IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 June 2023, accepted 31 July 2023, date of publication 7 August 2023, date of current version 15 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3302701

RESEARCH ARTICLE

A New Co-Optimized Hybrid Model Based on Multi-Objective Optimization for Probabilistic Wind Power Forecasting in a Spatio–Temporal Framework

MARKOS A. KOUSOUNADIS-KNOUSEN[1], IOANNIS K. BAZIONIS[1], DIMITRIOS SOUDRIS[1], (Member, IEEE), FRANCKY CATTHOOR[2,3], AND PAVLOS S. GEORGILAKIS[1], (Senior Member, IEEE)

[1]School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), 157 80 Athens, Greece
[2]IMEC, 3001 Leuven, Belgium
[3]Associated Division ESAT-INSYS (INSYS), Integrated Systems, KU Leuven, Heverlee, 3001 Leuven, Belgium

Corresponding author: Pavlos S. Georgilakis (pgeorg@power.ece.ntua.gr)

ABSTRACT Wind power generation is characterized by high intermittency and volatility owing to the stochastic nature of wind. In addition to forecasting accuracy, forecasting uncertainty quantification can have a major impact on power system energy management and operations planning. In this paper, a new fully co-optimized hybrid short-term probabilistic Wind Power Forecasting (WPF) model is proposed for the construction of Prediction Intervals (PIs) in a spatiotemporal framework. A Multi-Objective Improved Adaptive Particle Swarm Optimization (MOIAPSO) algorithm is developed to optimize the model's parameters. PIs are generated by nonlinear autoregressive networks with exogenous inputs (NARX) using the Lower Upper Bound Estimation (LUBE) method. Unlike previous related work, the components of the proposed hybrid NARX-LUBE-MOIAPSO model, as well as the initial settings and the configuration of the parameters, are determined based on a full co-optimization approach. The co-optimization is performed from a forecasting quality and training time trade-off perspective. Furthermore, a spatiotemporal framework is introduced to improve forecasting performance and comprehension of regional spatiotemporal uncertainty dynamics. The spatiotemporal framework comprises a novel conditional spatiotemporal forecasting methodology and the modeling of spatiotemporal dependencies based on a binary Probabilistic Forecasting Error (PFE) metric. The proposed model and spatiotemporal framework are tested on publicly available datasets consisting of turbine-specific measurements and generate accurate forecasts with efficient uncertainty quantification, while maintaining computational complexity at relatively low levels compared to other state-of-the-art hybrid probabilistic WPF models.

INDEX TERMS Co-optimization, improved adaptive particle swarm optimization (IAPSO), multi-objective optimization, prediction intervals (PIs), spatio–temporal, wind power forecasting (WPF).

NOMENCLATURE

Acronyms & Abbreviations

AI	Artificial intelligence.
ACF	Autocorrelation function.
CCELM	Chance constrained extreme learning machine.
CDF	Cumulative distribution function.
CRPS	Continuous ranked probability score.
CWC	Coverage width-based criterion.
ELM	Extreme learning machine.
GRU	Gated recurrent unit.
IAPSO	Improved adaptive particle swarm optimization.

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

KDE	Kernel density estimation.
LSTM	Long short-term memory.
LUBE	Lower upper bound estimation.
MLP	Multi-layer perceptron.
MOIAPSO	Multi-objective improved adaptive particle swarm optimization.
NARX	Nonlinear autoregressive network with exogenous inputs.
NSGA-II	Non-dominated sorting genetic algorithm II.
PFE	Probabilistic forecasting error.
PI	Prediction interval.
PICP	Prediction Interval Coverage Probability.
PINRW	Prediction Interval Normalized Root-mean-squared Width.
PL	Pinball Loss.
PSO	Particle swarm optimization.
VMD	Variational mode decomposition.
RES	Renewable energy sources.
WPF	Wind power forecasting.
WPP	Wind power plant.
WT	Wind turbine.

Sets & indices

P	Set of WPP turbines, indexed by g.
S	Swarm of particles, indexed by p.
T	Training set, indexed by t.
U	Set of NARX weight connections and biases, indexed by j.
g	Index of WTs.
j	Index of NARX weight connections and biases.
p	Index of particles.
t	Index of time intervals.

Parameters

I	Total iterations.
N_{WT}^1, N_{WT}^2	Total number of WT in case studies I and II, respectively.
N_t	Test set size.
N_{tr}	Training set size.
R	Range of actual output values.
X_{kt}	Exogenous input $k$at time t.
c_1^i, c_2^i	Acceleration coefficients of PSO.
cl	PIs nominal confidence level ϵ [0, 1].
h	Penalty parameter for CWC calculation.
i	Iteration number.
lag_1, lag_2	Amount of lagged time intervals for output feedback to NARX input.
m^i	Mutation rate (possibility) at iteration i.
n	Number of exogenous inputs.
n_{dim}	Number of search space dimensions.
r_1, r_2	Random numbers ϵ [0, 1].
w_{in}^i	Inertia weight at iteration i.
y_t	Actual wind power output at time t.

Variables

Z_g	Z-score of forecasting results related to WT g.
c_t	Binary PFE metric indicating whether y_t lies inside its corresponding PI.
d_t	Error metric used for the calculation of PL.
l_t/u_t	Lower/upper bound of PI at time t.
v_{jp}^i	Velocity of particle p corresponding to weight or bias j at iteration i.
w_{jp}^i	Weight or bias j of NARX, for particle p at iteration i.
$w_{j,swarm}^{i,best}$	Weight or bias j of NARX, of best position encountered in swarm up to iteration i.
$w_{j,p}^{i,best}$	Weight or bias j of NARX, of best position encountered by particle p up to iteration i.
z_t	Average of lower and upper bound of PI at time t.

I. INTRODUCTION

In an attempt to limit the consequences of global climate change and fossil fuel depletion, research focus has shifted towards RES over the past few years [1]. Wind power is one of the most popular RES; hence the contribution of wind power towards meeting global energy needs is increasing [2]. However, wind power generation is highly volatile and intermittent owing to the stochastic nature of wind fields. Increasing integration of wind power into electric power systems results in multiple challenges regarding grid stability and energy management operations [3]. Such operations range from long-term system upgrades and day-ahead reserve planning to real-time operation of smart microgrids, where intra-hour forecasting is essential to balance energy production, storage, and local loads. In any case, accurately forecasting wind power generation is of vital importance.

WPF models are based on physical, statistical, or hybrid approaches [3]. Physical approaches are suitable for long-term forecasting and involve computationally expensive models that rely heavily on expertise and high-quality data. On the other hand, statistical approaches are intended for short-term forecasting and use time series methods or machine learning. In recent years, machine learning methods have dominated research in WPF mainly because of their ability to identify complex, nonlinear relationships between data [4].

WPF usually relies on forecasted weather data which causes additional errors in addition to the already highly volatile WT operation. Forecast uncertainty quantification can have a major impact on energy management and operations planning through stochastic power flow analysis, reserve planning optimization, unit commitment, and economic dispatch [5]. Deterministic forecasting does not provide information about forecast uncertainty; thus, research focusing on probabilistic forecasting has increased [3]. Probabilistic forecasting includes parametric and nonparametric methods. Owing to the difficulty of accurately assuming predictive distributions of wind power, nonparametric

methods have attracted wide attention, even though they can be computationally demanding. Typical nonparametric methods include quantile regression, kernel density estimation, bootstrapping, LUBE, and ensemble methods [3], [6]. Multiple probabilistic WPF models have been proposed over the years for PIs generation, following different approaches to address the challenges derived from the volatile, complex nature of wind. Several studies have suggested probabilistic WPF models based on metaheuristic algorithms, to improve the identification of nonlinear relationships, often at the expense of computational complexity, convergence maturity, or even convergence assurance [7], [8]. In recent years, many researchers have shifted their focus to hybrid probabilistic WPF methods. In hybrid methods, several components are put together in an effort to mitigate each other's disadvantages [3]. Various hybrid methods have been suggested, that combine metaheuristics, deep learning, multi-objective optimization, and data preprocessing techniques. Further information regarding state-of-the-art hybrid probabilistic WPF models is provided in Section II.

Co-optimization of the components of a hybrid model can have a beneficial impact on forecasting accuracy [9]. Determining the type of components and their corresponding parameters with respect to each other's characteristics can further mitigate their respective disadvantages. However, few studies have focused on the co-optimization of the individual components of hybrid probabilistic WPF models [10], [11], [12], [13], [14], [15], [16], [17], [18]. Optimal probabilistic forecasting performance entails global searching, suitable model structures, well-balanced probabilistic forecasts (multi-objective optimization), and data interdependencies extraction. None of [10], [11], [12], [13], [14], [15], [16], and [17] have fully addressed and co-optimized the aforementioned probabilistic forecasting aspects. On the other hand, a full co-optimization is performed in [18]; however, completely disregarding computational complexity. Increased training time is often undesirable, particularly in online training and short-term forecasting horizons. Moreover, these studies ignore spatial wind patterns and hence do not take advantage of spatiotemporal dependencies among neighboring units. Forecasting performance is often improved in a spatiotemporal forecasting framework, while optimal forecasts generated by an efficient probabilistic WPF model facilitate interdependence structure modeling and enhance the comprehension of regional spatiotemporal dynamics.

This paper proposes a fully co-optimized hybrid WPF model that generates short-term probabilistic forecasts in a spatiotemporal framework. The proposed model introduces a MOIAPSO algorithm to enhance global search while maintaining a high convergence rate and training speed. All the parameters in IAPSO, i.e., the inertia weight and the acceleration coefficients, are fully adaptive. MOIAPSO is further improved by incorporating an adaptive grid, binary tournament selection, and Nguyen-Widrow initialization. NARX networks are used to capture time series dependencies in data while maintaining low computational complexity. Probabilistic forecasts are issued in the form of PIs using the LUBE method [19]. The choice of each individual component and the determination of its parameters are based on full co-optimization. In the full co-optimization approach, all hybrid components and their parameters are simultaneously tuned such that they complement each other, and individual weaknesses are mitigated. Furthermore, co-optimization is performed from a forecasting quality and training time trade-off perspective. The spatiotemporal framework consists of conditional spatiotemporal forecasting and modeling of the spatial dependencies of the PFE.

The main contributions of this paper are the following:

a. The introduction of a fully co-optimized hybrid probabilistic WPF model consisting of multi-objective optimization, advanced metaheuristics, the LUBE method, and NARX networks. The fully adaptive MOIAPSO algorithm enhances global search while ensuring fast and sufficient convergence. In contrast to existing co-optimized hybrid probabilistic WPF models, the proposed hybrid NARX-LUBE-MOIAPSO model is fully co-optimized from a forecasting quality and training time trade-off perspective.
b. The development of a novel conditional spatiotemporal forecasting methodology. The proposed approach is only applied to outliers in the forecasting results, to improve the overall forecasting performance while limiting the additional computational burden. Furthermore, spatiotemporal data are added in two waves, to avoid unnecessary additional information. To the authors knowledge, this is the first time a spatiotemporal framework is co-developed with a fully co-optimized hybrid probabilistic WPF model.
c. The modeling of the interdependence structure of WPPs, to enhance the comprehension of regional spatiotemporal dynamics, and potentially reduce the number of forecasts necessary for the target region. Interdependence structure modeling is based on a binary PFE metric, which is used for the first time to estimate the spatial dependencies of the PFE.

The remainder of this paper is organized as follows. A brief literature survey regarding state-of-the-art hybrid probabilistic WPF models and spatiotemporal methodologies is presented in Section II. Section III describes the proposed forecasting methodology, the evaluation metrics, and the spatiotemporal framework. Section IV provides a description of the case studies. Experimental results are presented and discussed in Section V. Concluding remarks are provided in Section VI.

II. LITERATURE SURVEY

Various hybrid probabilistic WPF models have been proposed in recent years. In [10], point predictions were generated using LSTM networks and k-means clustering of the input data. Nonparametric PIs were constructed using the KDE method. Several Gaussian process regression models were

used to construct combined PIs in [11]. VMD was applied to the input time series, and the weights of the combined models were determined based on area grey correlation. In [12], ELMs generated forecasts corresponding to the lower and upper bounds of noncentral PIs. The PIs were optimized using a chance-constrained optimization framework. A hybrid model comprising VMD and GRUs was used with the LUBE method in [13] to generate PIs. In [14], LSTM networks were combined with the LUBE method to generate PIs. The hybrid model was optimized using a fuzzy framework. Wavelet transformation was applied to the input time series of a hybrid PSO-LUBE-based model in [15]. In [18], multi-objective optimization was used to train the LSTM networks of a LUBE-based WPF model. NSGA-II was chosen as the optimization algorithm. A bidirectional GRU was used with LUBE for probabilistic WPF in [16]. In [17], two support vector machines were developed in a competitive fuzzy framework to generate PIs using the LUBE method. A distribution-free model was suggested in [20], for probability distribution function estimation. A parametric model was used to estimate the base distribution of wind power generation, which was then transformed using a spline-based normalizing flow.

Spatiotemporal dependencies regarding wind field patterns, forecasting error propagations, and uncertainty correlations, have also been exploited in several probabilistic WPF-related studies to improve forecasting performance and comprehension of regional spatiotemporal dynamics. In [21], spatiotemporal feature graphs were constructed, reflecting correlations among the data for different WPPs. The feature graphs were fed into a hybrid neural network to generate aggregated direct regional forecasts based on quantile regression. In [22], convolutional neural networks were trained with 2D images constructed using spatiotemporal wind field propagations, to generate deterministic and probabilistic forecasts. In [23], a k-nearest neighbor and conditional KDE were used to derive aggregated day-ahead probabilistic wind power forecasts over a region. The aggregated probabilistic forecasts were generated in the form of predictive densities, using the spatiotemporal correlations of the individual decentralized forecasts. The interdependence structure of five regions in western Denmark regarding forecasting uncertainty was modeled in [24]. The probabilistic predictions used were in the form of quantiles. In [25], two spatiotemporal forecasting methods were developed to exploit additional turbine-specific data. The first method was based on feature engineering whereas the second method implemented a hierarchical bottom-up approach.

None of the studies mentioned above performed a full co-optimization of all aspects of optimal hybrid probabilistic WPF. Some studies ignored certain aspects, such as the multi-objective nature of probabilistic forecasts optimization. In [18], where all components were co-optimized to some extent, co-optimization was performed with complete disregard for the computational complexity and training time required. Furthermore, the co-optimization of hybrid probabilistic WPF has never been co-developed with a spatiotemporal forecasting framework. In this paper, a hybrid probabilistic WPF model is fully co-optimized with respect to all aforementioned aspects, including computational complexity, and is co-developed with a holistic spatiotemporal framework including spatiotemporal forecasting and interdependence structure modeling. Table 1 summarizes the literature mentioned above in terms of the key aspects of this paper: full co- optimization, forecasting quality and training time trade-off, and spatiotemporal forecasting.

TABLE 1. Summary of literature survey including the proposed model, regarding the key aspects concerning this study.

Studies	Full co-optimization	Forecasting quality / training time trade-off	Spatiotemporal Framework
[11], [13], [14], [16], [17], [10], [20]	X	X	X
[12], [15]	X	√	X
[21]–[25]	X	X	√
[18]	√	X	X
Proposed	√	√	√

III. FORECASTING METHODOLOGY

A. CO-OPTIMIZED HYBRID PROBABILISTIC WPF MODEL

This subsection presents each component of the proposed hybrid probabilistic NARX-LUBE-MOIAPSO WPF methodology. The choice of each component is based on finding the global optimum of the probabilistic WPF problem while maintaining the computational complexity as low as possible. Furthermore, the components' structure and main parameters are co-optimized, with respect to each other's advantages and disadvantages.

1) MULTI-OBJECTIVE METAHEURISTICS-BASED OPTIMIZATION

WPF is a problem that consists of complex objective spaces with multiple local optima [26]. The choice of an appropriate optimization algorithm is crucial to avoid premature convergence to a local optimum. The traditional gradient descent-based backpropagation algorithms used in machine learning are prone to local optimum entrapment [7]. On the contrary, some metaheuristic algorithms employ search mechanisms that avoid premature convergence. Thus, a hybrid approach that combines machine learning and metaheuristics is appropriate for WPF.

In this paper, probabilistic forecasts are issued in the form of PIs. Optimal PIs are characterized simultaneously by maximum coverage and sharpness [5]. Ideally, PIs should cover all the corresponding observations while being as narrow as possible. However, coverage and sharpness maximization are two objectives in conflict. Thus, PI optimization is naturally a multi-objective optimization problem. Instead of converting

it to a single-objective optimization problem, maintaining the original multi-objective nature of PIs optimization has the advantage of reducing the overall bias, as it eliminates the bias introduced by artificial fixups used for fusing the optimization objectives into a single-objective optimization problem. Metaheuristic algorithms tend to be more efficient when less bias is introduced [7]; thus, a multi-objective optimization framework is chosen for the proposed hybrid model.

The multi-objective optimization framework is formulated as follows:

$$OBJ.1 : \max \frac{1}{N_{tr}} \sum_{t=1}^{N_{tr}} c_t \tag{1a}$$

$$OBJ.2 : \min \frac{1}{R} \sqrt{\frac{1}{N_{tr}} \sum_{t=1}^{N_{tr}} (u_t - l_t)^2} \tag{1b}$$

$$S.T. : \frac{1}{N_{tr}} \sum_{t=1}^{N_{tr}} c_t \geq cl \tag{1c}$$

$$c_t = \begin{cases} 1, \, if \; l_t \leq y_t \leq u_t \\ 0, \, else \end{cases} \tag{1d}$$

$$0 \leq l_t < u_t \leq 1 \; \forall t \in T. \tag{1e}$$

The problem objectives are given by (1a) and (1b), which represent the maximization of coverage and sharpness, respectively. Specifically, (1a) corresponds to PICP, i.e., the average coverage of the PIs, while (2a) corresponds to PINRW, which calculates the normalized root of the average squared width of the PIs [27]. The smaller the width of the PIs, the sharper they are. The multi-objective formulation of the problem consists of three constraints: The average coverage of the PIs should be greater than or equal to the pre-defined nominal confidence level (1c). Parameter c_t is a binary variable that indicates whether an actual observation lies inside the corresponding PI (1d). PI bounds are restricted between 0 and 1, corresponding to zero and nominal wind power generation, respectively. Furthermore, the upper and lower bounds of each PI should not be identical, to avoid generating point predictions (1e).

2) LUBE-BASED NARX NETWORK

The generation of optimal PIs for probabilistic WPF depends on the chosen PI generation approach. A nonparametric, direct PI generation method with little added bias to the overall forecasting procedure suits the metaheuristics-based optimization approach, described in subsection III-A.1. The LUBE method is a direct, simple, and fast PI generation approach, independent of strict mathematical formulations [19]. LUBE only requires a machine learning structure with two output units, representing the lower and upper bounds of the PI.

However, the LUBE method depends heavily on the parameters of the model. More parameters must be determined when dealing with complex machine learning structures, making it more difficult to optimize the LUBE method, while the computational complexity is significantly increased. Furthermore, metaheuristics and multi-objective optimization have already built a computationally expensive foundation for the overall framework.

Research focus has shifted greatly towards recurrent deep learning structures, such as LSTMs and GRUs, owing to their ability to capture long-term interdependencies in time series. However, an increase in dimensionality on its own does not necessarily avoid local optimum entrapment. Although hybrid deep learning – metaheuristics methods have been proven to work well, they require significant amounts of computational time and resources. In this paper, focus is driven towards finding a global optimization algorithm that fits the pursued objective of reaching a trade-off between the forecasting quality and training time.

The NARX network is a simple machine learning structure that maintains low complexity levels while still capturing temporal interdependencies in the time series. Thus, the NARX network is a suitable option for the proposed co-optimization approach. The core architecture of NARX is similar to that of feed-forward MLPs; however, lagged output feedback is provided to the input alongside other exogenous inputs.

The architecture of the NARX network, shown in Fig. 1, was chosen based on the co-optimization of the hybrid multi-objective metaheuristics – LUBE model. Extensive testing was performed to verify the optimal structure of the NARX network. A second hidden layer is necessary to assist the network in modeling more complex WPF input and output features. Further addition of hidden layers slightly improves the network's performance in some cases; however, it also significantly increases computational complexity. Feedback is provided by both outputs instead of fusing them into a single value, to facilitate the LUBE method with only a slight

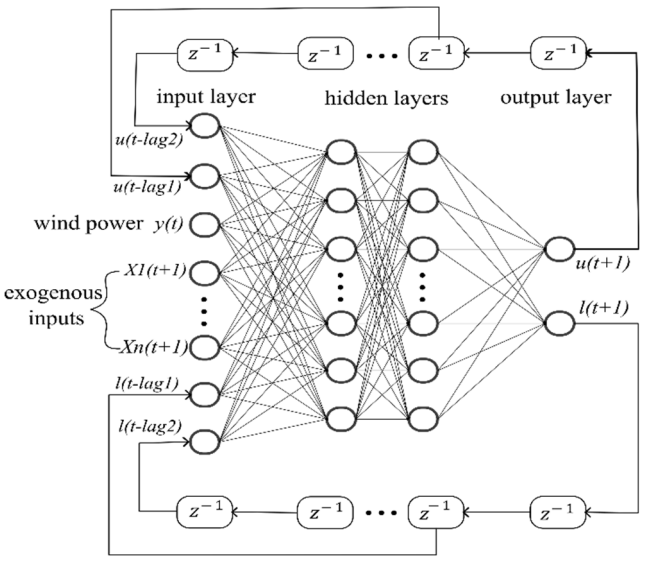


FIGURE 1. Optimized narx architecture for hybrid multi-objective metaheuristics - lube model.

computational increase. Furthermore, less bias is added to the model, which benefits metaheuristics-based optimization. Feedback is provided at two different lagged timesteps, which sufficiently captures the temporal interdependence of the wind power generation time series.

3) CO-OPTIMIZED IMPROVED ADAPTIVE PARTICLE SWARM OPTIMIZATION

Optimal PIs are reached by properly adjusting weights and bias connections of the NARX network. The choice of a suitable metaheuristic optimization algorithm is crucial for the optimal performance of the overall framework. PSO is known for its high convergence speed [28], which balances an already complex multi-objective optimization framework. Furthermore, PSO has relatively few parameters compared to other metaheuristics [7], which suits the highly parameter-sensitive LUBE method. According to PSO, each particle of the swarm represents a solution, i.e., a set of weights and biases connections. The positions of the particles correspond to sets of weights and biases, while the particle velocities indicate the step size and direction in the search space. For each particle, the weights and biases are updated based on the following equations:

$$w_{j,p}^{i+1} = w_{j,p}^{i} + v_{j,p}^{i+1} \forall j \in U; \forall p \in S \tag{2}$$

$$v_{j,p}^{i+1} = w_{in} v_{j,p}^{i} + c_1 r_1 \left(w_{j,p}^{i,best} - w_{j,p}^{i} \right) + c_2 r_2 \left(w_{j,swarm}^{i,best} - w_{j,p}^{i} \right) \forall j \in U; \forall p \in S. \tag{3}$$

However, PSO often suffers from premature convergence when dealing with complex objective spaces, such as those in the WPF problem [28]. An improved PSO algorithm is necessary to enhance global search while maintaining a high convergence speed and convergence rate.

A Gaussian mutation operator is applied to the positions of the particles to improve global searching capabilities. The mutation rate is co-optimized with the parameters of the hybrid model's components mentioned in subsections III-A.1 and III-AIII.A.2, and is given by the expression:

$$m^i = \begin{cases} 0.25 - \dfrac{i}{3I}, i \leq 0.75I \\ 0, i > 0.75I. \end{cases} \tag{4}$$

Initially, the mutation rate is sufficiently high to enhance stochastic search and avoid premature convergence. Specifically, 25% of the model's dimensions are mutated during the initial iteration. As the optimization proceeds, the mutation rate decreases linearly to allow convergence of particles. Mutation is eliminated after 75% of the total iterations to enhance local search.

Increasing global searching capability usually comes at the expense of the convergence rate. Adaptive determination of the algorithm's parameters helps ensure sufficient convergence while maintaining global searching and convergence speed at high levels. A fully adaptive metaheuristic algorithm tends to operate better because the parameters are optimally adjusted according to the optimization state.

Many studies have suggested methods for adaptively changing PSO parameters. In this paper, the adaptive inertia weight formula suggested in [29] is used for each particle in each iteration, which is given by the following expression:

$$w_{in}^{i} = w_{in}^{0} \left(1 - \frac{\sqrt{\sum_{j=1}^{n\_dim} (w_{j,swarm}^{i,best} - w_{j,p}^{i2})}}{\max_j (w_{j,swarm}^{i,best} - w_{j,p}^{i})} \right) \tag{5}$$

where w_{in}^{0} is a random number in interval [0.5, 1]. This formulation directs the particles towards the swarm's best position when they move too far away from it, to ensure sufficient convergence. Furthermore, the acceleration coefficients of each particle are adaptively determined in each iteration for each weight connection based on the following sigmoid function f, as suggested in [8]:

$$f(D) = \frac{0.5}{1 + e^{-aD}} + 1.5 \tag{6}$$

$$c_{1,j}^{i} = f \left(w_{j,p}^{i,best} - w_{j,p}^{i} \right) \forall j \in U \tag{7}$$

$$c_{2,j}^{i} = f \left(w_{j,swarm}^{i,best} - w_{j,p}^{i} \right) \forall j \in U. \tag{8}$$

In (6), a is equal to 0.0035 % of the search range. The adaptive strategies described in (5)-(8) are chosen owing to their ability to further improve the convergence rate while maintaining global searching capabilities.

4) MULTI-OBJECTIVE IMPROVED ADAPTIVE PARTICLE SWARM OPTIMIZATION

The MOIAPSO adopted in this paper is based on the classic multi-objective PSO proposed in [30], which is further improved by employing an adaptive grid [31] and binary tournament selection. These improvements are crucial for achieving a trade-off between forecasting quality and training time. The adaptive grid encourages global search by increasing the diversity among the stored non-dominated solutions, while the binary tournament method for the selection of the swarm leader per iteration reduces the overall training time.

The overall process of the MOIAPSO algorithm is shown in Fig. 2. Weights and biases are initialized using the Nguyen – Widrow technique. After the evaluation, a repository is created to store the non-dominated solutions. In the case of repository overflow, the least diverse solutions are discarded. The repository solutions are mapped according to their performance for each objective. The adaptive grid is created, covering all solutions, to determine their diversity. A crowding index is assigned to each hypercube of the grid based on the total number of solutions inside the hypercube. A high crowding index indicates that solutions inside the hypercube are less diverse. The grid dimensions change adaptively to cover all repository solutions. Binary tournament selection is applied to the repository solutions to select the current swarm leader based on the crowding index. Mutation is applied to the new positions, updated using (2) and (3). The personal best position of a particle is updated if it is

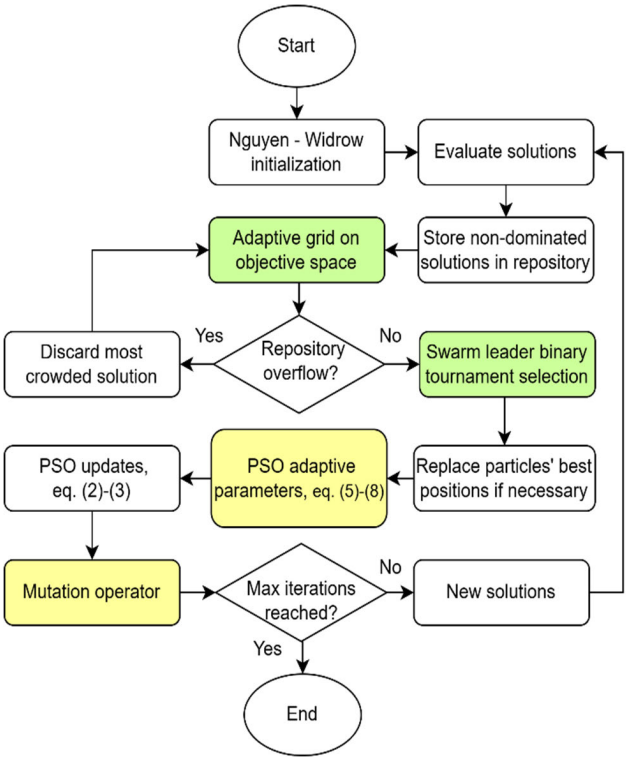


FIGURE 2. Flow chart of proposed MOIAPSO algorithm. Processes in yellow correspond to the general improvements of PSO while processes in green correspond to the improvements of the multi-objective version of the algorithm.

dominated by its current position. If non-domination occurs between the current position and the personal best position of a particle, the personal best position is chosen randomly.

B. EVALUATION METRICS

Reliable evaluation metrics are necessary to assess the forecasts generated by the proposed hybrid method efficiently. PIs should be evaluated not only in terms of coverage and sharpness but also based on their proximity to actual observations. Therefore, it is preferable to use all-around probabilistic evaluation metrics.

A widely acknowledged probabilistic evaluation metric, often used in WPF, is CRPS [32]. CRPS evaluates the distribution of forecasts in terms of both shape and location, by measuring the distance between the forecast probability distribution and the distribution of the observations. For the entire test set, CRPS is calculated using the following equation:

$$\mathrm{CRPS} = \frac{1}{N_t}\sum_{t=1}^{N_t}\int_{-\infty}^{+\infty}\left(F\left(z_t\right) - \mathbb{1}\left(z_t - y_t\right)\right)^2 dz \tag{9}$$

where F is the CDF of the generated forecasts and $\mathbb{1}$ is the Heaviside step function. The generated PIs are represented by variable z, which, in this paper, is calculated as the average of the lower and upper bounds of each PI.

Another all-around probabilistic evaluation metric is PL [32]. PL is mainly used to evaluate quantiles, by measuring the distance between the actual value and the forecasted quantile, weighted by the target quantile. PL can also be employed for PI evaluation using the following formula:

$$\mathrm{PL} = \frac{1}{N_t}\sum_{t=1}^{N_t} d_t \tag{10a}$$

$$d_t = \begin{cases} \left(l_t - y_t\right)cl, y_t < l_t \\ \left|z_t - y_t\right|\left(1 - cl\right), l_t \le y_t \le u_t \\ \left(y_t - u_t\right)cl, y_t \ge u_t. \end{cases} \tag{10b}$$

PL for PIs calculates the weighted distances between the actual observations and some point of the corresponding PIs. The weight values and points of the PIs chosen depend on whether the actual observations lie within their corresponding PIs.

A probabilistic evaluation metric often used in related literature is CWC. CWC considers the coverage and sharpness of the PIs, but not the shape of their distribution. CWC is calculated as follows:

$$\mathrm{CWC} = \frac{1}{N_t}\sum_{t=1}^{N_t}\left(u_t - l_t\right) + \gamma e^{-h\left(\frac{1}{N_t}\sum_{t=1}^{N_t}(c_t) - cl\right)} \tag{11}$$

where γ equals 1 when the average coverage of the PIs is less than the nominal confidence level; otherwise γ equals 0.

Certain concerns have been expressed regarding the sufficiency of CWC [33], mainly because of its inability to assess the distribution shape of PIs. On the other hand, PL is more sensitive to data outliers than CRPS. Furthermore, PL is a probabilistic evaluation metric that is difficult to interpret. Thus, while all these metrics (CRPS, PL, and CWC) are used to assess the generated PIs, CRPS is considered the main evaluation metric.

C. SPATIOTEMPORAL FRAMEWORK

1) SPATIOTEMPORAL FORECASTING

WPF in a spatiotemporal framework has proven effective for improving forecasting results [34]. However, additional spatiotemporal information does not always have a beneficial impact on the overall performance. The spatiotemporal forecasting performance depends on various factors, such as data quality, the WPF model used, feature engineering, and geographical characteristics.

In this paper, spatiotemporal data are only used to eliminate cases with extremely poor forecasting performance. Forecasts are generated multiple times to verify the inability of the corresponding WPF model to converge to a sufficiently good local optimum. These outliers in the forecasting results are mainly a consequence of either poor-quality data or the inability of the model to navigate through the search space because of the complex shape of the objective space. In these cases, spatiotemporal data can potentially provide additional useful information.

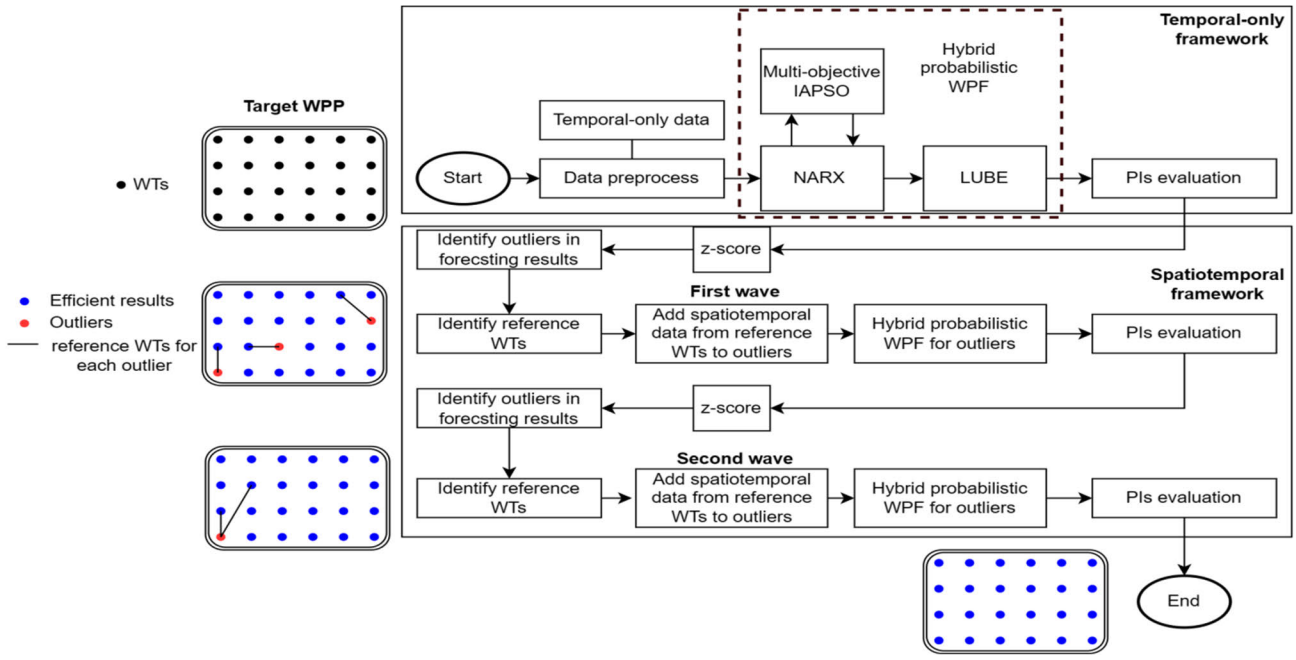


FIGURE 3. The proposed spatiotemporal forecasting methodology.

The spatiotemporal forecasting methodology is illustrated in Fig. 3. Initially, forecasts are generated for each WT of the WPP using only the corresponding turbine-specific data. The hybrid probabilistic WPF model is trained using the MOIAPSO algorithm presented in Fig. 2. PIs are generated and evaluated for each WT of the WPP. To identify outliers in the forecasting results, the z-score of each WT is calculated as follows:

$$Z_g = \frac{\text{CRPS}_g - E_P(\text{CRPS})}{S_P(\text{CRPS})} \forall g \in P \tag{12}$$

where CRPS_g is calculated from (9) and $E_P(\text{CRPS})$, $S_P(\text{CRPS})$ are the mean and standard deviation of crps values respectively, for all wts in the wpp. WTS WITH $Z_g > 3$, are considered outliers. next, a reference wt is assigned to each outlier based on the correlation of their data. the data of the reference wts are added to the corresponding outlier wts (first wave of additional spatiotemporal input). pis are generated again, this time only for outlier cases. After evaluation, outliers are identified (if any), and the spatiotemporal process is repeated. Correlation is measured on another feature if available, otherwise, the second highest-correlated wt, based on the feature selected in the first wave, is chosen as reference. The new datasets are formulated (second wave of additional spatiotemporal input), and pis are generated and evaluated for the remaining outlier cases.

The spatiotemporal forecasting methodology is designed with the aim of eliminating outliers in the forecasting results while keeping the computational complexity of the overall framework relatively low. Outliers have a significant negative effect on overall forecasting accuracy; hence, it is crucial to limit them as much as possible. Furthermore, spatiotemporal data are added in two waves to reduce the amount of overall data required. More waves of additional spatiotemporal data could be used; however, this would further increase the computational complexity without necessarily improving the results.

2) SPATIAL DEPENDENCE IN PFE

Modeling the interdependence structure of a WPP can significantly improve the comprehension of the spatiotemporal dynamics of the region. Furthermore, in regions with high dependencies between units, the overall number of forecasts necessary to estimate the aggregated energy yield can potentially be reduced.

In this paper, the spatial dependence between the WTs of the WPP is modeled based on the PFE. The chosen PFE metric is c_t of 1(d), and dependence is calculated using the Pearson correlation coefficient. To the best of our knowledge, this PFE binary metric has never been used to model the interdependence structure of WPPs.

IV. CASE STUDIES

The proposed NARX-LUBE-MOIAPSO WPF forecasting model and spatiotemporal methodology are tested in two different case studies to ensure efficient evaluation. Each case study consists of turbine-specific measurements of the WTs located in the same WPP. The datasets for both case studies are publicly available.

A. CASE STUDY I

The first case study consists of data regarding 200 randomly selected WTs of a WPP and three meteorological

masts. The WTs are located in a flat field in the United States [35]. Hourly wind speed and power output data are provided for each WT. Similarly, hourly wind speed and direction measurements are provided for each meteorological mast. The measurements were conducted over one year, from September 2010 to August 2011.

The relative positions of the WTs and meteorological masts are shown in Fig. 4. All coordinates are shifted from their actual values, to keep the WPP's real location confidential. However, the layout of WTs inside the WPP reflects their actual relative positions. To simplify the modeling of spatial dependence in the PFE, the WPP is divided into eight clusters using the k-means clustering technique. The clustering results are shown in Fig. 4.

In the temporal-only framework, the wind speed and power output measurements of each WT are used as inputs to its corresponding NARX network. In the spatiotemporal forecasting framework, additional data are added as exogenous inputs to the NARX networks. The first wave of additional spatiotemporal data consists of wind direction measurements, to provide different types of information to the NARX networks. Specifically, for each WT, the wind direction measurements of the closest meteorological mast are added as input. Since no other meteorological feature is available, the second wave of additional spatiotemporal data consists of wind speed measurements. For each WT, the wind speed measurements of the highest correlated reference WT are added as input. The Pearson correlation coefficient is used to calculate the correlations between the wind speed datasets of the WTs.

B. CASE STUDY II

The second case study consists of SCADA measurements of a coastal WPP located in the United Kingdom, with an installed capacity of 28.7MW [36]. The WPP consists of 13 WTs, the relative positions of which are shown in Fig. 5. The dataset contains measurements from 2016 to 2021 provided in 10-minute intervals.

Similar to case study I, wind speed and power output measurements of each WT are used as input to the corresponding NARX network in the temporal-only framework. Wind direction and wind speed data measured at the location of WTs, which are highly correlated with the target WT, are added as inputs in the spatiotemporal forecasting framework. Specifically, the first wave of additional spatiotemporal data consists of wind direction measurements that have the highest correlation with the wind direction measured at the location of the target WT. Correspondingly, the third wave of additional spatiotemporal data consists of the highest correlated wind speed measurements. All correlations are calculated using Pearson's correlation coefficient.

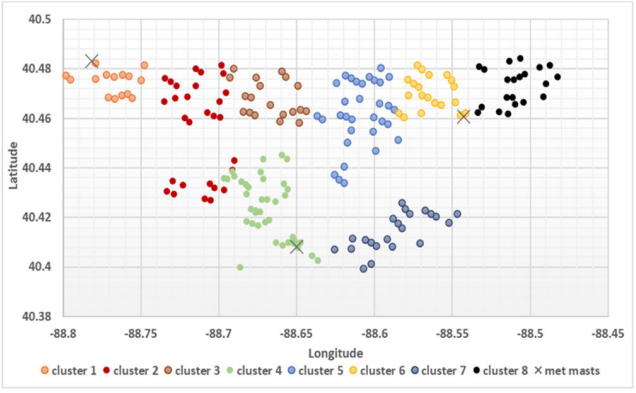


FIGURE 4. Relative positions of WTs and meteorological masts as well as clustering results of case study I.

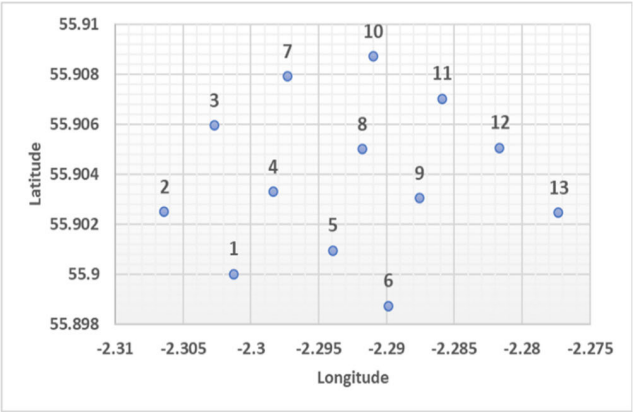


FIGURE 5. Relative positions of WTs of case study II.

V. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETTINGS AND COMPARISON MODELS

This subsection provides indicative structure and parameter optimization results that prove the efficiency of the co-optimization choices presented in Section III. Furthermore, it describes the benchmark probabilistic WPF models used. All runs were conducted on an Intel(R) Core (TM) i7-8700 CPU (3.20GHz, 6 cores) desktop computer with 8 GB of RAM and all WPF models were developed in Python.

1) PARAMETERS OPTIMIZATION

The results for different versions of the PSO algorithm are presented in Table 2. Specifically, PSO with mutation only (mPSO), mutated PSO with adaptive inertia weight (mAIWPSO), mutated PSO with adaptive acceleration coefficients (mAACPSO), and fully adaptive mutated PSO (proposed IAPSO) are compared. All versions of PSO are based on single-objective optimization and the results are generated for one WT of both case studies. The results are generated 20 times for each case study, and the empirical CDF of the PIs is constructed for CRPS calculation. The adaptive selection of PSO parameters improves the results of the mutated PSO, while the proposed fully adaptive PSO generates the most accurate results.

Indicative comparative results of the optimization procedure of a WT of case study I, between the proposed

TABLE 2. Indicative results for 1 WT of each case study for different versions of single-objective PSO.

Model	CRPS	
	Case study I	Case study II
IAPSO (proposed)	0.052988	0.049640
mPSO	0.058847	0.058655
mAIWPSO	0.057468	0.058538
mAACPSO	0.057069	0.054505

MOIAPSO and single-objective IAPSO, are presented in Fig. 6. The solution of the repository with the smallest CRPS value is chosen in each iteration for MOIAPSO. While single-objective IAPSO converges faster, the proposed MOIAPSO performs slightly better at searching globally. The convergence rates of both the models are similar. Furthermore, the proposed MOIAPSO algorithm generates a set of solutions instead of a single solution, thereby offering additional flexibility to the system operator.

Both time lags for the feedback of the NARX network are selected based on the ACF of the wind power generation time series. The ACF values of lags of up to 72 hours for one WT in each case study are presented in Fig. 7. In both cases studies, the smaller the time lag for the first 24 hours, the higher the ACF value. Thus, the first feedback value is only lagged 1 timestep. The second peak of the ACF values differs between the case studies. In case study I, the second peak is observed after 72 hours, whereas in case study II, the second peak is observed after 38 hours. Thus, the second feedback is lagged by 72 and 38 timesteps for case studies I and II, respectively.

2) COMPARISON MODELS

The proposed NARX-LUBE-MOIAPSO model is compared with several other hybrid probabilistic WPF models to ensure efficient evaluation, such as the LSTM-based model proposed in [10]. The LSTM-based model uses k-means clustering for the input features and forms PIs by applying KDE to the point predictions generated. The CCELM model proposed in [12] is

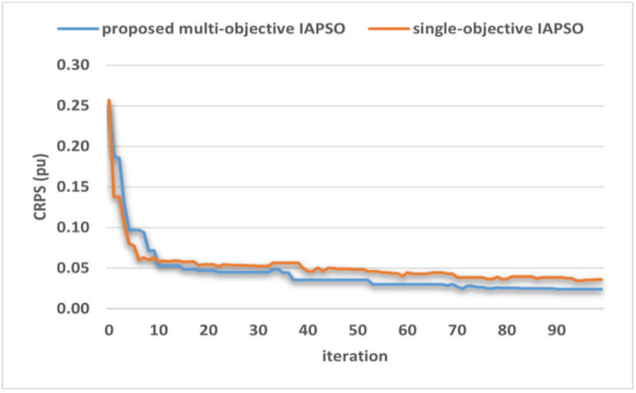


FIGURE 6. Comparative results of optimization procedure between single-objective IAPSO and the proposed MOIAPSO.

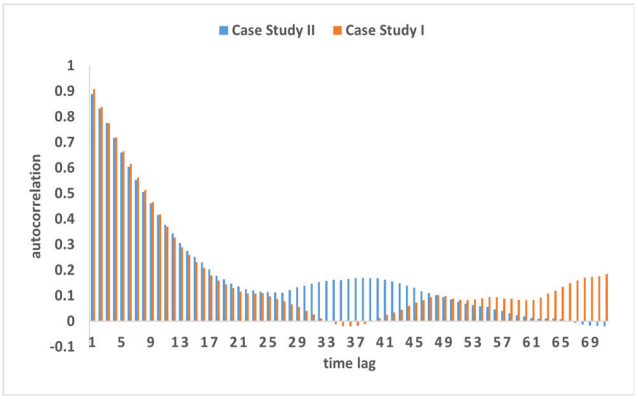


FIGURE 7. ACF values of the wind power generation time series of a WT for both case studies.

also selected for comparison. CCELM comprises two ELMs representing the lower and upper bounds of the PIs, optimized using a chance-constrained optimization framework. Furthermore, the hybrid model suggested in [18] is selected as a benchmark. The model is based on LSTMs and LUBE for the construction of PIs, which are optimized using a multi-objective optimization framework with NSGA-II.

B. TEMPORAL-ONLY FRAMEWORK

All input variables are normalized to the interval $[-1, 1]$ while the wind power output is normalized to the interval $[0, 1]$. The datasets used in both case studies are yearlong. Initially, results are generated for both case studies without considering the spatiotemporal forecasting methodology, to compare the performance of each probabilistic WPF model. The multi-objective optimization models generate a set of final solutions, which are all evaluated. The solution with the lowest CRPS value is selected for comparison with the solutions generated by the rest of the models. Fig. 8 presents an example of the final repository solutions generated by NARX-LUBE-MOIAPSO, for 1-hour ahead forecasts and a 90% confidence level, for case study I. The solutions are mapped according to the performance of their objective functions. The solution with the lowest CRPS evaluation value is represented as the red particle.

Perfect wind speed forecasts and power output measurements are provided as input to the models for 1-hour ahead forecasting. All the input data are turbine-specific, and no spatiotemporal reference data are used. The confidence level is set to 90%. The average value of CRPS throughout each WPP is calculated in both case studies as follows:

$$\mathrm{CRPS}^{j}_{avg} = \frac{1}{N^{j}_{WT}} \sum_{g=1}^{N^{j}_{WT}} \mathrm{CRPS}_{g}, j = 1, 2. \tag{13}$$

The average values of PL and CWC are calculated similarly. The CDF for calculating CRPS of each WT in (9) is constructed by assuming a normal distribution of the PIs. In each timestep t, the mean of the distribution is equal to z_t,

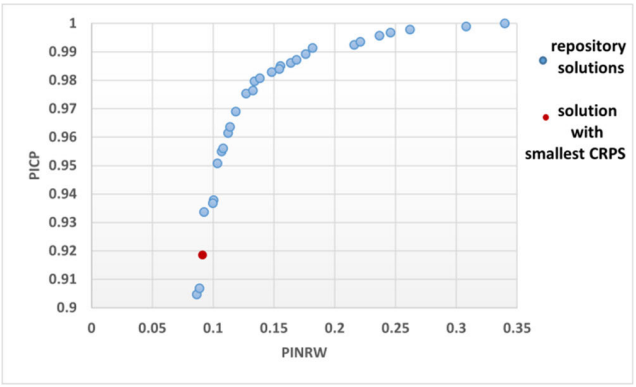


FIGURE 8. Final repository solutions of NARX-LUBE-MOIAPSO, for a WT of case study I. The solutions are mapped according to their objective functions' scores. The solution represented by the red particle is the solution that generates PIs with the smallest CRPS value. The forecasting horizon is 1 hour, and the confidence level is set to 90%.

and the standard deviation is calculated as:

$$s_t = \frac{u_t - l_t}{2 \cdot z_{crit}(cl)} \tag{14}$$

where $z_{crit}(cl)$ represents the critical value of cl in the normal distribution. The aggregated results for both the case studies are presented in Table 3. The proposed model clearly outperforms LSTM-KDE and CCELM in both case studies for all evaluation metrics. LSTM-KDE and CCELM are the main competitors of the proposed model, because of their relative proximity regarding their computational complexity. Compared to LSTM-LUBE-NSGA-II, the proposed model generates slightly worse results. Specifically, the proposed model's performance regarding CRPS is 6.7% and 1.03% worse than the performance of LSTM-LUBE-NSGA-II for case studies I and II respectively. However, the average training time required for the proposed model is approximately 75% less than that required for LSTM-LUBE-NSGA-II. Furthermore, the proposed model outperforms LUBE-LSTM-NSGA-II in case study II in terms of PL and CWC. Provided with the higher time-resolution data of case study II, the proposed model has a performance similar to that of the state-of-the-art. Only CCELM outperforms the proposed model in terms of required training time, owing to the extremely fast nature of ELMs. Compared to all other benchmarks, the proposed model requires much less training time. Therefore, from a forecasting quality and training time trade-off perspective, the proposed model shows superior performance.

The VIKOR method [37] is used to further verify the forecasting quality and training time trade-off analysis. VIKOR is a popular multi-criteria-decision-making method that ranks alternatives according to their performance for each criterion and selects the closest alternative to the ideal solution. The results of the VIKOR analysis are shown in Fig. 9. Different combinations of the relative weight values are tested for the two criteria. The sum of the forecasting quality and training time weights is always equal to one. The metric Q is a normalized aggregating function of the group utility and individual regret. Alternatives are ranked from worst to best in decreasing order according to their Q values. The smaller the value of Q is, the closer the alternative is to the ideal solution. Multiple alternatives can be considered compromise solutions. Further information on the VIKOR method can be found in [37]. The encircled points in Fig. 9 indicate the promising compromise solutions for each weight combination. In both case studies, the proposed NARX-LUBE-MOIAPSO model is considered a compromise solution for all weight combinations in the interval [0.1, 0.9]. This proves

TABLE 3. Aggregated results in temporal-only framework.

Model	$CRPS_{avg}$	PL_{avg} (%)	CWC_{avg}	Avg training time (s)
Case study I				
NARX-LUBE-MOIAPSO (proposed)	0.0492	0.7778	31.053	353
LSTM-LUBE-NSGA-II	0.0459	0.7662	28.447	1459
LSTM-KDE	0.0551	0.8089	36.700	967
CCELM	0.075	1.209	38.119	91
Case study II				
NARX-LUBE-MOIAPSO (proposed)	0.0491	0.8161	12.194	2108
LSTM-LUBE-NSGA-II	0.0486	0.8458	12.431	8707
LSTM-KDE	0.0562	0.9532	28.923	5012
CCELM	0.0652	1.0523	97.315	538

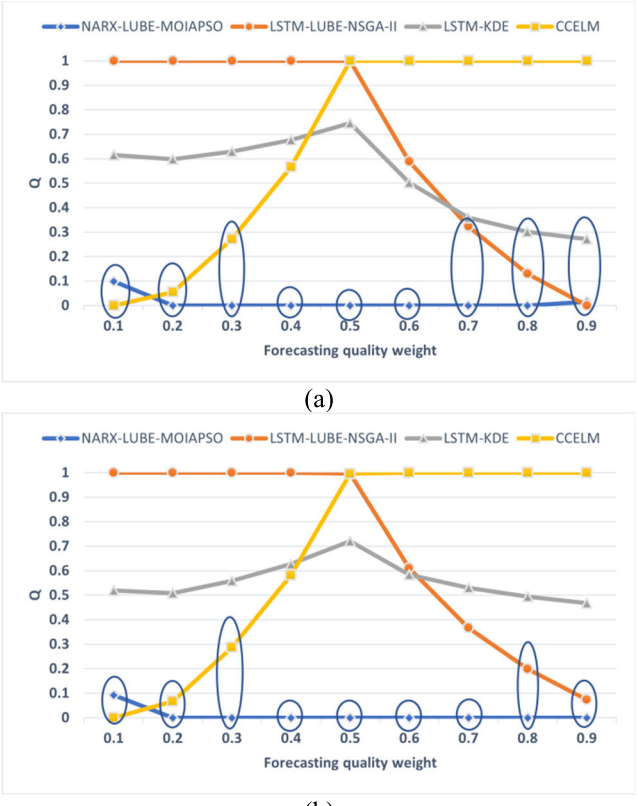


FIGURE 9. Results of the VIKOR analysis. (a) case study I (b) case study II.

that the proposed model achieves the desired forecasting quality and training time trade-off. Furthermore, as shown in Fig. 9, the proposed model achieves the best Q-score for the majority of weight combinations. The only exceptions are when training time or forecasting quality are considered extremely dominant criteria (> 0.8). In the first case, CCELM outperforms the proposed model in both case studies, whereas in the latter case, LSTM-LUBE-NSGA-II outperforms the proposed model in case study I.

C. SPATIOTEMPORAL FRAMEWORK

1) SPATIOTEMPORAL FORECASTING

In this subsection, the results are generated using the spatiotemporal forecasting methodology shown in Fig. 3. The z-score is calculated for both case studies for all models, to identify outliers in the forecasting results of the temporal-only framework. The z-score of all the WTs in case study I is presented in Fig. 10. The points inside the circle are considered outliers. The number of outliers of each model in both case studies, before and after the application of the spatiotemporal methodology, is presented in Table 4. It can be seen that 7 out of 200 forecasting cases of the proposed model are considered outliers in case study I, while one case is considered as outlier in case study II. After applying the spatiotemporal forecasting methodology, all outlier cases of the proposed model are eliminated for both case studies. The improved results after applying the spatiotemporal forecasting methodology are presented in Table 5. The proposed model improves its overall performance and manages to outperform LSTM-LUBE-NSGA-II in case study II while maintaining a limited increase in required training time (4.53% AND 13.57% in case studies I and II, respectively). The performance of LSTM-KDE and CCELM is also significantly improved, and only two outlier cases remain in case study I. It is evident that the proposed spatiotemporal forecasting methodology eliminates almost all outlier cases, significantly improving the stability of the probabilistic WPF models. Furthermore, the forecasting quality and training time trade-off perspective is maintained, as only a limited additional computational burden is introduced. Even for CCELM, which has the most outlier cases in case study I, the required training time is only increased by 69.23%, preserving it as the fastest model.

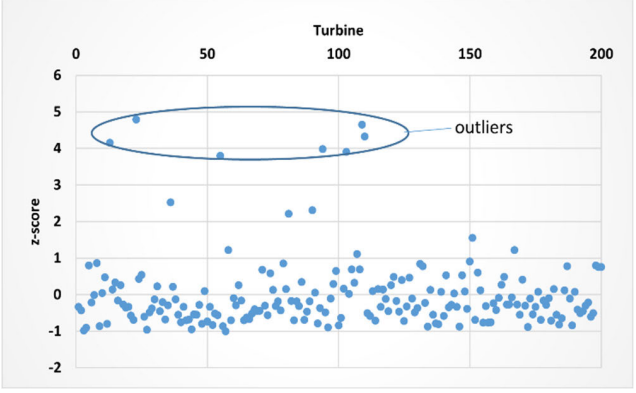


FIGURE 10. Z-score of WTs in case study I. The points inside the circle represent the outlier cases.

TABLE 4. Outlier cases before and after the spatiotemporal forecasting methodology application for both case studies.

Model	Before spatiotemporal forecasting		After spatiotemporal forecasting	
	Case study I	Case study II	Case study I	Case study II
NARX-LUBE-MOIAPSO (proposed)	7	1	0	0
LSTM-LUBE-NSGA-II	0	0	0	0
LSTM-KDE	13	2	2	0
CCELM	27	2	2	0

Perfect meteorological forecasts (exogenous inputs) were considered in all results presented above. However, this assumption is unrealistic. Meteorological data used as input are either measurements at time t when forecasts are issued, or forecasts for time t + 1,2,..., depending on the forecasting horizon. In both cases, the actual meteorological values realized on the forecasting horizon are unknown, thus an additional error is inherited from the input of the WPF model. Efficient WPF evaluation necessitates testing the performance of the model under realistic circumstances. Table 6 presents forecasting results considering the forecasting horizon and the error inherited by the input data used. Specifically, forecasts are issued at time t with a 1-hour forecasting horizon, using only data known at time t. This is the worst-case scenario since no information regarding time $t + 1$ is available; thus, the inherited input error is maximum. Only the results of the two best probabilistic WPF models are presented, i.e., the proposed model and LSTM-LUBE-NSGA-II. Again, the forecasting quality of the models is similar, obviously worsened compared to the case where perfect exogenous input parameters are available. After the application of the spatiotemporal forecasting methodology, all outlier cases of the proposed co-optimized model are eliminated. Furthermore, the proposed model outperforms LSTM-LUBE-NSGA-II in both case studies.

The PIs of different confidence levels generated by the proposed model are shown in Fig. 11. Specifically, PIs are generated for 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, and 99% confidence levels for 24 hours-ahead forecasting. This time, forecasts of exogenous meteorological parameters are used as input. Fig. 11 illustrates the PIs generated for 1 WT of case study I on a relatively "difficult" day, in which wind power fluctuations are frequent and intense. The larger the forecasting horizon, the higher the uncertainty of the forecast. This is mainly due to the inherited error of the meteorological parameter forecasts, which is propagated to the probabilistic WPF model. The proposed WPF model captures the wind power time series behavior sufficiently up to about an 8 hours-ahead forecasting horizon. Hence, the proposed model is suitable for short-term forecasts, from one to several hours ahead (intra-day forecasting).

TABLE 5. Aggregated results in spatiotemporal framework.

Model	$CRPS_{avg}$	PL_{avg} (%)	CWC_{avg}	Avg training time (s)
Case study I				
NARX-LUBE-MOIAPSO (proposed)	0.04664	0.6523	31.246	369
LSTM-LUBE-NSGA-II	0.0459	0.7662	28.447	1459
LSTM-KDE	0.05238	0.6923	23.298	1087
CCELM	0.07042	1.006	38.263	154
Case study II				
NARX-LUBE-MOIAPSO (proposed)	0.04807	0.7843	12.304	2394
LSTM-LUBE-NSGA-II	0.0486	0.8458	12.431	8707
LSTM-KDE	0.05294	0.9842	23.637	5796
CCELM	0.05998	1.0502	94.5902	626

TABLE 6. Results for 1-hour ahead forecasts issued for the case where no information about future values of exogenous variables is available (worst-case scenario).

Model	Before spatiotemporal forecasting		After spatiotemporal forecasting	
	$CRPS_{avg}$	outliers	$CRPS_{avg}$	outliers
Case study I				
NARX-LUBE-MOIAPSO (proposed)	0.08414	5	0.08241	0
LSTM-LUBE-NSGA-II	0.08377	0	0.08377	0
Case study II				
NARX-LUBE-MOIAPSO (proposed)	0.0847	1	0.08366	0
LSTM-LUBE-NSGA-II	0.08374	0	0.08374	0

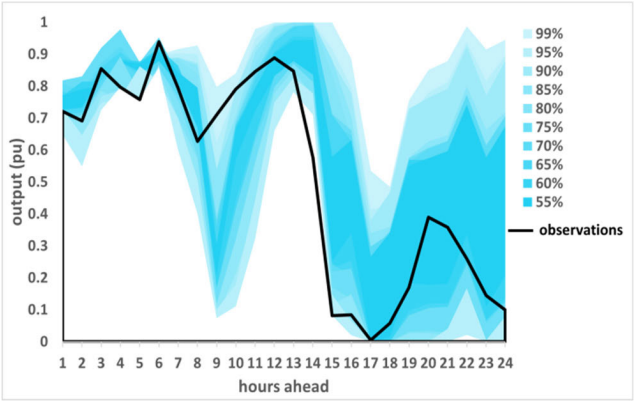


FIGURE 11. 24 hours-ahead PIs generated for different confidence levels in case study I, as well as the real power output of the corresponding WT. Forecasted meteorological data are used as input.

2) SPATIAL DEPENDENCE IN PFE

The spatial dependence in the PFE estimation is based on the results generated only by the proposed model. A PFE occurs when the PI does not cover the corresponding actual output value. The interdependence structures of the WPPs are modeled using PFE correlation coefficient matrices, as shown in Fig. 12. In case study I, the percentage of PFE is calculated in each cluster.

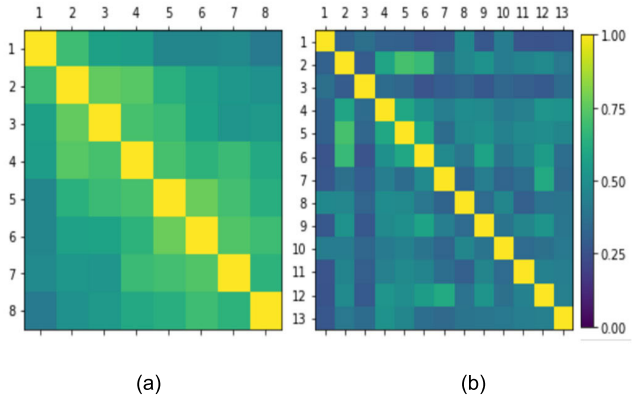


FIGURE 12. Correlation coefficient matrices regarding PFE. (a) Clusters of case study I. (b) WTs of case study II.

The average correlation coefficient values are 0.6145 and 0.4114 for case studies I and II, respectively. The relatively higher average correlation coefficient value measured for case study I is a result of the clustering aggregation effect. Furthermore, the WPP of case study I is located in an inland flat field, whereas the WPP of case study II is located near the coastline where strong, volatile wind fields are common. Thus, the interdependence structural complexity of the WPP in case study II is higher, and higher-level statistical tools are necessary to model it sufficiently. The highest correlation coefficient values in case study II are measured for WT 2. The area covered by turbines 2, 4, 5, and 6 is the area with the highest interdependence, probably because of the prevailing wind fields in the area. On the other hand, the interdependence structure in case study I is simpler, and the patterns are easier to identify. The correlation between neighboring clusters is higher. Clusters that lie in the center of the WPP have higher correlations on average, compared to clusters lying at the edges. Thus, the PFE propagates in a smoother manner resulting in greater dependence between the units of the WPP.

The interdependence structure modeling provides additional information to the system operator about the quality of the forecasts. Uncertainty is further quantified since the PFE-based spatiotemporal dependence modeling warns about future PIs coverage or non-coverage in highly correlated areas. In these areas, the total number of necessary forecasts could potentially be reduced, since a single forecast would provide information about several WTs. However, the focus of this paper was to investigate whether the PFE metric could sufficiently model the interdependence structures of the WPPs. Developing a methodology to exploit spatiotemporal dependencies is out of the scope of this study.

VI. CONCLUSION

Accurately forecasting wind power generation is essential to ensure stable operation of electric power systems. Interest

in hybrid probabilistic approaches that are able to capture uncertainties, is increasing. In this paper, a co-optimized hybrid model consisting of multi-objective optimization, metaheuristics, LUBE, and NARX networks is developed for short-term probabilistic WPF. The PIs generated are optimized using an improved adaptive version of PSO. Furthermore, a conditional spatiotemporal forecasting methodology and a spatial dependence in the PFE modeling approach are suggested to improve forecasting performance and comprehension of spatiotemporal dynamics.

The experimental results indicate the superiority of the proposed fully co-optimized hybrid model when viewed from the perspective of a forecasting quality and training time trade-off. A similar forecasting performance to that of deep learning-based hybrid models can be achieved when focus is driven towards co-optimization, while significantly less training time is required. The proposed spatiotemporal forecasting methodology overall improves the forecasting performance since it eliminates the majority of outlier cases. Furthermore, interdependence structures are sufficiently modeled by a binary PFE metric, resulting in an enhanced understanding of the regional spatiotemporal dynamics.

Future research efforts should consider including hybrid versions of optimization algorithms in the full co-optimization approach of probabilistic WPF. Combining different metaheuristic, evolutionary, or even gradient-descent based algorithms can have positive effects on global optimization capabilities. Moreso, some aspects of this paper could be further investigated, such as the impact of the temporal resolution of data on the forecasting results, or the applicability of the proposed spatiotemporal framework to real energy management optimization tasks. Furthermore, it would be useful to validate the spatiotemporal framework on datasets that are more coarse-grained spatially, as well as on larger regions that include more than a single WPP.

REFERENCES

[1] R. J. Bessa, P. Pinson, G. Kariniotakis, D. Srinivasan, C. Smith, N. Amjady, and H. Zareipour, "Guest editorial for the special section on advances in renewable energy forecasting: Predictability, business models and applications in the power industry," *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 1166–1168, Apr. 2022.

[2] V. Prema, M. S. Bhaskar, D. Almakhles, N. Gowtham, and K. U. Rao, "Critical review of data, models and performance metrics for wind and solar power forecast," *IEEE Access*, vol. 10, pp. 667–688, 2022.

[3] I. K. Bazionis and P. S. Georgilakis, "Review of deterministic and probabilistic wind power forecasting: Models, methods, and future research," *Electricity*, vol. 2, no. 1, pp. 13–47, Jan. 2021.

[4] K. L. Jørgensen and H. R. Shaker, "Wind power forecasting using machine learning: State of the art, trends and challenges," in *Proc. IEEE 8th Int. Conf. Smart Energy Grid Eng. (SEGE)*, Aug. 2020, pp. 44–50.

[5] H. Quan, A. Khosravi, D. Yang, and D. Srinivasan, "A survey of computational intelligence techniques for wind power uncertainty quantification in smart grids," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4582–4599, Nov. 2020.

[6] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255–270, Apr. 2014.

[7] P. Lu, L. Ye, Y. Zhao, B. Dai, M. Pei, and Y. Tang, "Review of meta-heuristic algorithms for wind power prediction: Methodologies, applications and challenges," *Appl. Energy*, vol. 301, Nov. 2021, Art. no. 117446.

[8] W. Liu, Z. Wang, Y. Yuan, N. Zeng, K. Hone, and X. Liu, "A novel sigmoid-function-based adaptive weighted particle swarm optimizer," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 1085–1093, Feb. 2021.

[9] F. Kılıç, İ. H. Yılmaz, and Ö. Kaya, "Adaptive co-optimization of artificial neural networks using evolutionary algorithm for global radiation forecasting," *Renew. Energy*, vol. 171, pp. 176–190, Jun. 2021.

[10] B. Zhou, X. Ma, Y. Luo, and D. Yang, "Wind power prediction based on LSTM networks and nonparametric kernel density estimation," *IEEE Access*, vol. 7, pp. 165279–165292, 2019.

[11] N. Huang, Y. Wu, G. Lu, W. Wang, and X. Cao, "Combined probability prediction of wind power considering the conflict of evaluation indicators," *IEEE Access*, vol. 7, pp. 174709–174724, 2019.

[12] C. Wan, C. Zhao, and Y. Song, "Chance constrained extreme learning machine for nonparametric prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3869–3884, Sep. 2020.

[13] R. Wang, C. Li, W. Fu, and G. Tang, "Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 3814–3827, Oct. 2020.

[14] A. Almutairi and O. Alrumayh, "An intelligent deep learning based prediction model for wind power generation," *Comput. Electr. Eng.*, vol. 101, Jul. 2022, Art. no. 108000.

[15] I. Bazionis, M. Kousounadis-Knudsen, T. Konstantinou, and P. Georgilakis, "A WT-LUBE-PSO-CWC wind power probabilistic forecasting model for prediction interval construction and seasonality analysis," *Energies*, vol. 14, no. 18, p. 5942, Sep. 2021.

[16] F. Liu, Q. Tao, D. Yang, and D. Sidorov, "Bidirectional gated recurrent unit-based lower upper bound estimation method for wind power interval prediction," *IEEE Trans. Artif. Intell.*, vol. 3, no. 3, pp. 461–469, Jun. 2022.

[17] S. D. Dehnavi, A. Shirani, H. Mehrjerdi, M. Baziar, and L. Chen, "New deep learning-based approach for wind turbine output power modeling and forecasting," *IEEE Trans. Ind. Appl.*, early access, Jun. 12, 2020, doi: 10.1109/TIA.2020.3002186.

[18] M. Zhou, B. Wang, S. Guo, and J. Watada, "Multi-objective prediction intervals for wind power forecast based on deep neural networks," *Inf. Sci.*, vol. 550, pp. 207–220, Mar. 2021.

[19] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE Trans. Neural Netw.*, vol. 22, no. 3, pp. 337–346, Mar. 2011.

[20] H. Wen, P. Pinson, J. Ma, J. Gu, and Z. Jin, "Continuous and distribution-free probabilistic wind power forecasting: A conditional normalizing flow approach," *IEEE Trans. Sustain. Energy*, vol. 13, no. 4, pp. 2250–2263, Oct. 2022.

[21] Y. Yu, X. Han, M. Yang, and J. Yang, "Probabilistic prediction of regional wind power based on spatiotemporal quantile regression," *IEEE Trans. Ind. Appl.*, vol. 56, no. 6, pp. 6117–6127, Nov. 2020.

[22] L. Cheng, H. Zang, Y. Xu, Z. Wei, and G. Sun, "Augmented convolutional network for wind power prediction: A new recurrent architecture design with spatial–temporal image inputs," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6981–6993, Oct. 2021.

[23] N. Mararakanye, A. Dalton, and B. Bekker, "Incorporating spatial and temporal correlations to improve aggregation of decentralized day-ahead wind power forecasts," *IEEE Access*, vol. 10, pp. 116182–116195, 2022.

[24] G. Papaefthymiou and P. Pinson, "Modeling of spatial dependence in wind power forecast uncertainty," in *Proc. 10th Int. Conf. Probablistic Methods Appl. to Power Syst.*, May 2008, pp. 1–9.

[25] C. Gilbert, J. Browell, and D. McMillan, "Leveraging turbine-level data for improved probabilistic wind power forecasting," *IEEE Trans. Sustain. Energy*, vol. 11, no. 3, pp. 1152–1160, Jul. 2020.

[26] P. Pinson, "Wind energy: Forecasting challenges for its operational management," *Stat. Sci.*, vol. 28, no. 4, pp. 564–585, Nov. 2013.

[27] H. M. D. Kabir, A. Khosravi, M. A. Hosen, and S. Nahavandi, "Neural network-based uncertainty quantification: A survey of methodologies and applications," *IEEE Access*, vol. 6, pp. 36218–36234, 2018.

[28] S. Mirjalili, H. Faris, and I. Aljarah, "Evolutionary and swarm-based feature selection for imbalanced data classification," in *Evolutionary Machine Learning Techniques: Algorithms and Applications*, 1st ed. Singapore: Springer, 2020, pp. 231–250.

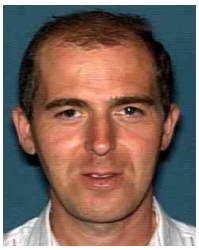
[29] M. Isiet and M. Gadala, "Self-adapting control parameters in particle swarm optimization," *Appl. Soft Comput.*, vol. 83, Oct. 2019, Art. no. 105653.

[30] C. A. C. Coello and M. S. Lechuga, "MOPSO: A proposal for multiple objective particle swarm optimization," in *Proc. Congr. Evol. Comput. (CEC)*, May 2002, pp. 1051–1056.
[31] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 256–279, Jun. 2004.
[32] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *J. Amer. Stat. Assoc.*, vol. 102, no. 477, pp. 359–378, Mar. 2007.
[33] P. Pinson and J. Tastu, "Discussion of 'prediction intervals for short-term wind farm generation forecast' and 'combined nonparametric prediction intervals for wind power generatio,'" *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 1019–1020, Jul. 2014.
[34] A. Tascikaraoglu, "Evaluation of spatio-temporal forecasting methods in various smart city applications," *Renew. Sustain. Energy Rev.*, vol. 82, pp. 424–435, Feb. 2018.
[35] Y. Ding, (Sep. 2021), "Wind time series dataset," Zenodo, doi: 10.5281/zenodo.5516539.
[36] C. Plumley, (Feb. 2022), "Penmanshiel wind farm data," Zenodo, doi: 10.5281/zenodo.5946808.
[37] S. Opricovic and G.-H. Tzeng, "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS," *Eur. J. Oper. Res.*, vol. 156, no. 2, pp. 445–455, Jul. 2004.

MARKOS A. KOUSOUNADIS-KNOUSEN received the Diploma degree in electrical and computer engineering from the National Technical University of Athens (NTUA), Greece, in 2021, where he is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering. His research interests include power systems management optimization, renewable power generation forecasting, and photovoltaic power integration into electric power systems. He is a member of the Technical Chamber of Greece.

IOANNIS K. BAZIONIS received the Diploma degree in geology and geo-environment from the National and Kapodistrian University of Athens, Greece, in 2015, and the M.Sc. degree in energy production and management from the National Technical University of Athens (NTUA), Greece, in 2018, where he is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering. His research interests include wind power forecasting and day-ahead scheduling of power distribution networks in electricity market environment considering the forecasting errors in the production of renewable energy sources.

DIMITRIOS SOUDRIS (Member, IEEE) received the Diploma and Ph.D. degrees in electrical and computer engineering from the University of Patras, Greece, in 1987 and 1992, respectively. Since 1995, he has been a Professor with the Department of Electrical and Computer Engineering, Democritus University of Thrace, Xanthi, Greece, for 13 years. He is currently an Associate Professor with the School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece. He has published more than 500 papers in international journals and conferences. He is the author and an editor of eight books in Kluwer and Springer. His research work has been cited more than 5000 times. He is a leader and a principal investigator in numerous research projects (more than 65) funded by the European Commission, ENIAC-JU, European Space Agency, and the Greek Government and Industry. His research interests include embedded systems design, reconfigurable architectures, network-on-chip architectures, and low-power VLSI design. He is a member of the VLSI Systems and Applications Technical Committee of IEEE CAS and ACM.



FRANCKY CATTHOOR received the Ph.D. degree in EE from Katholieke University Leuven (KU Leuven), Belgium, in 1987. From 1987 to 2000, he headed several research domains in the area of synthesis techniques and architectural methodologies. Since 2000, he has been strongly involving in other activities with IMEC, Leuven, Belgium, including co-exploration of application, computer architecture and deep submicron technology aspects, biomedical systems and IoT sensor nodes, and photo-voltaic modules combined with renewable energy systems. Currently, he is an IMEC Senior Fellow. He is also a part-time Full Professor with the EE Department, KU Leuven. He has been an associate editor of several IEEE and ACM journals.

PAVLOS S. GEORGILAKIS (Senior Member, IEEE) received the Diploma and Ph.D. degrees in electrical and computer engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 1990 and 2000, respectively. In September 2009, he joined as a Faculty Member with the School of Electrical and Computer Engineering, NTUA, where he is currently a Full Professor. From 2004 to 2009, he was an Assistant Professor with the School of Production Engineering and Management, Technical University of Crete, Greece. From 1994 to 2003, he was with Schneider Electric AE, the Greek subsidiary of Schneider Electric, where he was a Quality Control Engineer, a Transformer Design Engineer, the Research and Development Manager, and the Low-Voltage Products Marketing Manager. His current research interests include optimization algorithms and computational intelligence techniques for the optimal operation and planning of smart distribution systems. He is a member of the Technical Chamber of Greece.