	Radial Basis Function Network
GPE	Global Percentage Error	RMSE	Root Mean Square Error
HVdc	High-Voltage Direct Current	SCADA	Supervisory Control and Data Acquisition
LSTM	Long Short-Term Memory Network	STLF	Short-Term Load Forecasting
MAPE	Mean Absolute Percentage Error	BEM	Basic Ensemble Method
MAE	Mean Absolute Error	ARIMA	Autoregressive Integrated Moving Average
MDE	Mixture Design of Experiments	DMS	Distribution Management System
MLP	Multilayer Perceptron	MOP	Multi-Objective Optimization Problem
NBI	Normal-Boundary Intersection	CHIM	Convex Hull of Individual Minima
SVD	Singular Value Decomposition	CNN	Convolutional Neural Network

tolerance to noisy data [3], [4]. However, the accuracy of a single network is highly sensitive to its topology and training regimen. Ensemble learning mitigates this sensitivity by combining the strengths of diverse base learners, but existing schemes often rely on heuristic weight selection or ad hoc voting rules and do not fully exploit the multivariate structure of the forecast-error space [5], [6].

Classical ensemble-learning frameworks for STLF, such as bagging, boosting, and stacking, generally promote diversity via resampling or sequential reweighting of base learners. Their outputs are then combined through single-objective loss minimization or simple voting rules. In contrast, the present work adopts a design-of-experiments perspective in which the ensemble mixing weights themselves are explored through a simplex-lattice mixture design and tuned via multi-objective optimization over multiple error metrics, providing a structured alternative to these traditional schemes.

This work proposes a systematic multivariate optimization framework for building an NN ensemble dedicated to STLF in distribution substations. A linear NN, a multilayer perceptron and a radial-basis-function are blended through a $\{3,5\}$ simplex-lattice mixture design. Correlated error metrics are first reduced via principal component analysis (PCA). The compromise between Shannon entropy and overall percentage error is then resolved using the normal boundary intersection (NBI) method.

The proposed approach makes three main contributions: (i) it casts the definition of ensemble weights as a mixture-design problem on the probability simplex, using a $\{3,5\}$ simplex-lattice to systematically explore combinations instead of ad hoc or single-objective tuning; (ii) it integrates PCA-assisted reduction of multiple error metrics with NBI multi-objective optimization and an entropy/GPE scalarization that jointly balances forecast accuracy and weight diversification along the Pareto frontier; and (iii) it demonstrates, on four real Brazilian distribution substations, that this design-of-experiments perspective yields up to 20 % reductions in MAPE relative to the best individual network in an operational STLF setting.

The remainder of this article is organized as follows. Section II reviews the material and methods relevant in the literature; Section III details the proposed methodology; Section IV presents and discusses the experimental results; and Section V draws the main conclusions and outlines avenues for future work.

II. MATERIAL AND METHODS

A. NNS FOR STLF

NNs are universal, non-linear function approximators that can learn complex input–output mappings directly from data— making them well suited to capturing seasonality, abrupt demand swings, and other non-stationary behaviors in electric-load time series [7]. Their capacity to model such effects has made them a mainstay in STLF.

NN architectures differ in how neurons are connected, how weights are updated and which activation functions are employed. Feed-forward multilayer perceptrons (MLPs), radial-basis-function (RBF) networks and recurrent designs—particularly long short-term memory (LSTM) networks—are among the most widely used. MLPs excel at modelling global non-linear trends, RBF networks capture localized patterns, and recurrent architectures learn explicit temporal dependencies.

In typical STLF applications, lagged load observations together with exogenous variables (calendar, weather, socioeconomic indicators, etc…) could form the input vector, while the network outputs one-step-ahead or multi-step forecasts. Once trained, NNs help utilities anticipate daily peaks, schedule generation assets, and design demand-response programs more accurately than classical statistical models.

Early empirical evidence confirmed these advantages: Bhattacharyya and Thanh demonstrated sizeable accuracy gains with a feed-forward NN in Northern Vietnam [8], and Cavallaro obtained similarly promising results for an isolated Italian island using an MLP [9]. Subsequent studies refined forecast precision by integrating evolutionary optimization, hybrid preprocessing or ensemble learning.

For instance, Muralitharan et al. combined genetic algorithms and particle-swarm optimization to tune NN weights [10], while Voronin and Partanen fused wavelet transforms, ARIMA and NNs to improve both load and price forecasts [11]. Additional wavelet-NN hybrids reported by Ghayekhloo et al. [12] and Kumar et al. [13] further illustrate the value of multi-model approaches in capturing multi-scale load dynamics.

Nevertheless, single NNs remain sensitive to hyperparameter choice and data noise. Ensemble techniques that average or selectively weight the outputs of diverse networks mitigate this fragility and generally yield more reliable and generalizable STLF results. A neural network ensemble is a pathway to increased robustness and accuracy.

To the best of our knowledge, none of the existing studies uses exactly the same proprietary substation-level dataset considered here, which is drawn from a Brazilian distribution utility. Consequently, only qualitative comparisons with the literature are possible. Nevertheless, Table 2 summarizes a set of representative ensemble-based STLF and multi-objective optimization works on similar types of data, highlighting their data characteristics, model families, and reported performance metrics, and thereby positioning the proposed mixture-design + PCA + NBI framework within the existing body of research.

TABLE 2. Representative ensemble-based STLF and optimization studies in the literature.

Reference	*Data type / application*	*Models / ensemble approach*	*Optimization / combination scheme*	*Main reported metrics*
Jovanović et al. [5]	Heating-energy consumption of buildings	Ensemble of various neural networks	Simple averaging / voting of NN outputs	MAPE, RMSE
Saviozzi et al. [19]	Distribution-management system load forecasts	ANN ensembles for distribution feeders	Heuristic ensemble design within DMS platform	MAPE, RMSE
Jetcheva et al. [21]	Building-level electricity-load forecasts	Neural-network model ensembles	Weighted combination of NN models	MAPE, RMSE
Cervone et al. [20]	Short-term photovoltaic-power forecasting	ANN plus analog ensemble	Analog ensemble selection and combination	MAE, RMSE
H. G. R. et al. [18]	Residential short-term load for demand response	XGBoost–ANN ensemble	Gradient-boosting based feature /model combination	MAPE, RMSE

Recent works by Yousaf et al. on DC fault protection for meshed HVdc grids illustrate a complementary approach, in which discrete wavelet transform–based features are combined with Bayesian optimization to tune neural-network and LSTM hyperparameters such as depth, number of neurons, and learning rate for fault detection and fault location tasks [14], [15], [16]. In contrast, the present study keeps the individual network hyperparameters fixed after preliminary tuning and concentrates the optimization effort on a statistically guided mixture-design and NBI-based procedure for ensemble weight selection. These differences in scope position our contribution as orthogonal to Bayesian hyperparameter optimization in meshed HVdc protection schemes, while suggesting that such techniques could be naturally integrated into the proposed STLF framework in future work.

B. NN ENSEMBLE

Hansen Hansen and Salamon [17] were the first to demonstrate that the generalization capacity of a neural system could be improved by combining several neural networks. Figure 1 shows the basic structure of an ensemble. In the first step, each NN $_i$ ($i = 1, 2, \ldots, k$) (with the same or different input data) belonging to the ensemble is trained to produce an output $y_1, y_2, \ldots, y_k$, which will be weighted to generate the ensemble output. In this work, we mix the residual series from each network rather than the raw outputs; the combined residual is then added back to obtain the final forecast.

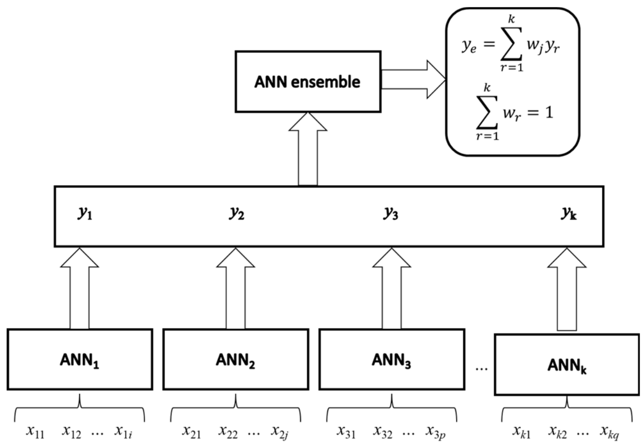


FIGURE 1. -Conceptual structure of the neural-network ensemble. Schematic diagram of the proposed ensemble for short-term load forecasting, in which three base models (Linear, RBF, and MLP) receive the same input features and their outputs are combined through optimized mixing weights to produce the final load forecast.

The literature includes studies on using NN ensembles to make predictions [18]. Saviozzi et al. [19] presented the design and implementation of two functionalities in a real distribution system: load forecasting and load modeling. The algorithms were based on an NN ensemble, performing well. In this work, load prediction was performed using the basic ensemble method (BEM), which defines the combined output as the average of the individual network outputs. The gap our research fills is determining how to weigh individual network outputs using a mixture design.

Cervone et al. [20] present a methodology based on a neural network and an analog ensemble to generate forecasts for three solar power plants in Italy and indicate that the combined solution produces the best results. Jetcheva et al. [21] demonstrated an ensemble model of a neural network for next-day electricity forecasting, surpassing a seasonal ARIMA model by 50%. This article also describes an automated method for selecting the model's parameters, facilitating its implementation in real problems. The following section presents the dimensionality reduction and multiobjective optimization tools that will be used to determine the weighting used in the NN ensemble.

C. FACTOR ANALYSIS

When many response variables are studied, they are often correlated. This motivates the use of statistical techniques that summarize the correlation structure in a smaller set of variables. According to Johnson and Wichern [22] factor analysis is a multivariate statistical technique that describes correlated variables in groups of factors, that is, it is a technique that describes the covariance structure between the response variables $y_1, y_2, \ldots, y_q$ in terms of factors $F_1, F_2, \ldots, F_s$ $(s < q)$. This technique reduces redundant information across variables by representing them with a smaller number of latent variables. According to Almeida et al. [23], the latent and original variables can be related to (1).

$$\mathbf{X}-\mu=\mathbf{LF}+\epsilon \tag{1}$$

where $X = [X_1, X_2,.., X_q]^T$ is an observable random vector with q components with mean vector $\mu = [\mu_1, \mu_2, \ldots,\mu_q]^T$, $F = [F_1, F_2, \ldots, F_s]^T$ is a random vector of unobservable variables $(s < q)$, $\epsilon = [\epsilon_1, \epsilon_2, \ldots,\epsilon_q]$ is the vector containing errors and L is the factor loading matrix ($q \times s$ dimension), whose elements represent the covariance (or correlation) between the latent and response variables [24], obtained through a spectral decomposition of the variance and covariance matrix Σ. Thus, the factor loadings are the eigenvectors of the variance-covariance matrix Σ, as in (2), scaled by $\sqrt{\lambda_q}$. Being $\Sigma = E[(x-\mu)(x-\mu)^T$, where T denotes the transposed matrix, it follows that:

$$\begin{aligned}\mathbf{\Sigma} &= E\left[(x-\mu)(x-\mu)^T\right] = E\left[(LF-\epsilon)(LF-\epsilon)^T\right]\\ &= E\left[LF\left(L^TF^T\right)+\epsilon\left(L^TF^T\right)+(LF)\epsilon^{T}+\epsilon^T\right]\\ &= \frac{\mathrm{L}\underbrace{E\left[FF^T\right]}L^T}{I}+\frac{\underbrace{E\left[\in F^T\right]}L^T}{0}+\frac{L\underbrace{E\left[F\in^T\right]}}{0}\\ &\quad+\frac{\underbrace{E\left[\in\in^T\right]}}{\Psi}=LL^T+\Psi\end{aligned} \tag{2}$$

where Ψ_{qxq} is a diagonal matrix formed by the specific variances ψ_i, as in (3), it soon follows that:

$$\sigma_i = Var(X_i) = \left(l_{i1}^2 + l_{i2}^2 + \ldots + l_{iq}^2\right) + \psi_i = h_i^2 + \psi_i \tag{3}$$

where $h_i^2 = l_{i1}^2 + l_{i2}^2 + \ldots + l_{iq}^2$ is called commonality (part of the variance of X_i explained by common factors). To simplify the interpretation of the factor loadings l_{ij}, the factor loading matrix is often rotated. Among the rotation methods, the most used is varimax rotation [24]. This rotation method selects a matrix $\mathbf{T}$ that maximizes the variance function (4):

$$\mathrm{V} = \frac{1}{p}\sum_{j=1}^{s}\left[\sum_{i=1}^{q}\tilde{l}_{ij}^4 - \left(\sum_{i=1}^{q}\tilde{l}_{ij}^2\right)^2 \Big/ p\right] \tag{4}$$

where $\tilde{l}_{ij} = \frac{l_{ij}}{\sqrt{h_i^2}}$ is the factor loading rotated and scaled by the square root of the i th commonality [24]. Considering a matrix of standardized data Z of X and the matrix of loadings L, the matrix of factor scores (uncorrelated) F is given by (5):

$$F = Z\left(L\left(L^TL\right)^{-1}\right) \tag{5}$$

To determine s, Jolliffe [25] suggests Kaiser's rule [26]. This rule selects only the main components with eigenvalues greater than one (using the correlation matrix), and a cumulative variance of at least 80%.

D. MULTIOBJECTIVE OPTIMIZATION

In a multiobjective optimization (MOP) problem, multiple responses can present conflicting objectives, and individual optimization can lead to different solutions [27]. Thus, the goal is to find a solution that satisfies the objective functions and their restrictions. There are several techniques for solving MOP. Among them, the normal boundary intersection method (NBI) deserves attention. This method, proposed by Das and Dennis [28], obtains a Pareto boundary in MOP that does not depend on the relative scales of objective functions, producing equispaced Pareto-optimal solutions. However, this method can lead to inadequate results if the responses to be optimized are correlated or have conflicting objectives. This problem can be solved if the Pareto boundary is generated from non-correlated functions, represented by functions that represent the scores of the factors in factor analysis through the extraction by principal components (PCA) [29]. The PCA technique obtains the eigenvalues from the variance and covariance matrix or the correlation matrix, using them as weights to multiply the default values of the original dataset [30]. In this work, we use NBI with rotated factor-score functions to determine the NN ensemble weights. The NBI method is presented and summarized in Figure 2. Figure 3 illustrates the NBI method in normalized space for two objective functions.

The NBI method starts with the calculation of the elements of the matrix payoff as in (6) [32]:

$$\Phi = \begin{bmatrix} f_1^*\left(x_1^*\right) & \cdots & f_1\left(x_i^*\right) & \cdots & f_1\left(x_m^*\right)\\ \vdots & \ddots & & & \vdots\\ f_i\left(x_1^*\right) & \cdots & f_i^*\left(x_i^*\right) & \cdots & f_i\left(x_m^*\right)\\ \cdots & & & \ddots & \vdots\\ f_m\left(x_1^*\right) & \cdots & f_m\left(x_i^*\right) & \cdots & f_m^*\left(x_m^*\right)\end{bmatrix} \tag{6}$$

whose main diagonal elements represent the individual optimal values of the objective functions. Together, these values form a utopia point $f^U = \left[f_1\left(x^{1*}\right)\ldots f_i\left(x^{i*}\right)\ldots f_j\left(x^{j*}\right)\right]$.

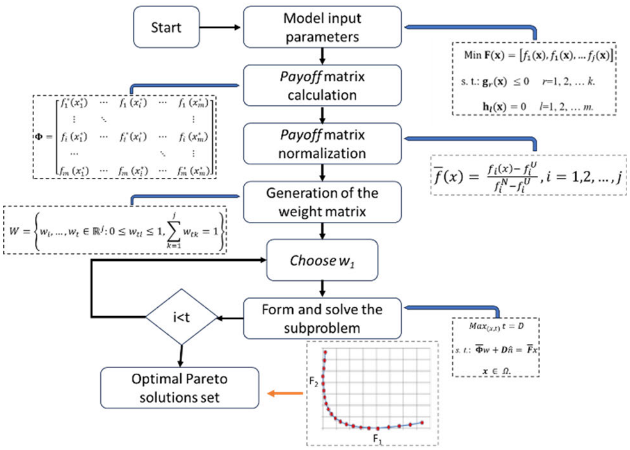


FIGURE 2. Main steps of the NBI multiobjective optimization method. Overview of the normal-boundary-intersection (NBI) procedure used in this work to generate Pareto-efficient solutions for the multiobjective optimization of ensemble weights.

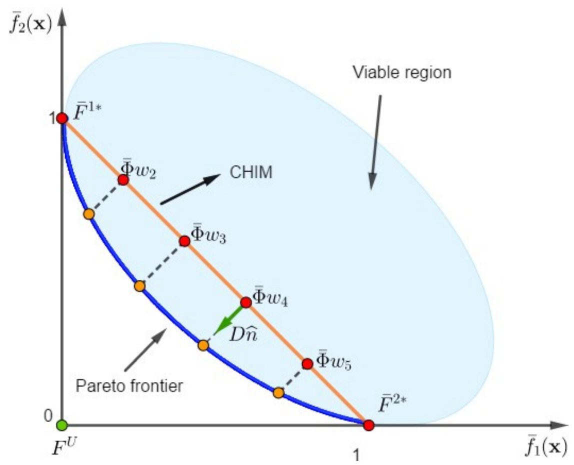


FIGURE 3. Illustration of the NBI method in normalized objective space. Graphical representation of the NBI approach showing the construction of the convex hull of individual minima, the generation of evenly spaced reference points on the utopia line, and the determination of the corresponding Pareto-optimal solutions in normalized space (adapted from Vahidinasab and Jadid [31]).

Additionally, Nadir point, denoted by $f^N = \left[f_1^N \ldots f_i^N \ldots f_j^N \right]$ are the worst value found for each i_{th} objective function (worst values of each line in the Φ matrix). When the objective functions have different scales, they are normalized using (7):

$$\bar{f}\,(x) = \frac{f_i\,(x) - f_i^U}{f_i^N - f_i^U}, i = 1, 2, \ldots, j \tag{7}$$

to get the normalized matrix $\bar{\Phi}$. According to Vahdinasab and Jadid [31] the convex combinations of each row of the normalized payoff matrix will form the set of points in "The Convex Hull of Individual Minima" (CHIM), also called the Utopia line (8), i.e, for $w_i = (w_{i1}, \ldots, w_{ij}) \in R^j$ and

$$W = \left\{ w_i, \ldots, w_t \in \mathbb{R}^j \colon 0 \leq w_{ti} \leq 1, \sum_{k=1}^{j} w_{tk} = 1 \right\} \tag{8}$$

where t is the number of optimization subproblems to be solved, the CHIM set is as in (9).

$$\bar{\Phi}_w = \left\{ \bar{\Phi} w_1, \bar{\Phi} w_2, \ldots, \bar{\Phi} w_t \right\}. \tag{9}$$

Since $\hat{n}$ is a normal vector to the utopia line at the points $\bar{\Phi}_w$ in the direction of the origin, $\bar{\Phi}_w + D\hat{n}, D \in R$ is the vector equation of the normal line to the utopia line passing through the points of $\bar{\Phi}_w$. The point at which $\bar{\Phi}_w + D\hat{n}$ intersects the boundary of the viable region closest to the origin will correspond to the maximization of the distance (D) between the utopia line and the boundary of the viable region, thus marking an ideal solution point on that frontier. Thus, the NBI method can be written as a nonlinear programming problem as in (10) [33]:

$$\begin{aligned} & Max_{(x,t)} t = D \\ & s.t.: \overline{\Phi} w + D\hat{n} = \overline{F} x \\ & \quad x \in \Omega. \end{aligned} \tag{10}$$

And it must be resolved iteratively to generate an equispaced boundary. For the bi-objective case, a choice to generate the weight matrix is through the relationship $w_j = 1 - \sum_{i=1}^{n} w_i$ [34], and problem 10 can be written in the form (11):

$$\begin{aligned} & Min \bar{f}_1(x) \\ & s.t.: \bar{f}_1\,(x) - \bar{f}_2\,(x) + 2w - 1 = 0 \\ & \quad g_r\,(x) \leq, r = 1, 2, \ldots, k \\ & \quad 0 \leq w \leq 1 \end{aligned} \tag{11}$$

Let $\{z(s)\}_{s=1}^{S}$ denote the set of Pareto solutions returned by the NBI procedure, ordered along the front. We use the index s to refer to the s-th NBI solution.

E. SHANNON ENTROPY AND GLOBAL PERCENTAGE ERROR

According to Taboada et al. [35] the optimal Pareto solutions set includes all rational decisions, with the decision maker being responsible for choosing a solution among the existing solutions. However, due to the complexity of assigning the weights to the objective functions, decision makers find it difficult to choose a solution. Thus, several multiobjective decision-making techniques can be used. Among them, we can highlight the metric developed by Rocha et al. [36], combining the global percentage error index and the Shannon entropy index.

Entropy quantifies uncertainty in a probability vector and reaches its maximum when all outcomes are equally probable. In our NN mixture, the ensemble-weight vector $\mathbf{w} = (w_1, \ldots, w_z)$, with $w_i \geq 0$ and $\sum w_{\text{i}} = 1$, represents the mixing proportions assigned to the base models (RBF, Linear, MLP). Following Shannon [37], we compute Eq. (12) and use it to favor diversified ensembles (higher entropy) over overly concentrated ones.

$$H = -\sum_{i=1}^{m} x_i ln\,(x_i) \tag{12}$$

Here, x_i denotes the i-th component of the ensemble mixing-weight vector $w = (w_1, \ldots, w_z)$, that is, the weight assigned to the i-th base network used in the entropy calculation.

Another strategy for determining weighting is the Global Percentage Error (GPE). The GPE calculates the sum of the differences in solutions at the Pareto frontier in relation to their targets. GPE is calculated using the expression (13):

$$GPE = \sum_{i=1}^{j} \left| \frac{y_i^* - T_i}{T_i} \right| \tag{13}$$

where y_i^* represents the values of the Pareto-optimal responses, T_i are the defined targets, and j *is* the number of objectives.

In this work, to choose a solution among the Pareto-optimal solutions, a measure that uses both Shannon's entropy and GPE will be used; more specifically, the criterion will be the maximum ratio between entropy and global percentage error as in (14) [38]:

$$\zeta = \frac{H}{GPE} \tag{14}$$

The maximum value of ζ will correspond to the highest entropy (resulting in weight diversification) and the minimum distance between the optimal Pareto solution and the targets.

To further assess the robustness of this choice, we also implemented an alternative scalarization based on the Euclidean distance to the utopia point in the objective space. For each Pareto-efficient solution s, we computed $d_U(s) = \sqrt{\sum_j (f_j^{(s)} - f_j^U)^2}$, where $f_j^{(s)}$ denotes the normalized value of the j-th objective and f_j^U is its utopia target, and ranked the solutions by increasing d_U. Comparing the operating points selected by the entropy/GPE ratio and by the utopia-distance criterion shows that both approaches typically identify solutions lying in the same short segment of the frontier, with very similar forecast errors. However, in regions where several Pareto solutions have comparable distances to the utopia point, the entropy/GPE-based criterion systematically favors ensembles with more diversified weight vectors, providing a better balance between accuracy and model diversification. In the present STLF context—where robustness to model misspecification and data perturbations is important in addition to point-forecast accuracy—this property makes the entropy/GPE scalarization preferable to a purely distance-based selection rule.

III. METHODOLOGY

In what follows, we adopt the following notation (summarized in Table 3). For each time instant t, y_t denotes the actual load and $\hat{y}_t^{(m)}$ the forecast produced by the m-th base network, with residual $e_t^{(m)} = y_t - \hat{y}_t^{(m)}$. The ensemble forecast is written as $\hat{y}_t = \sum_{m=1}^{z} w_m \hat{y}_t^{(m)}$, where $w = (w_1, \ldots, w_z)$ is the vector of non-negative mixing weights satisfying $\sum_m w_m = 1$. The performance metric (MAPE) computed from $\{e_t^{(m)}\}$ are collected in the response vector $F(w)$, and the set S denotes the Pareto-efficient solutions returned by the NBI procedure. All symbols are defined here before being used in the subsequent equations.

TABLE 3. Main mathematical symbols and notation.

Symbol	Meaning
t	Time index (hour).
y_t	Actual load at time t.
$\hat{y}_{i,t}$	Forecast produced by base model $i$at time t.
$r_{i,t}$	Residual of model $i$at time t, $r_{i,t} = y_t - \hat{y}_{i,t}$.
$\hat{y}_t^{\text{ens}}$	Ensemble forecast at time t.
w_i	Mixing weight assigned to base model i.
$\mathbf{w}$	Vector of mixing weights.
z	Number of base models in the ensemble (here $z = 3$).
$\mathbf{f}(\mathbf{w})$	Vector of performance metrics as a function of $\mathbf{w}$.
$\mathbf{F}$	Reduced factors obtained from PCA on the set of metrics.
$\mathcal{P}$	Set of Pareto-efficient solutions returned by NBI.
$H(\mathbf{w})$	Shannon entropy of the weight vector $\mathbf{w}$.
GPE	Global Percentage Error index.
d_s^U	Distance to utopia point for solution $s$in objective space.
Additional indices and parameters	
Symbol	**Meaning**
i	Index of base model (Linear, RBF, MLP).
j	Index of substation / response (or objective).
k	Index of mixture-design weight vector (mixture point).
s	Index of Pareto-optimal solution returned by NBI.
p	Number of mixture-design points in the simplex-lattice.
$\{q, m\}$	Simplex-lattice mixture-design parameters.
Time series and residuals	
Symbol	**Meaning**
S_j	Load time series of substation j, $j = 1, \ldots, 4$.
r_{ij}	Residual series of base model $i$for substation j.
r_{ckj}	Combined residual series for substation $j$and mixture-design point k.
Targets and factors	
Symbol	**Meaning**
T_i	Target value for objective $i$used in the GPE definition.
F_1, F_2	First and second reduced factors obtained from the four substation MAPEs.

To determine the weights for linearly combining the residual series produced by the neural networks (and then forming the ensemble forecast), this study follows the steps illustrated in Figure 4.

The steps of the flowchart are:

a) Data collection of electrical load time series of distribution substations.
b) Selection of z neural networks to be part of the ensemble. Then, the residual series is generated for each of these neural networks, and each time series is considered in step (a).
c) A mixture design of experiments (MDE) [38] is defined to combine the residual series obtained in the

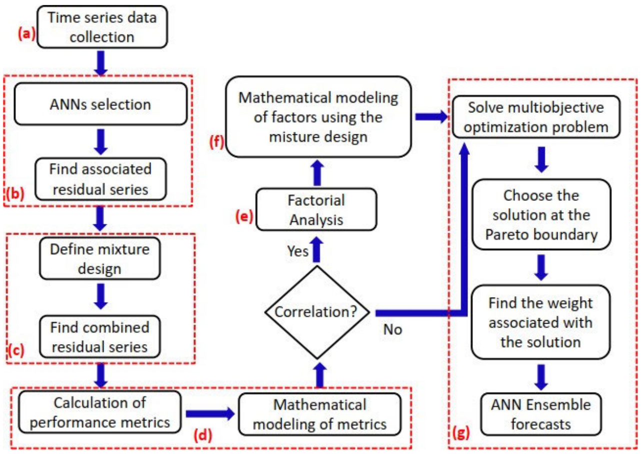


FIGURE 4. Flowchart of the proposed ensemble-optimization methodology. Block diagram of the complete workflow used to construct and optimize the neural-network ensemble, including data preprocessing, feature construction, mixture-design generation, NBI-based multiobjective optimization, and final evaluation.

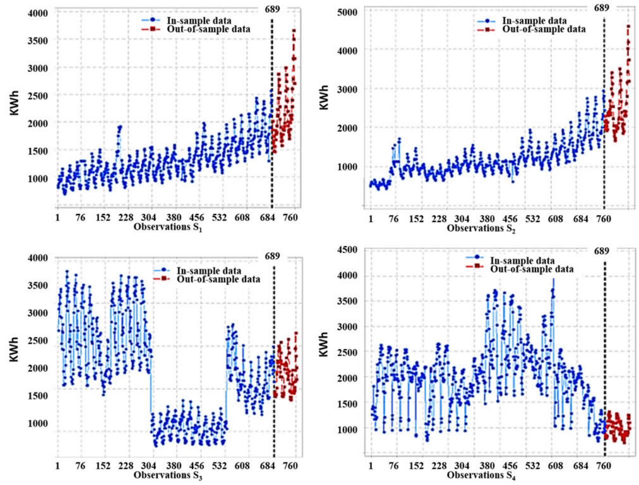


FIGURE 5. Training and validation segments for the substations. Standardized hourly load time series for Substations 1–4 over the ~32-day observation window, indicating the 689-sample training segment and the 72-sample validation segment used in the experiments.

previous step. More specifically, through a lattice simplex design $\{q, m\}$, p sets of weight vectors containing z weights are used to make a linear combination of the residual series obtained in the previous step to obtain the combined residual series.

d) Calculation of forecast performance metrics. In this step, j metrics are calculated for each combined residual series obtained in step (c), obtaining the mathematical models representing these metrics.
e) If there is a correlation between the modeled responses, the factor analysis technique (considering a linear factor analysis model with extraction by principal components and Varimax rotation) will be applied to reduce the dimensionality of the problem as well as the repetition of information, being the number of factors retained determined by the Kaiser's rule [23], [26]. If there is no correlation, we apply step (g).
f) Mathematical modeling of factors using mixture design.
g) Solving the multiobjective optimization problem and selecting the solution that maximizes the entropy of the ensemble-weight vector (Shannon), in ratio with GPE.

According to Figure 4, the methodology for short-term electric load forecast proposed in this study begins with the collection of data from the time series S_1, S_2, S_3, and S_4 of the power distribution substations 1, 2, 3, and 4 shown in Figure 5 (step (a)).

Each time series has ~32 days of hourly measurements (761 points). According to Figure 5, 689 data points were used in training neural networks, and the remainder (72 observations) were used for validation. We split each series chronologically: the first 689 hours for training (model fitting and mixture-design/NBI optimization), and the last 72 hours as a validation segment for performance reporting. No data from the validation period are used for model fitting.

In this study we did not adopt full k-fold cross-validation or rolling-window evaluation because each run requires retraining three neural architectures and executing the mixture-design plus NBI-based ensemble optimization; repeating this entire pipeline over many folds or forecast origins would substantially increase computational cost and complexity for the comparative analysis.

This choice reflects an operational scenario in which STLF models are trained on approximately four weeks of the most recent data and evaluated over a 72-hour horizon matching the longest forecast range considered, thereby capturing several repetitions of daily and weekly load patterns while keeping the experimental design computationally tractable. We acknowledge that this relatively short data span limits the assessment of longer-term generalization and regard extensions with longer training and evaluation windows as an important direction for future work.

The neural networks were trained on the 689-sample training segment using mini-batch stochastic gradient descent. For the MLP, we used a learning rate of 0.01, a mini-batch size of 32, and a total of 292 epochs (100 epochs of backpropagation followed by 192 epochs of conjugate-gradient refinement). During the backpropagation phase, we employed momentum of 0.9, while the conjugate-gradient phase followed the default line-search settings of the training toolbox. The Linear and RBF networks were estimated in closed form via singular-value decomposition (SVD) on the full batch of training data; therefore, iterative hyperparameters such as learning rate and epoch schedule do not apply to these models.

All simulations and analyses were carried out in Python. Neural-network models were implemented using the Keras API with a TensorFlow backend, while data preprocessing, feature construction, and evaluation metrics relied on NumPy, pandas, and scikit-learn, with SciPy used for statistical tests and Matplotlib for visualization.

In step (b), three neural networks were selected to be part of the ensemble: a linear neural network, radial basis function (RBF), and multilayer perceptron (MLP), whose architecture and training algorithms are shown in Table 4.

By combining a Linear Network, a Radial-Basis-Function (RBF) Network and a Multilayer Perceptron (MLP) into a single ensemble, we deliberately exploit the bias–variance trade-off and structural diversity that ensemble theory identifies as prerequisites for error reduction: the Linear model, with high bias and very low variance, captures nearly linear global trends and offers immediate interpretability; the RBF, whose locality and moderate, centre-controlled variance let it respond to sharp events such as feeder switching or public-holiday demand spikes, fills the gaps that global models typically smooth out; and the MLP, with low bias and higher variance, represents complex nonlinear and seasonal interactions thanks to its universal approximation power. Beyond covering different regions of influence (global vs. local), each network is trained with a distinct algorithm (pure gradient descent, clustering + least squares, back-propagation), injecting estimation noise that lowers residual-error correlation—an empirically demonstrated driver of RMSE/MAE improvements in recent load-forecasting studies.

In short-term load-forecasting, where quasi-linear temperature trends, abrupt operational changes and strongly nonlinear consumption patterns coexist, this trio spans the full representational spectrum and delivers more stable, accurate forecasts at negligible extra cost: the Linear model trains in milliseconds, the RBF in seconds and the MLP in minutes. Consequently, the ensemble is grounded both in classic bias–variance theory and in contemporary evidence on heterogeneous ensembles, while its strengths map cleanly onto the operational realities of electric-power systems—a robust rationale for its adoption.

For the MLP, we employed L2 weight decay ($\lambda = 1 \times 10.4$) together with dropout regularization, using a dropout rate of 0.2 on the first hidden layer and 0.1 on the second, to mitigate overfitting on the relatively short training window. For the RBF network, a small ridge term ($\lambda = 1 \times 10.4$) was added to the SVD-based solution of the output weights to stabilize the inversion of the design matrix.

It is worth noting that the ensemble deliberately focuses on three relatively simple yet structurally diverse base models (Linear, MLP, and RBF). This controlled choice keeps the computational burden manageable and allows us to clearly isolate the contribution of the mixture design, PCA-based dimensionality reduction, and NBI-based multi-objective weight optimization, while acknowledging that more complex deep learning architectures are not included in the present study.

To clarify the training configuration, the MLP was optimized using mini-batch stochastic gradient descent with a fixed learning rate of 0.01 and momentum of 0.9 during the backpropagation phase, followed by a conjugate-gradient refinement phase with the same mini-batch size and default line-search parameters. The Linear and RBF networks were trained by solving a regularized least-squares system via SVD on the full training batch, which corresponds to a single-epoch, full-batch optimization without iterative learning-rate updates.

TABLE 4. Architectures and training settings of the base neural networks. Summary of the architectures and training characteristics of the three neural networks used in the ensemble (Linear, RBF, and MLP), including number of layers and neurons, activation functions, training algorithms, and regularization parameters.

	RBF	Linear	MLP
Input layer	144 neurons	144 neurons	144 neurons
Hidden layers	152 neurons	-	9 neurons in hidden layer 1 7 neurons in hidden layer 2
Output layer	4 neurons	4 neurons	4 neurons
Activation function	Gaussian	Linear	Hyperbolic
Training algorithm	Singular Value Decomposition	Singular Value Decomposition	100 epochs of backpropagation followed by 192 epochs of conjugate gradient descent

In this study, NNs are used to analyze time series. Thus, the input is a sample $(y_{t-1}, y_{t-2}, \ldots, y_{t-k})$ of the time series, with a delay of k time units. Since the series has 24-hour seasonality (identified via a Fast Fourier Transform), we set the delay to ($k = 24$) time units. Data referring to the measurement period and day were associated with these measures. To facilitate the detection of important relationships and assist in learning neural network patterns, the time series inputs y_i have been standardized.

The input layer of each neural network has 144 neurons, corresponding to a 24-hour sliding window of substation load and calendar information. For each forecast time, we build a 24×6 input matrix in which each row contains six features: the day of the week, the time-of-day shift (morning, afternoon, evening, night), and the standardized loads of the four substations (S1–S4). This 24×6 matrix is then stacked and vectorized into a 144-dimensional input vector that feeds the networks. The day of the week is encoded as an integer from 1 (Sunday) to 7 (Saturday), and the time-of-day shift is encoded from 1 to 4 for morning, afternoon, evening, and night, respectively.

Each network produced four outputs referring to the forecast of a step forward for each substation and each lagged input $(y_{t-1}, y_{t-2}, \ldots, y_{t-k})$. After training the neural networks, each network produced four residual series $\mathbf{r}_{ij}$ (where i= 1, 2, 3 represents the RBF, Linear, and MLP networks, respectively, and j= 1,…,4 substations, in that order) totaling 12 series of residues with 665 observations. In step (c), a mixture design of experiments (MDE) [38] was chosen to combine the residual series obtained in the previous step. More specifically, using a simplex lattice {3, 5} design (see Figure 6), 25 weight vectors $w_k = [w_{1k}\ w_{2k}\ w_{3k}]$, $k = 1, \ldots, 25$, were defined. For each vector w_k and each residual series obtained in step (b), the combined residual series $\mathbf{rc}_{kj}$ (with 665 observations) were calculated using (15):

$$rc_{kj} = \sum_{t=1}^{3} w_{tk} r_{ij} k= 1, \ldots, 25, j= 1, \ldots, 4. \tag{15}$$

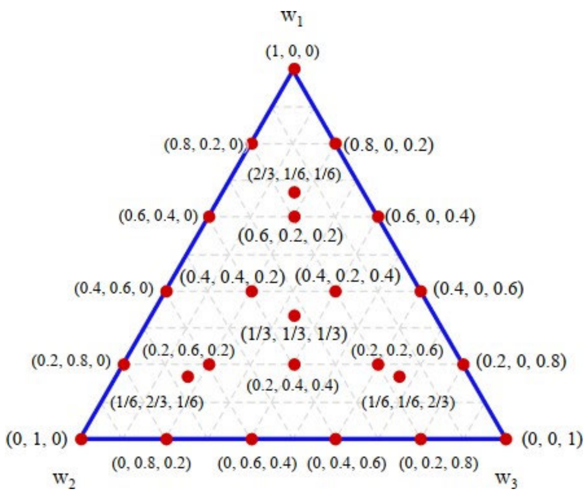


FIGURE 6. Simplex-lattice {3,5} mixture design for ensemble weights. Simplex-lattice {3, 5} design on the unit simplex used to generate candidate combinations of the three ensemble weights (w_1, w_2, w_3)for residual mixing and subsequent multiobjective optimization.

IV. RESULTS AND DISCUSSION

A. ENSEMBLE-WEIGHT OPTIMIZATION

Factor analysis reduced the four sub-station MAPEs to two orthogonal factors F1 and F2. Applying the NBI algorithm to these factors produced a Pareto frontier (Figure 7). The solution that maximizes the Shannon-entropy/GPE ratio $s = 11$ (with s indexing the NBI Pareto solutions $\{z(s)\}_{s=1}^{S}$) which yields the weights w = [0.333, 0.412, 0.255] to the RBF, Linear and MLP residuals respectively (first row of Table 5). Here, 'Entropy' refers to H(**w**) computed on the ensemble mixing weights (not on objective-function weights). This balanced combination is used in all subsequent tests.

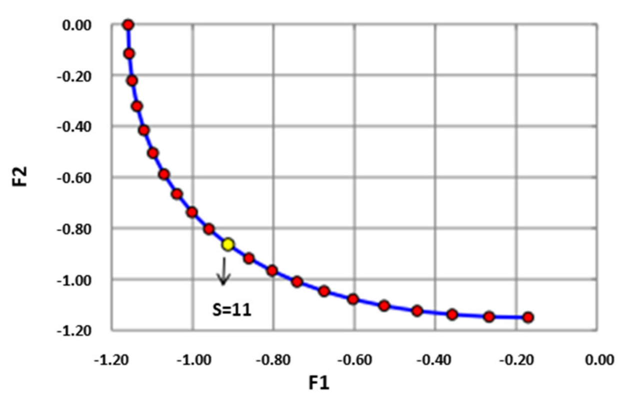


FIGURE 7. Pareto frontier in the reduced objective space. Pareto frontier illustrating the trade-off between the two reduced factors F_1 and F_2 obtained from PCA of the multi-metric responses, for the ensemble-weight optimization problem; each point corresponds to one weight combination from the mixture design.

TABLE 5. Optimization results along the Pareto frontier. Normal-boundary-intersection (NBI) optimization results for the selected Pareto-efficient solutions, reporting the ensemble weight vector (w_1, w_2, w_3), the corresponding values of the reduced factors F_1 and F_2, and the associated entropy and global percentage error (GPE) indices for each solution.

s	β_i	$1-\beta_i$	$\overline{F_1}$	$\overline{F_2}$	F_1	F_2
1	0.000	1.000	1.000	0.000	-0.170	-1.150
2	0.050	0.950	0.903	0.003	-0.266	-1.147
3	0.100	0.900	0.810	0.010	-0.358	-1.139
4	0.150	0.850	0.722	0.022	-0.445	-1.124
5	0.200	0.800	0.640	0.040	-0.526	-1.104
6	0.250	0.750	0.562	0.062	-0.603	-1.079
7	0.300	0.700	0.490	0.090	-0.675	-1.047
8	0.350	0.650	0.422	0.122	-0.742	-1.010
9	0.400	0.600	0.360	0.160	-0.804	-0.967
10	0.450	0.550	0.302	0.202	-0.860	-0.918
11	0.500	0.500	0.250	0.250	-0.912	-0.864
12	0.550	0.450	0.202	0.302	-0.959	-0.803
13	0.600	0.400	0.160	0.360	-1.001	-0.737
14	0.650	0.350	0.122	0.422	-1.039	-0.666
15	0.700	0.300	0.090	0.490	-1.071	-0.588
16	0.750	0.250	0.062	0.562	-1.098	-0.505
17	0.800	0.200	0.040	0.640	-1.120	-0.416
18	0.850	0.150	0.022	0.722	-1.137	-0.321
19	0.900	0.100	0.010	0.810	-1.150	-0.221
20	0.950	0.050	0.003	0.903	-1.157	-0.115
21	1.000	0.000	0.000	1.000	-1.159	-0.003

s	w_1	w_2	w_3	H	GPE	$\xi=H/GPE$
1	0.309	0.223	0.469	0.000	0.854	0.001
2	0.312	0.241	0.447	0.086	0.773	0.112
3	0.315	0.260	0.425	0.141	0.701	0.201
4	0.317	0.279	0.404	0.184	0.639	0.287
5	0.320	0.298	0.383	0.217	0.586	0.371
6	0.322	0.317	0.361	0.244	0.542	0.451
7	0.325	0.336	0.340	0.265	0.508	0.523
8	0.327	0.355	0.319	0.281	0.482	0.583
9	0.329	0.374	0.297	0.292	0.466	0.627
10	0.331	0.393	0.276	0.299	0.460	0.650
11	0.333	0.412	0.255	0.301	0.462	0.652
12	0.334	0.432	0.234	0.299	0.474	0.631
13	0.336	0.451	0.213	0.292	0.495	0.591
14	0.337	0.471	0.192	0.281	0.525	0.535
15	0.338	0.490	0.172	0.265	0.565	0.470
16	0.339	0.510	0.151	0.244	0.614	0.398
17	0.339	0.530	0.131	0.217	0.672	0.323
18	0.339	0.550	0.111	0.184	0.740	0.248
19	0.339	0.570	0.091	0.141	0.817	0.173
20	0.339	0.590	0.071	0.086	0.902	0.096
21	0.338	0.611	0.051	0.000	0.998	0.000

Figure 8 shows various weight combinations w_1, w_2, and w_3 within the doable region ($w_1 + w_2 + w_3 = 1$, $w_i \leq 0$, and $w_i \geq 1$) and their respective values for F_1 and F_2. These points are dominated by Pareto-optimal solutions, which determine the boundary.

We can see from the overlaid contour graph on Figure 9 that the chosen solution (s= 11, highlighted in red) belongs to the viable region and has the shortest distance from the optimum individual functions F_1 and F_2 (highlighted in black dots). The graph also shows how the solutions behave for different values of β in the optimization problem as in (11). The more a function is prioritized, the closer the solution is to the optimum individual.

Using the criterion of the maximum ratio ζ, the best solution among the solutions was $s = 11$. To assess the stability of this choice, we performed a sensitivity analysis of the

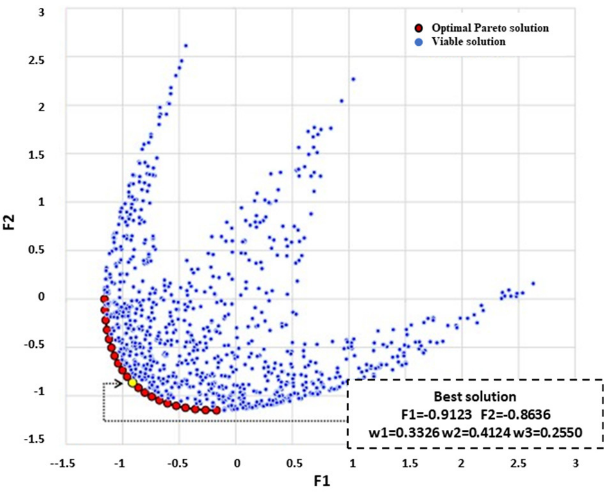


FIGURE 8. Viable weight combinations and Pareto-optimal boundary. Feasible weight solutions in barycentric coordinates (w_1, w_2, w_3) mapped over the mixture-design space, together with the Pareto-optimal boundary that delineates the set of nondominated ensemble configurations.

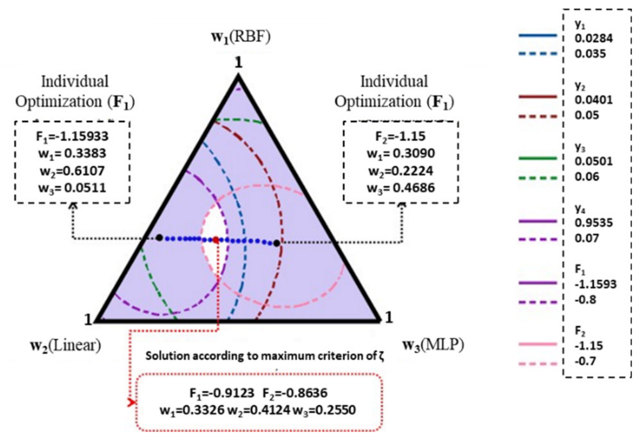


FIGURE 9. Contour plot and selected ensemble solution in the mixture space. Overlaid contour plot of the reduced factors F_1 and F_2 on the simplex-lattice design, highlighting the chosen ensemble solution ($s = 11$) relative to the region of best compromise between the two objectives.

entropy/GPE-based selection rule. Starting from the Pareto set reported in Table 5 and the contour representation in Figure 9, we perturbed the exponents and weights that control the relative importance assigned to entropy and GPE over a wide, yet reasonable, range and recomputed the resulting scalarization for all Pareto solutions. In the vast majority of configurations, the preferred solution remained $s = 11$; in the few cases where nearby solutions were selected, the corresponding weight vectors (w_1, w_2, w_3)and performance metrics were very close to those of $s = 11$. These findings indicate that the ensemble-weight vector associated with the chosen operating point is only weakly affected by moderate changes in the scalarization parameters, confirming the robustness of the weight-selection procedure. However, all other points on the Pareto Frontier also represent Pareto-optimal solutions. The vector of viable solutions ($s = 11$) is Pareto optimal (Figure 10), since no viable point can reduce any objective function without causing a simultaneous increase of at least some objective function. Note that it is impossible to reduce the value of y_1 (and/or y_2) without increasing the value of y_3 (and/or y_4).

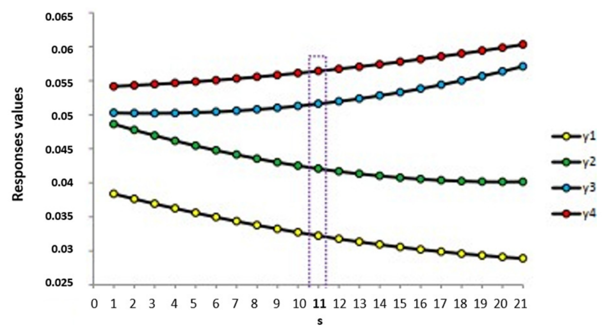


FIGURE 10. Behavior of original responses along the Pareto frontier. Variation of the original performance responses y_1–y_4 (MAPE) across successive Pareto solutions, showing how improvements in one metric impact the others along the frontier.

B. VALIDATION PERFORMANCE

Table 6 compares validation-set performance of the ensemble with its three constituent networks for 24-, 48- and 72-step horizons. Across all substations the ensemble lowers MAPE by 15 – 25 % relative to the best single model (e.g. Substation 1, 24-step: 0.027 vs 0.033; Substation 4, 48-step: 0.079 vs 0.085). Similar gains are observed for RMSE and MAE.

To statistically assess the performance gains of the optimized ensemble, we conducted paired t-tests comparing the ensemble errors with those of the individual RBF, Linear, and MLP models over the 12 evaluation cases. In all comparisons, the ensemble achieved a lower average error (e.g., mean MAPE of 0.0501 versus 0.0645, 0.0707, and 0.0650 for the RBF, Linear, and MLP models, respectively). The paired t-tests (one-sided, testing whether the ensemble error is smaller) yielded p-values of 4.2×10.6, 3.7×10^{-3}, and 7.3×10.4 for RBF, Linear, and MLP, respectively, indicating statistically significant improvements at the 5% level. These results were confirmed by Wilcoxon signed-rank tests, which also favored the ensemble in all cases. The corresponding paired Cohen's d values (0.95–2.25) further indicate large to very large effect sizes, underscoring the practical relevance of the observed performance gains.

From a methodological standpoint, the proposed framework is complementary to contemporary ensemble-learning techniques such as boosting (e.g., XGBoost), bagging, stacking, and Bayesian model averaging. While these approaches mainly focus on how base learners are generated or combined, our contribution lies in the systematic tuning of ensemble weights via mixture design and NBI-based multi-objective optimization. Future work will explicitly combine these strategies with the present framework to provide a more direct empirical comparison and to quantify the added value of each component.

TABLE 6. Validation-set forecasting performance (MAPE). Mean absolute percentage error (MAPE) on the 72-hour validation segment for Substations 1–4 at 24-, 48-, and 72-step horizons, comparing the three individual neural networks (Linear, RBF, MLP) and the proposed ensemble; lower values indicate better accuracy.

	Substation 1			Substation 2		
Steps ahead	24	48	72	24	48	72
RBF	0,0426	0,0439	0,0390	0,0820	0,0791	0,0655
Linear	0,0329	0,0305	0,0299	0,0640	0,0572	0,0519
MLP	0,0435	0,0416	0,0357	0,1078	0,0805	0,0682
Ensemble	0,0270	0,0255	0,0228	0,0618	0,0560	0,0472
	Substation 3			Substation 4		
Steps ahead	24	48	72	24	48	72
RBF	0,0457	0,0477	0,0558	0,0761	0,0925	0,1046
Linear	0,0542	0,0601	0,0663	0,1376	0,1332	0,1310
MLP	0,0425	0,0446	0,0544	0,0753	0,0853	0,1005
Ensemble	0,0399	0,0416	0,0398	0,0739	0,0791	0,0864

C. FORECAST TRACE ILLUSTRATION

Figure 11 juxtaposes the ensemble prediction (red) against the actual load (blue) for Substation 1. The model captures both the diurnal envelope and short-lived spikes with minimal phase lag, illustrating its operational usefulness for day-ahead scheduling.

D. PRACTICAL IMPLICATIONS

The optimized weights exploit complementary strengths: the Linear network anchors the global trend, the RBF model sharpens local corrections, and the MLP absorbs residual nonlinearities. Because the weight vector sits near the center of the Pareto set, small deviations in the training data are unlikely to substantially degrade accuracy—an asset for real-time deployment.

E. SOLUTION DEPLOYMENT

From a deployment perspective, the proposed ensemble can be integrated as an additional forecasting module within a distribution utility's existing STLF workflow. In an operational setting, the models are trained or updated off-line using recent historical load data and the corresponding calendar/holiday features, whereas the on-line stage only requires feeding the most recent load measurements and time-related covariates to the three base networks and applying the precomputed ensemble weights. This separation between off-line optimization and lightweight on-line inference keeps the computational cost compatible with near real-time operation on standard utility-grade servers.

The required input data streams—hourly substation load, basic calendar information (day, hour, weekday/weekend, holidays), and any optional exogenous variables if available—are routinely collected in modern SCADA/EMS infrastructures. The ensemble forecasts can be exported as time-stamped load profiles to existing decision-support tools for scheduling and dispatch, with integration implemented via standard database queries, file-based interfaces, or web APIs. In this way, the proposed ensemble outputs can be interfaced with existing SCADA/EMS environments without changes to downstream applications, which simply consume the forecast trajectories produced by the model.

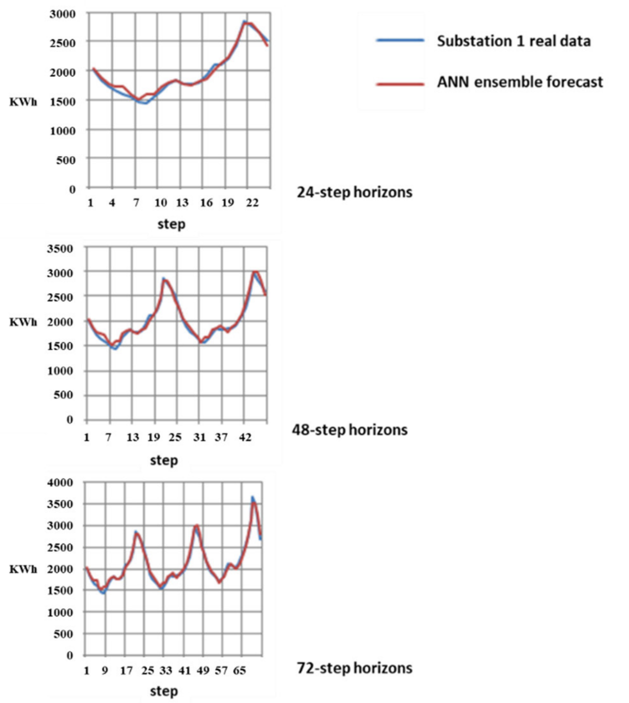


FIGURE 11. Forecasting results for Substation 1 with uncertainty information. Comparison between actual load and ensemble forecasts for Substation 1 at 24-, 48-, and 72-step horizons; similar behavior was observed for Substations 2–4.

V. CONCLUSION

This study proposed a multivariate optimization framework for short-term load forecasting at the distribution-substation level, based on a neural-network ensemble whose weights are tuned through mixture design of experiments, factor analysis with principal-component extraction, and NBI multi-objective optimization. By explicitly modeling the ensemble-weight vector on the probability simplex and operating on orthogonalized error factors, the methodology provides a structured alternative to heuristic or single-metric ensemble-weight selection.

Applied to hourly data from four Brazilian distribution substations over 24-, 48-, and 72-step horizons, the optimized ensemble consistently outperformed its constituent Linear, RBF, and MLP networks. Reductions in MAPE of up to 20% relative to the best individual model were observed, with similar gains for RMSE and MAE. Paired and nonparametric tests indicated that these improvements are statistically meaningful. From an operational standpoint, the framework is compatible with standard SCADA/EMS infrastructures: the computationally intensive weight-optimization and model-training steps are performed off-line, whereas on-line deployment requires only lightweight inference of three base networks followed by a fixed linear combination.

The work also contributes methodologically by integrating mixture-design concepts, factor-analysis/PCA-based dimensionality reduction, and entropy/GPE-based scalarization along the Pareto frontier into a coherent ensemble-construction pipeline. This perspective complements conventional ensemble approaches (bagging, boosting, stacking, Bayesian averaging), which typically optimize a single aggregate loss or rely on ad hoc voting schemes, by highlighting how multi-criteria trade-offs and weight diversification can be handled in a principled way.

At the same time, several limitations suggest avenues for future research. The empirical study considered a limited number of substations and a relatively short training window, and focused on three classical neural-network architectures without performing a full ablation study of the mixture-design, PCA, and NBI components. The primary evaluation metric was MAPE, and comparisons with more advanced ensemble and deep-learning schemes (e.g., LSTMs, CNNs, Transformer-based models, stacking or boosting frameworks) were not exhaustively explored. Extending the analysis to larger and more heterogeneous datasets, richer model families, broader sets of error and probabilistic metrics, and systematic ablation/benchmarking studies would provide a more detailed assessment of the individual and joint contributions of each block in the proposed pipeline.

Overall, the multivariate optimization-based ensemble presented here offers a practical and scalable tool for improving distribution-level STLF. Its reliance on standard neural-network toolchains, modest run-time requirements, and clear integration path with existing SCADA/EMS environments make it a strong candidate for real-world deployment. Future work will focus on larger-scale case studies, automated retraining and monitoring procedures, and tighter coupling with probabilistic and risk-aware decision-support tools in modern smart-grid operations.

DATA AND CODE AVAILABILITY

The data and code that support the findings of this study are openly available in the institutional repository of Federal University at Itajubá at EnsembleSTLF.

REFERENCES

[1] S. Karamolegkos and D. E. Koulouriotis, "Advancing short-term load forecasting with decomposed Fourier ARIMA: A case study on the Greek energy market," *Energy*, vol. 325, Jun. 2025, Art. no. 135854, doi: 10.1016/j.energy.2025.135854.

[2] N. Giamarelos and E. N. Zois, "Boosting short term electric load forecasting of high & medium voltage substations with visibility graphs and graph neural networks," *Sustain. Energy, Grids Netw.*, vol. 38, Jun. 2024, Art. no. 101304, doi: 10.1016/j.segan.2024.101304.

[3] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 303–315, Feb. 2014, doi: 10.1109/TNNLS.2013.2276053.

[4] A. Jahani, K. Zare, and L. Mohammad Khanli, "Short-term load forecasting for microgrid energy management system using hybrid SPM-LSTM," *Sustain. Cities Soc.*, vol. 98, Nov. 2023, Art. no. 104775, doi: 10.1016/j.scs.2023.104775.

[5] R. Ž. Jovanović, A. A. Sretenović, and B. D. Živković, "Ensemble of various neural networks for prediction of heating energy consumption," *Energy Buildings*, vol. 94, pp. 189–199, May 2015, doi: 10.1016/j.enbuild.2015.02.052.

[6] A. Graefe, J. S. Armstrong, R. J. Jones, and A. G. Cuzán, "Combining forecasts: An application to elections," *Int. J. Forecasting*, vol. 30, no. 1, pp. 43–54, Jan. 2014, doi: 10.1016/j.ijforecast.2013.02.005.

[7] L. Fausett, *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, 1994.

[8] S. C. Bhattacharyya and L. T. Thanh, "Short-term electric load forecasting using an artificial neural network: Case of northern Vietnam," *Int. J. Energy Res.*, vol. 28, no. 5, pp. 463–472, Apr. 2004, doi: 10.1002/er.980.

[9] F. Cavallaro, "Electric load analysis using an artificial neural network," *Int. J. Energy Res.*, vol. 29, no. 5, pp. 377–392, 2005, doi: 10.1002/er.1054.

[10] K. Muralitharan, R. Sakthivel, and R. Vishnuvarthan, "Neural network based optimization approach for energy demand prediction in smart grid," *Neurocomputing*, vol. 273, pp. 199–208, Jan. 2018, doi: 10.1016/j.neucom.2017.08.017.

[11] S. Voronin and J. Partanen, "Forecasting electricity price and demand using a hybrid approach based on wavelet transform, ARIMA and neural networks," *Int. J. Energy Res.*, vol. 38, no. 5, pp. 626–637, Apr. 2014, doi: 10.1002/er.3067.

[12] M. Ghayekhloo, M. B. Menhaj, and M. Ghofrani, "A hybrid short-term load forecasting with a new data preprocessing framework," *Electric Power Syst. Res.*, vol. 119, pp. 138–148, Feb. 2015, doi: 10.1016/j.epsr.2014.09.002.

[13] R. Kumar, S. Aeri, and D. Baleanu, "Artificial neural networks for the wavelet analysis of lane-emden equations: Exploration of astrophysical enigma," *Int. J. Model. Simul.*, vol. 45, no. 5, pp. 1–12, Sep. 2025, doi: 10.1080/02286203.2023.2301126.

[14] M. Z. Yousaf, S. Khalid, M. F. Tahir, A. Tzes, and A. Raza, "A novel DC fault protection scheme based on intelligent network for meshed DC grids," *Int. J. Electr. Power Energy Syst.*, vol. 154, Dec. 2023, Art. no. 109423, doi: 10.1016/j.ijepes.2023.109423.

[15] M. Z. Yousaf, H. Liu, A. Raza, and A. Mustafa, "Deep learning-based robust DC fault protection scheme for meshed HVDC grids," *CSEE J. Power Energy Syst.*, pp. 2423–2434, 2023, doi: 10.17775/cseejpes.2021.03550.

[16] M. Z. Yousaf, M. F. Tahir, A. Raza, M. A. Khan, and F. Badshah, "Intelligent sensors for DC fault location scheme based on optimized intelligent architecture for HVdc systems," *Sensors*, vol. 22, no. 24, p. 9936, Dec. 2022, doi: 10.3390/s22249936.

[17] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990, doi: 10.1109/34.58871.

[18] G. R. Harikrishnan and S. Sreedharan, "Advanced short-term load forecasting for residential demand response: An XGBoost-ANN ensemble approach," *Electric Power Syst. Res.*, vol. 242, May 2025, Art. no. 111476, doi: 10.1016/j.epsr.2025.111476.

[19] M. Saviozzi, S. Massucco, and F. Silvestro, "Implementation of advanced functionalities for distribution management systems: Load forecasting and modeling through artificial neural networks ensembles," *Electric Power Syst. Res.*, vol. 167, pp. 230–239, Feb. 2019, doi: 10.1016/j.epsr.2018.10.036.

[20] G. Cervone, L. Clemente-Harding, S. Alessandrini, and L. Delle Monache, "Short-term photovoltaic power forecasting using artificial neural networks and an analog ensemble," *Renew. Energy*, vol. 108, pp. 274–286, Aug. 2017, doi: 10.1016/j.renene.2017.02.052.

[21] J. G. Jetcheva, M. Majidpour, and W.-P. Chen, "Neural network model ensembles for building-level electricity load forecasts," *Energy Buildings*, vol. 84, pp. 214–223, Dec. 2014, doi: 10.1016/j.enbuild.2014.08.004.

[22] R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed., Upper Saddle River, NJ, USA: Prentice-Hall, 2007.

[23] F. A. D. Almeida, G. F. Gomes, J. H. D. Gaudêncio, J. H. D. F. Gomes, and A. P. D. Paiva, "A new multivariate approach based on weighted factor scores and confidence ellipses to precision evaluation of textured fiber bobbins measurement system," *Precis. Eng.*, vol. 60, pp. 520–534, Nov. 2019, doi: 10.1016/j.precisioneng.2019.09.010.

[24] A. C. Rencher, *Methods of Multivariate Analysis*. Hoboken, NJ, USA: Wiley, 2002, doi: 10.1002/0471271357.

[25] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., New York, NY, USA: Springer, 2002, doi: 10.1007/b98835.

[26] H. F. Kaiser, "The application of electronic computers to factor analysis," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 141–151, Apr. 1960, doi: 10.1177/001316446002000116.

[27] J. H. de Freitas Gomes, A. R. S. Júnior, A. P. de Paiva, J. R. Ferreira, S. C. da Costa, and P. P. Balestrassi, "Global criterion method based on principal components to the optimization of manufacturing processes with multiple responses," *J. Mech. Eng.*, vol. 58, no. 5, pp. 345–353, May 2012, doi: 10.5545/sv-jme.2011.136.

[28] I. Das and J. E. Dennis, "Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems," *SIAM J. Optim.*, vol. 8, no. 3, pp. 631–657, Aug. 1998, doi: 10.1137/s1052623496307510.

[29] N. Bratchell, "Multivariate response surface modelling by principal components analysis," *J. Chemometrics*, vol. 3, no. 4, pp. 579–588, Sep. 1989, doi: 10.1002/cem.1180030406.

[30] Y. Lim, Y. Park, and H.-S. Oh, "Robust principal component analysis via ES-algorithm," *J. Korean Stat. Soc.*, vol. 43, no. 1, pp. 149–159, Mar. 2014, doi: 10.1016/j.jkss.2013.07.002.

[31] V. Vahidinasab and S. Jadid, "Normal boundary intersection method for suppliers' strategic bidding in electricity markets: An environmental/economic approach," *Energy Convers. Manage.*, vol. 51, no. 6, pp. 1111–1119, Jun. 2010, doi: 10.1016/j.enconman.2009.12.019.

[32] A. L. G. Pires, P. Rotella Junior, L. C. S. Rocha, R. S. Peruchi, K. Janda, and R. D. C. Miranda, "Environmental and financial multi-objective optimization: Hybrid wind-photovoltaic generation with battery energy storage systems," *J. Energy Storage*, vol. 66, Aug. 2023, Art. no. 107425, doi: 10.1016/j.est.2023.107425.

[33] S. C. Streitenberger, E. L. Romão, A. P. Paiva, P. P. Balestrassi, J. H. G. Freitas, and V. C. Paes, "Normal boundary intersection with factor analysis approach for multiobjective stochastic optimization of a cladding process focusing on reduction of energy consumption and rework," *J. Cleaner Prod.*, vol. 333, Jan. 2022, Art. no. 129915, doi: 10.1016/j.jclepro.2021.129915.

[34] Z. Jia and M. G. Ierapetritou, "Generate Pareto optimal solutions of scheduling problems using normal boundary intersection technique," *Comput. Chem. Eng.*, vol. 31, no. 4, pp. 268–280, Feb. 2007, doi: 10.1016/j.compchemeng.2006.07.001.

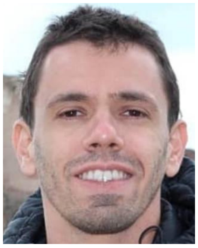
[35] H. A. Taboada, F. Baheranwala, D. W. Coit, and N. Wattanapongsakorn, "Practical solutions for multi-objective optimization: An application to system reliability design problems," *Rel. Eng. Syst. Saf.*, vol. 92, no. 3, pp. 314–322, Mar. 2007, doi: 10.1016/j.ress.2006.04.014.

[36] L. C. S. Rocha, A. P. D. Paiva, P. P. Balestrassi, G. Severino, and P. Rotela Junior, "Entropy-based weighting for multiobjective optimization: An application on vertical turning," *Math. Problems Eng.*, vol. 2015, pp. 1–11, Jan. 2015, doi: 10.1155/2015/608325.

[37] L. C. S. Rocha, P. Rotela Junior, G. Aquila, A. P. de Paiva, and P. P. Balestrassi, "Toward a robust optimal point selection: A multiple-criteria decision-making process applied to multi-objective optimization using response surface methodology," *Eng. Comput.*, vol. 37, no. 4, pp. 2735–2761, Oct. 2021, doi: 10.1007/s00366-020-00973-5.

[38] J. Cornell, *Experiments With Mixtures: Designs, Models, and the Analysis of Mixture Data*, 3rd ed., Hoboken, NJ, USA: Wiley, 2002, doi: 10.1002/9781118204221.

FRANCO B. ROCHA received the master's degree in physics and applied mathematics and the Ph.D. degree in electrical engineering sciences from the Federal University of Itajubá, Brazil. He is currently a Faculty Member with the Federal University of Alfenas, Brazil.



PAULO ROTELLA JUNIOR received the Ph.D. degree in production engineering from UNIFEI. He is currently an Adjunct Professor with the Production Engineering Department, Federal University of Paraíba (UFPB), Brazil. His research interests include engineering economics, renewable-energy investment, managerial accounting, finance, and optimization.

PEDRO PAULO BALESTRASSI received the M.Sc. degree in electrical engineering from the Federal University of Itajubá (UNIFEI), Brazil, and the Ph.D. degree in industrial engineering from the Federal University of Santa Catarina. He is currently a Full Professor with UNIFEI. His research interests include applied statistics, quality engineering, and neural networks.

FARID MELGANI (Fellow, IEEE) received the Ph.D. degree in electronic and computer engineering from the University of Genoa, Italy. He is currently a Full Professor with the University of Trento, Italy, where he leads the Signal Processing and Recognition Laboratory. His research interests include machine learning, remote sensing, and image/signal processing.

ANTONIO C. ZAMBRONI DE SOUZA (Senior Member, IEEE) received the Ph.D. degree from the University of Waterloo, Canada. He is currently a Full Professor of electrical engineering with the Federal University of Itajubá, Brazil. His work centers on power-system analysis and optimization. He is a fellow of the Institution of Engineering and Technology (IET).