ORIGINAL PAPER

Wind Power Scenario Generation Considering Spatiotemporal Correlations: A Distribution Free Hybrid VARMA–Copula Approach

Kailash Chand Sharma[1] **· Archee Gupta**[2] **· Rohit Bhakar**[2]

Abstract

Wind power scenario forecast is a primary step for probabilistic modelling of power systems' operation and planning problems in stochastic programming framework considering uncertainties. Several models have been proposed in the literature to generate wind power scenarios using statistical and machine learning approaches. Most of these models are univariate and do not consider dependency between the Wind Farms (WFs), resulting in generated scenarios lacking spatiotemporal correlation. Furthermore, most scenario generation approaches assume a parametric distribution (i.e., Normal or Weibull distribution) for wind power uncertainty, which is unrealistic. This paper proposes a novel distribution-free hybrid approach that combines multivariate Vector Autoregressive Moving Average (VARMA) and Copula models to generate wind power scenarios. The VARMA model is used to forecast wind power considering spatiotemporal correlation, and then a regular vine (R-Vine) copula is applied on WFs' residuals to retain spatial correlation. The proposed approach is independent of the distribution type because the R-Vine copula can decompose residuals into a copula function and a marginal distribution. The proposed model is illustrated through a realistic case study based on nine Australian WFs. The results obtained are compared with benchmark models that show the efficacy of the proposed model for the generation of wind power scenarios. Minimum energy score, nearly accurate Kendall's correlation, and cross correlation plots show that the proposed method can produce high-quality wind power scenarios without sacrificing spatiotemporal correlation or making distribution assumptions. Generated scenarios using the proposed approach can help WFs and system operators improve decisions in the stochastic programming framework.

Keywords Copula theory · Probabilistic forecasting · Scenario generation · Uncertainty · VARMA model · Wind power forecasting

Introduction

Renewable energy sources, specifically wind and solar, are progressively being integrated into global power systems to prevent the greenhouse gas emissions produced by fossil fuel-based generators while meeting growing demand for power at the lowest possible cost. According to the Global Wind Council, the total global wind power installed capacity is up to 837 GW by 2022, allowing the globe to save nearly 1.2 billion tons of CO_2 yearly [1]. Such massive wind power has the potential to supply global energy demand sustainably while also achieving net-zero objectives by 2050 [1].

Large-scale integration of wind energy in power systems poses considerable challenges for the system operators and wind power producers due to its unpredictable and highly variable nature. Modelling wind power uncertainty into numerous decision-making problems, such as generation scheduling, market clearing, power trading, reserve management, etc. is a formidable challenge [2]. Decision-making problems under wind uncertainty are generally modelled through stochastic programming, robust and chance-constrained programming approaches. The stochastic programming approach is accorded utmost importance because of its accurate representation of wind uncertainty through scenarios. Scenarios are possible sets of random wind power inputs with definite probability [2, 3]. The generation of quality scenarios is essential to model wind power uncertainty in decision-making problems through a stochastic programming approach.

✉ Rohit Bhakar
    rbhakar.ee@mnit.ac.in

1   Department of Electrical Engineering, Dr B R Ambedkar
    National Institute of Technology Jalandhar, Punjab, India

2   Centre for Energy and Environment, Malaviya National
    Institute of Technology Jaipur, Jaipur, India 302017

Several methods have been proposed in the literature to generate wind power scenarios. These are fundamentally categorized as path-based methods, movement matching, and internal sampling. In path-based methods, scenarios are generated by combining forecasted wind power and random error matrices. Machine learning and statistical time-series models are commonly used to predict wind power [2, 3]. Movement-matching methods generate a discrete distribution of statistically dependent random variables by comparing sample distribution with original distributions. Internal sampling is a continuous sampling of the actual distribution of random variables. Due to the use of advanced forecasting methods for scenario generation, path-based methods can accurately represent the stochastic nature of wind power. This paper focuses on expanding the use of path-based concept for wind power generation scenarios considering spatiotemporal correlation between multiple WFs.

An in-depth analysis of wind power scenario generation techniques for efficient use of renewable energy systems is provided [2, 3]. Short-term wind power scenarios are generated for a single WF using multivariate Gaussian distribution-based error covariance matrix [4]. Further, wind speed scenarios are generated for WFs using a univariate Autoregressive Moving Average (ARMA) model and stationary variance–covariance matrix [5, 6]. ARMA model is further translated into the state-space model to generate wind power scenarios and analyze dependencies for multiple WFs [7]. An inverse transform sampling approach was proposed to generate wind power scenarios considering statistical uncertainty and variability [8]. Empirical cumulative distribution functions characterized by uncertainty and variability are modelled by sampling from multivariate normal distributions of forecast errors. The above-discussed scenario generation methods assume parametric distributions for errors and separately modelled temporal and spatial correlations. As a result, significant correlation loss in the generated scenarios occurs due to these assumptions.

A generalized dynamic factor model preserves correlations between generated load and wind power scenarios [9]. However, capturing the temporal correlations through the first or second-order statistics is difficult. Electric load, photovoltaic (PV) and wind production scenarios can be generated using Artificial Neural Network (ANN) based methodology [10]. Data-driven approaches can also generate wind and solar power scenarios using Generative Adversarial Networks (GANs) [11, 12]. In this approach, two deep neural networks are fused as a generator and discriminator. It assumes no parametric distribution of errors but requires another forecasting method to generate the tedious forecasting errors. GANs are modified by imposing Lipschitz constraints on discriminator networks for wind power generation [12]. The GANs based scenario generation approach is further improved by using a conditional improved Wasserstein generative adversarial network (WGAN) [13]. The support vector classifier (SVC) predicts data labels in this approach. The wind power scenarios are generated by

integrated non-separable spatiotemporal covariance function and fluctuation-based clustering [14]. The historical data is grouped into clusters with different fluctuations using the K-means clustering algorithm to estimate the covariance matrix precisely. The machine learning models performance is better than time series models but require a large amount of historical data for learning and have over and underfitting issues. Covariance matrix-based approaches have limitations due to skewness, excess kurtosis and asymmetrical dependencies present in the wind power data. Also, the size of covariance matrix increases with the data dimensions or number of WFs.

Copulas are commonly used to model dependencies among high-dimensional random variables [15–31]. This approach has gained significant attention recently for WFs' dependency and uncertainty modelling. Copulas combine an arbitrary number of univariate distributions into a joint probability density function (PDF), from which useful information can be extracted. The advantageous feature is that it does not impose any restrictions on the data's marginal distribution [16–21]. This property of copula allows it to model asymmetrical dependencies between any numbers of WFs while considering spatiotemporal correlation. A Drawable Vine (D-Vine) copula generates time-coupled wind power infeed scenarios for aggregation of WFs [21]. D-Vine copula was combined with residuals of the support vector regression (SVR) model to represent spatial dependence in the probabilistic forecasts of wind power [22]. A mixture vine copula method that combined the K-means, C-vine, and D-vine copulas is used to analyze the dependence of the multi-wind power output for wind/hydrogen production scheduling. [23]. Regular-Vine (R-Vine) copulas are flexible and have better choices of decomposing dependence structures than D-vine copulas and canonical vine (C-vine) copulas. The R-Vine copula produces better fitting and prediction performance than C-Vine and D-Vine copulas [24–27]. R-Vine copula models, kernel density estimation, time series and hybrid models are used to capture spatiotemporal correlations of multiple WFs for the generation of wind power, wind speed, and market price scenarios. Kernel density estimation, ARIMA, generalized autoregressive conditional heteroskedasticity (GARCH) and ARIMA-GARCH-t models can estimate the marginal distributions for scenario generation [24–28]. A two-stage spatiotemporal sampling approach employs R-Vine copula as a spatial sampling method to generate wind power scenarios for multiple WFs. [29]. R-Vine copula is further combined with a variance reduction method to generate time-coupled wind power scenarios. [30]. R-Vine copula is also used in conjunction with the random forest method to identify load patterns in smart grids. [31].

In the above-mentioned methods, R-Vine copula-based scenario generation methods used a univariate or multivariate distribution to model temporal correlations and copula for spatial correlation. However, the simultaneous modeling of temporal and spatial correlations through the multivariate forecasting models and further recovery of spatial correlation

from the residuals using the R-Vine copula model can generate quality wind power scenarios retaining spatiotemporal correlations. Table 1 highlights the summary of the literature review on scenario generation and provides comparison between the existing work and the proposed work. The models that use the vine copula approach for scenario generation are highlighted in grey in this table. The table clearly shows that the advanced multivariate time series model, *i.e.*, VARMA, is not a hybrid with the R-Vine copula for generating wind power scenarios that consider spatial and temporal correlation simultaneously. The combination of VARMA and R-Vine copula can provide two-stage modeling for spatial correlation, which improves the quality of generated scenarios. The wind power scenario generation method can be further improved by incorporating the R-Vine copula and the multivariate time series forecasting model, which capture the asymmetrical tail dependency that occurs in wind generation without making any assumptions about distribution types.

This paper proposes a hybrid, distribution-free VARMA-Copula approach for generating wind power scenarios for multiple WFs with spatiotemporal correlations in the very short-term horizon. For spatiotemporal correlations, the multivariate VARMA model is used to generate marginal distributions of WFs residuals. The R-vine copula then uses the residuals to capture the asymmetrical dependencies between WFs. Modelling marginal distributions through a multivariate model can simultaneously consider the spatial and temporal propagation and forecast errors. The use of the R-vine copula makes the proposed model distribution-free and better compared to other benchmark models. The proposed algorithm is implemented on publicly available data of nine Australian WFs. The obtained results are compared with benchmark models such as VAR, VARMA, ANN [10] and GANs [11]. The proposed model retains the spatiotemporal correlation in the generated scenarios, as demonstrated by the minimum energy score, cross-correlation plots, and Kendall's correlation plots.

The rest of the paper is organized as follows. Section II describes the mathematical formulation of VARMA and R-Vine copula to model spatiotemporal correlations. The proposed scenario generation algorithm is discussed in Section II. Section III describes energy score and cross-correlation functions to evaluate the quality of generated scenarios using proposed and benchmark models. Results are discussed in the case study section IV. Finally, the proposed work has been concluded in section V.

Table 1 Summary of literature review on scenario generation

Ref	Authors	Model	Scenarios	Model type	Distribution	Temporal Correlation	Spatial Correlation
[27]	G. Giovani *et. al*	GARCH + R-Vine	Wind speed	Univariate	Independent	✓	✓
[5]	J. Morales *et.al*	ARMA + stationary covariance matrix	Wind speed	Hybrid	Gaussian	✓	✓
[7]	G. Diaz *et. al*	ARMA + State-space	Wind speed	Hybrid	Gaussian	✓	✓
[4]	P. Pinson *et.al*	Covariance matrix	Wind power	Multivariate	Gaussian	✓	✗
[6]	K.C. Sharma *et. al*	ARMA	Wind power	Univariate	Gaussian	✓	✗
[25]	Z. Wang *et. al*	ARIMA + R-Vine copula	Wind power	Univariate	Independent	✓	✓
[28]	Q. Tu *et. al*	ARIMA-GARCH-t + R-Vine Copula	Wind Power	Univariate	Independent	✓	✓
[8]	X. Y. Ma *et. al*	Inverse transform sampling	Wind Power	Multivariate	Empirical + t-location scale + Normal	✓	✓
[12]	C. Jiang *et. al*	Modified GANs	Wind Power	Multivariate	Gaussian	✓✓	✓✓
[13]	Y. Zhang *et. al*	WGAN	Wind Power	Multivariate	Independent	✓✓	✓✓
[14]	J. Tan *et. al*	Covariance matrix	Wind Power	Multivariate	Independent	✓	✓
[22]	M. A. Prieto *et. al*	SVR + D-Vine Copula	Wind Power	Multivariate	Independent	✓	✓
[29]	*Y. Li* et. Al	R-Vine + Sampling	Wind Power	Multivariate	Independent	✓	✓
[30]	*A. B. Krishna* et.al	R-Vine + Variance reduction	Wind Power	Multivariate	Independent	✓	✓
[9]	D. Lee *et. al*	Generalized dynamic factor + VAR	Load & Wind Power	Multivariate	Gaussian	✓✓	✓✓
[10]	S. Vagropoulos *et. al*	ANN	Load + PV + Wind power	Multivariate	Gaussian	✓✓	✓✓
[11]	Y. Chen *et. al*	GANs	Wind + PV Power	Multivariate	Gaussian	✓✓	✓✓
[31]	M. Fang *et. al*	R-Vine + Random Forest	Load	Multivariate	Independent	✓	✓
Proposed work		VARMA + R-Vine Copula	Wind power	Multivariate	Independent	✓✓	✓✓

✓ Considered ✗ Not considered ✓✓ simultaneously considered.

Spatiotemporal Correlation Modeling in Scenario Generation

VARMA Model

The power output of N-WFs is considered as stochastic process and modelled through the multivariate n-dimensional $VARMA(p, q)$ model. Compared to the VAR model, the VARMA model can forecast wind power output using temporal propagation of historical observations and forecast errors [32–34]. The number of historical observations and errors used in the VARMA process depends on the estimated order of VA terms p and MA terms q. The spatial correlation between WFs is modelled using the VARMA model parameter matrix. The N-dimensional $VARMA(p, q)$ is mathematically expressed as follows:

$$\hat{y}_t^n = \sum_{l=1}^p \phi_l y_{t-1} + a_t + \sum_{m=1}^q \Theta_m a_{t-m} \tag{1}$$

where, $\hat{y}_t^n$ is forecasted conditional mean power output of nth WF at time t. ϕ_l and Θ_m are $N \times N$ AR and MA parameter matrices for l and m lags, respectively. The size of these matrices depends on the number of WFs and the order of the fitted VARMA model. a_t denotes N-dimensional white-noise vector at a time t with zero mean and $N \times N$ non-singular contemporaneous covariance matrix cov(a). The Extended Cross-Correlation Matrix and Maximum likelihood approach are used to estimate order and parameters of the VARMA model, respectively [34, 35]. The estimated model parameters are locally optimal. The estimated order of the VARMA model is further verified by different information criterions such as Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC) and Hannan–Quinn information criterion (HQC). The minimum value of these criteria is desired for the selected order of the VARMA model [34, 35].

Under the following statistical assumptions, the VARMA model can be used directly to generate wind power scenarios for multiple WFs while accounting for spatiotemporal correlations.

- Historical wind power data input is assumed to be stationary and linear. Thus, input data is preprocessed to make it stationary using either differentiation and/or logarithm transformation.
- Furthermore, input wind power data is assumed to follow parametric distributions such as Gaussian or Weibull distributions with symmetrical tails.
- The fitting of the VARMA model to time series data involves epistemic uncertainty regarding model order and parameter estimations.
- The VARMA model parameters are assumed to be time-independent. The consideration of time-varying model parameters necessitates adaptive models for parameter estimation, which is beyond the scope of the current paper.

Due to the aforementioned statistical assumptions, VARMA model-based scenario generation algorithms will fail to retain the spatial correlation between the WFs. The reasons for this are as follows:

1. VARMA models primarily focus on capturing the temporal dynamics within a time series and do not inherently account for the spatial interdependencies between different WFs. Spatial correlations in wind power data arise from factors such as wind patterns, geographical proximity, and shared weather conditions. By not explicitly incorporating these spatial dynamics, VARMA models alone cannot adequately capture or retain the spatial correlations.

2. VARMA models often make assumptions about the statistical properties of the data, such as stationarity, linearity, and parametric distributions with symmetrical tails. However, wind power data often exhibits non-Gaussian characteristics and non-stationary behavior, which are not accurately captured by these assumptions. The oversimplification of the underlying statistical properties can lead to the failure of VARMA models to capture the true spatial correlations.

3. VARMA models require decisions regarding the model order selection and parameter estimation, both of which introduce uncertainty. The choice of model order and estimation of parameters have a significant impact on the model's ability to capture spatial correlations accurately. Incorrect model order selection or inaccurate parameter estimation can result in the omission or misrepresentation of spatial dependencies, leading to a failure in retaining spatial correlations.

Based on the reasons discussed above, the VARMA model-based scenario generation algorithms alone are unable to adequately retain spatial correlations in wind power data. To address this limitation, the integration of copula models, such as the regular vine (R-Vine) copula, with the VARMA model is proposed in this paper. VARMA and copula models are hybridized through residuals. Residuals of the VARMA model are inputs for the copula Model. The copula model captures the residual correlations between WFs, thus improving the representation of spatial correlations in the generated wind power scenarios. The following equation can be used to calculate residuals:

$$u_t^n = y_t^n - \hat{y}_t^n \tag{2}$$

Mathematically, residuals u_t^n are the difference between fitted or forecasted values $\hat{y}_t^n$ and observed values y_t^n of the nth WF's power outputs. Uncorrelated residuals signify the best fitting of

the VARMA model. But if residuals are correlated, means that VARMA model has not fully captured spatiotemporal correlation. The dependency between the residuals of WFs is modelled using the R-Vine copula model in this paper. Copula models the dependency in terms of marginal and copula probability density functions separately which is free from an assumption of a specific distribution.

R-vine Copula Model

The primary function of a copula in mathematics is to capture and quantify the dependency or correlation between variables. It provides a way to study and analyze the relationship between random variables separately from their individual distributions. For example, random variables in the power system such as wind speed, wind power, load, and solar generation are following the different probability distributions like Gaussian, Weibull, Gamma, Beta etc. Therefore, modeling of correlation between such variables is difficult by the conventional correlation measures. By using copulas, it becomes possible to model and simulate multivariate data while preserving the marginal distributions (not having correlation) and capturing the correlation structure. Copula has widely used for the following purpose in the literature:

1. Copulas allow for the modeling of various types of dependence structures between random variables, such as positive or negative correlation, tail dependence, and rank correlation.
2. Copulas can be used to construct multivariate distributions by combining specified marginal distributions with a copula function that represents the dependence structure.
3. Copulas provide a framework for estimating and testing the dependence structure between variables based on observed data. They enable the estimation of copula parameters and the assessment of goodness-of-fit.

Mathematically, copula is a flexible approach to model multivariate distributions for different univariate time series. As per Sklar's theorem, the multivariate cumulative distribution function (CDF) can be expressed as a function of marginal distributions of each variable and copula that defines the dependency between the variables [18, 27]. Let $u_1, u_2 ...... u_n$ are estimated residuals of N-WFs at time t and $F_1, F_2 ...... F_n$ are corresponding marginal CDFs. According to Sklar's theorem, the joint distribution can be mathematically represented as follows:

$$F\left(u_1, u_2, ..., u_n\right) = C\left\{F_1\left(u_1\right), F_2\left(u_2\right), ..., F_n\left(u_n\right)\right\} \quad (3)$$

where, C is n-dimensional copula and a joint CDF of n-dimensional residuals with uniform marginals. Similarly,

the joint probability density function (PDF) for residuals of n WFs can be expressed as follows:

$$f\left(u_1, u_2, ..., u_n\right) = c\left\{F_1\left(u_1\right), F_2\left(u_2\right), ..., F_n\left(u_n\right)\right\} \cdot f_1\left(u_1\right), f_2\left(u_2\right), ..., f_n\left(u_n\right) \quad (4)$$

where, c is the copula PDF and f_n represents the marginal PDF of n residuals. The parametric normal distribution is used in this paper to estimate marginals' PDF of residuals. Several classes of copula functions are available in the existing literature. These are broadly classified as Elliptical and Archimedean copula [25–28]. Gaussian copulas and t-copulas are standard multivariate Elliptical copulas used to model data has symmetrical tail dependence. While Clayton, Gumbel, and Joe copula are bivariate Archimedean copulas that can capture asymmetrical tail dependence *i.e.* common for wind power and wind speed distributions. Elliptical copulas can model high dimensional distributions with symmetric tail dependence while Archimedean copulas are limited bidimensional distributions with asymmetric tail dependence.

The efficient modelling of high dimensional distribution requires a copula model which can extract all variety of tail dependence without restricting to two dimensions. So, vine copula also named pair copula construction (PCC) emerged as a powerful tool to model the high dependence structure in form of bivariate copulas. This extensibility enhances the performance of the vine copula in exhibiting the arbitrary dimensional structure better than the elliptical copula. Vine copula is more efficient in capturing the wind power data structure which possesses different characteristics of tail dependence. PCCs provide a way to explore the excellence of bivariate copula by extending it to variable dimensions.

The multivariate N-dimensional copula is shown in Eq. (3) and (4) are decomposed into $N(N − 1)/2$ bivariate or pair copula as a Vine copula. The graphical representation of decomposed copula is called the R-Vine copula. In the R-Vine copula, the relationship between pair copula is depicted by a set of N-1 trees. Each tree T_i consists of N_i number of nodes and E_i edges. The first or initial tree has several nodes equal to the number of WFs. Each tree should satisfy the essential properties as described in [24, 27].

1. Initial tree T_1 with the node $N_1 = \{1, 2, ...n\}$ and edge E_1.
2. For $i = 2, ..., n − 1$ tree T_i has the node $N_i = E_{i−1}$.
3. Proximity condition: If two nodes are joined by an edge in a tree T_i, their corresponding edges in the tree $T_{i−1}$ must share a common node.

Further, an edge E_i is denoted by $e = \{j(e), k(e)|D(e)\}, \forall n, j \neq k$ where $j(e)$ and $k(e)$ are conditional nodes and $D(e)$ is a conditioning set associated with an edge e. The elements

of these sets are all nodes. The conditioning set $D(e)$ of the initial tree is empty [27]. The density function of the bivariate copula associated with each edge is represented by $c_{j(e),k(e)|D(e)}$.

The joint density function of an n-dimensional residual vector as shown in Eq. (4), can be expressed as a product of the marginal density functions and bivariate copula density functions shown in Eq. (5)-(7):

$$f(u_1, u_2, .., u_n) = c\{u|v, B, \alpha\}.f_1(u_1), f_2(u_2), .., f_n(u_n) \quad (5)$$

$$f(u_1, u_2, ..., u_n) = c\{u|v, B, \alpha\} \times \left[\prod_{k=1}^{n} f_k(u_k)\right] \quad (6)$$

$$c\{u|v, B, \alpha\} = \prod_{i=1}^{n} \prod_{e \in E_i} c_{j(e),k(e)|D(e)} \begin{pmatrix} F_{j(e),D(e)}(u_{j(e)}|u_{D(e)}), \\ F_{k(e),D(e)}(u_{k(e)}|u_{D(e)}) \end{pmatrix} \quad (7)$$

where, $u_{D(e)} = \{u_i | i \in D(e)\} c(x|v, B, \alpha)$ is the R-vine copula density function, v is the R-vine structure, B is the set of bivariate copulas associated with the R-vine, and α is the parameter vector for the bivariate copulas. The sequential approach is used in this paper to select the R-vine structure [24, 27]. In this approach, the trees are selected in such a way that the chosen pairs model the strongest pairwise dependencies present. The algorithm for the sequential approach is described in the Algorithm 2. The parameters of each bivariate copula are estimated by Maximum Likelihood Estimation [27].

Proposed Scenario Generation Algorithm

R-Vine Copula and VARMA models are used in this paper to generate wind power scenarios considering spatiotemporal correlations. The proposed scenario approach is summarized in Algorithm 1. The first VARMA model is fitted on collected historical wind power data of N WFs using Eq. (1) then standard residuals are formed using Eq. (2). The residuals are separated for 24-steps ahead. Before being fed into the VARMA model, WF's historic data is normalized based on installed capacity. To check stationary in the historical power output data of WFs, the Augmented Dickey-Fuller test is used [35]. If data is found to be non-stationary, then differencing and log transformation are applied to make data stationary. VARMA model is fitted on stationary data time series and model parameters and residuals are estimated.

Input: Historical power outputs of N-WFs
Output: Wind power scenarios for N-WFs

1. Data Stationary test and transformation
     if (Data = = Stationary)
     Go to the next step
     else
     Data transformation
     end if

2. VARMA model fitting: Estimate order and parameters
 for t = 1:5:120
 3. Forecast 24-step ahead wind power output of N-WFs considering spatiotemporal correlation using Eq. (1).
 4. Estimate residuals of N-WFs using Eq. (2) for randomly generated white noises.
 end for
 Initialize time counter t =1
 while (t ≤24)
 5. Prepare copula data by decomposing residuals into uniform marginals [0,1] $^{N.}$
 6. Select R-copula with minimum AIC and estimate copula parameters using the sequential method presented in Algorithm 2.
 7. Estimate joint copula pdf for N-WFs using Eq. (6)
 8. Generate wind power scenarios using forecasted WFs outputs and sampling from joint PDF.
 update time counter t =t+1
 end while
 9. Evaluate the quality of generated scenarios using energy score and cross-correlation plots.

Algorithm 1 VARMA-Copula based algorithm for generation of wind power scenarios

Due to the normalized data, the residuals will range from -1 to + 1. As a result, the unit hypercube constraint $[0, 1]^N$ is violated. Therefore, in step 5, residuals must be converted into copula data. After the joint PDF has been constructed, the copula data is transformed back into the original data using Probability Integral Transformation in Step 8. Once the desired number of scenarios are generated, their quality check has been done using the energy score and spatiotemporal correlation are depicted by the cross-correlation plot, describes in section III. An energy score with the minimum value is the required result to prove the superiority of the model. Algorithm 2 presents the sequential method for R-Vine copula selection and parameter estimation. This algorithm's input is the residuals of the WFs under consideration, and its output is the joint PDF. The sequential method captures most of the dependence in lower-level trees while leaving no or minimal dependence in higher-level trees.

(3) homoscedasticity. However, for the case of wind speed and power time series, meeting these assumptions is challenging due to the following characteristics of wind energy: (1) a nonlinear relationship between wind power and wind speed in the wind turbine power curves; (2) wind power follows an approximately Weibull distribution instead of a normal distribution; and (3) wind power time series are typically characterized by heteroscedasticity rather than homoscedasticity. Consequently, using Pearson's linear correlation to measure the correlation between two nonlinear wind power time series may result in errors. In contrast, Kendall correlation does not rely on the above preconditions, making it suitable for use in the proposed research to analyze the correlation between multiple WFs.

First, the empirical Kendall's tau is computed for every possible variable pair combination. The total number of combinations for each time step is the square of the number of WFs. Following that, the sum of absolute empirical

---

Input: Residuals of N wind farms ($u_1, u_2 ......u_n$)

Output: Joint PDF using R-vine copula

1. Calculate the empirical Kendall's tau $\hat{\tau}_{j(e),k(e)}$ for all possible variable pairs $\left(u_{j(e)}, u_{k(e)}\right), 1 \leq j \neq k \leq n$.

2. For T_1 choose the spinning tree that maximize the sum of absolute empirical Kendall's tau.

$$\max \sum_{E_1 = \{j(e),k(e)\}} \left|\hat{\tau}_{j(e),k(e)}\right|$$

3. For an edge $E_1 = \left\{j(e), k(e)\right\}$ in the selected tree T_1, choose and estimate bivariate copula that minimize the AIC.

Initialize tree counter $i = 2$

While ($i \leq n - 1$)

4. For i^{th} tree T_i, calculate the empirical Kendall's tau $\hat{\tau}_{j(e),k(e)|D(e)}$ for all possible variable pairs $\left(u_{j(e)} | u_{D(e)}, u_{k(e)} | u_{D(e)}\right)$.

5. Select the tree that satisfy the proximity condition and maximize the sum of absolute empirical Kendall's tau.

$$\max \sum_{E_i = \{j(e),k(e);D(e)\}} \left|\hat{\tau}_{j(e),k(e)|D(e)}\right|$$

6. For an edge in the selected tree T_i, choose and estimate bivariate copula that minimize the AIC.

Update tree counter $i = i + 1$

End While

Algorithm 2 Sequential method for R-Vine copula selection and estimation [27]

Pearson and Kendall are two main correlation measures that have been used in the existing literature to demonstrate the correlation between two variables, namely electricity prices, renewable generation, and loads [36, 37]. Kendall's tau is a non-parametric measure, meaning it does not assume any specific distribution for the variables. On the other hand, Pearson is a linear correlation measure, and it relies on three key assumptions: (1) linearity; (2) normality of variables; and

Kendall's taus is maximized to form the initial tree. To minimize AIC, a bivariate copula is chosen from a predefined set of copula families for each edge E_1 of the initial tree. For other than initial tree edge E_i is denoted by $E_i = \{j(e), k(e) | D(e)\}$, where $j(e)$ and $k(e)$ are conditional nodes and $D(e)$ is the conditional set associated with edge e.

The R-vine is a flexible copula that constructs a tree-like dependence structure using a sequential modelling

Table 2 WFs id and installed capacity [40]

WFs	AEMO ID	Installed capacity	Latitude	Longitude
WF1	MTMILLAR	70	-30.85	136.58
WF2	STARHLWF	34.5	-35.57	138.16
WF3	LKBONNY2	159	-37.56	140.27
WF4	LKBONNY1	80.5	-37.76	140.40
WF5	WPWF	90.75	-35.11	137.72
WF6	HALLWF1	94.5	-33.37	138.73
WF7	HALLWF2	71.4	-33.52	138.87
WF8	CLEMGPWF	57	-33.50	138.10
WF9	BLUFF1	52.5	-33.35	138.80

approach. This sequential modelling enables a more flexible representation of dependency patterns. Each level of the vine structure can have a different bivariate copula family, allowing for the modelling of a wide range of dependencies. Different bivariate copula families can be used to model the dependencies at each level of the R-vine copula. The bivariate copula family includes the Gaussian, Student's t, Gumbel, Clayton, Frank, Joe, Tawn, independent copulas, and their hybrid. Copulas with rotations of 90, 180, and 270 degrees are modified versions of the base or survival copula that capture asymmetrical tail and negative dependence. More information on vine copulas, including their CDF equations and parameter ranges, can be found in [27, 38]. Finally, parameters for selected bivariate copulas are estimated for each tree edge. Because the presented paper focuses on multistep scenario forecasting, copula parameters are calculated for each time step separately rather than for the entire time period. The procedure is repeated for the remaining tree set.

Evaluation of Scenarios

The quality of scenarios is of utmost importance as they impact the decision-making in stochastic programming. Various scenario quality assessment methods such as reliability diagrams, sharpness, skill score and energy score are discussed [39]. Most of these quality assessment methods are limited to univariate probabilistic marginal distribution and overlooking the temporal independence structure. Overcoming the drawback, the various multivariate verification tools can be used. In this paper, an energy skill score is used to evaluate the quality of generated scenarios.

Energy Score:

The energy score is a proper scoring rule for verification of multivariate scenario forecasting. It is used in the evaluation of time-coupled power infeed scenarios [39]. It is a negatively oriented score, *i.e.* lower the better. Providing the real observations of wind power data y and a matrix of corresponding generated scenarios y_s, the Energy score can be calculated as:
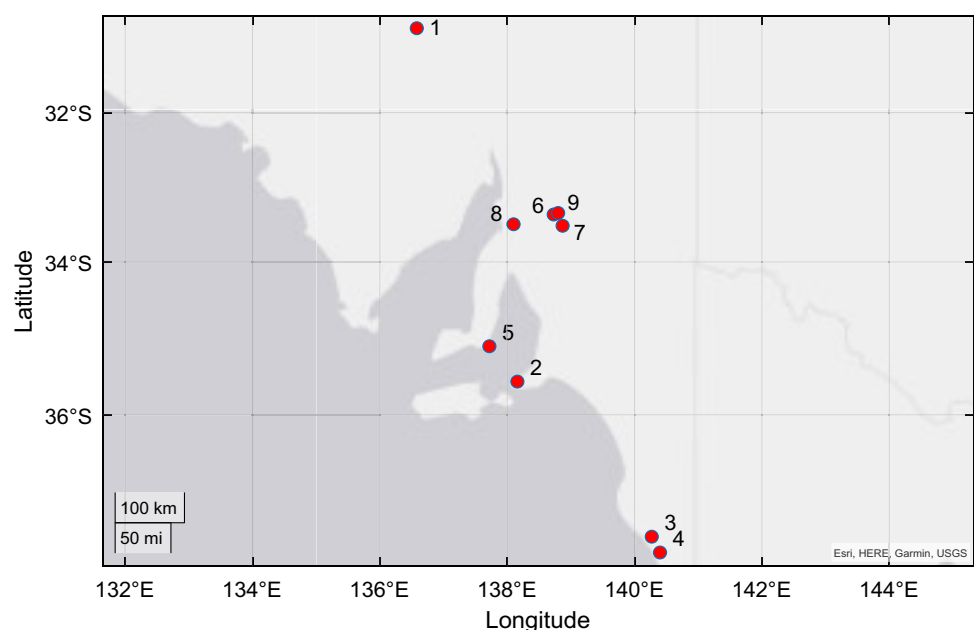
Fig. 1 Location of WFs [40]

Table 3 Estimated AR and MA Coefficients for three representative WFs

WFs	AR Coefficients			MA Coefficients		
	WF2	WF5	WF8	WF2	WF5	WF8
WF2	0.9854	0.00087	0.000417	0.0895	-0.0130	0.0144
WF5	-0.0018	0.99261	0.002621	-0.0080	0.1768	0.0056
WF8	-0.0035	0.00082	0.941704	0.0096	-0.0030	-0.0790

$$ES(Y_s, y) = \frac{1}{m} \sum_{i=1}^{m} \|y_{s,i} - y\| - \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \|y_{s,i} - y_{s,j}\| \quad (8)$$

where, m is the number of observations and $\|.\|$ is the Euclidean norm.

Cross-Correlation Function

The cross-correlation function is a way to find the degree of dependency between two-time series in different times and spaces. It shows how one-time series behave when other shifts relative to time and space. Let sample scenario time series for i^{th} and j^{th} WFs are X_i and, X_j respectively. Both time series consists of m number of observations.

The cross-correlation $\rho_{i,j}$ between X_i and X_j is defined as the ratio of covariance to root-mean variance [34, 35]. The cross-correlation is mathematically expressed as follows:

$$\rho_{i,j} = \frac{\hat{\gamma}_{i,j}}{\sqrt{\sigma_i^2 \sigma_j^2}} \quad (9)$$

where, $\hat{\gamma}_{i,j}$ is sample covariance, σ_i^2 is the variance of time series X_i and σ_j^2 is the variance of time series X_j. The cross-covariance of time series X_i and X_j with some lag l is given by using Eq. (10).

$$\hat{\gamma}_{i,j} = \frac{1}{m} \sum_{t=1}^{m-l} \left[\left(X_i^t - \overline{X}_i \right) \left(X_j^{t+l} - \overline{X}_j \right) \right] \quad (10)$$
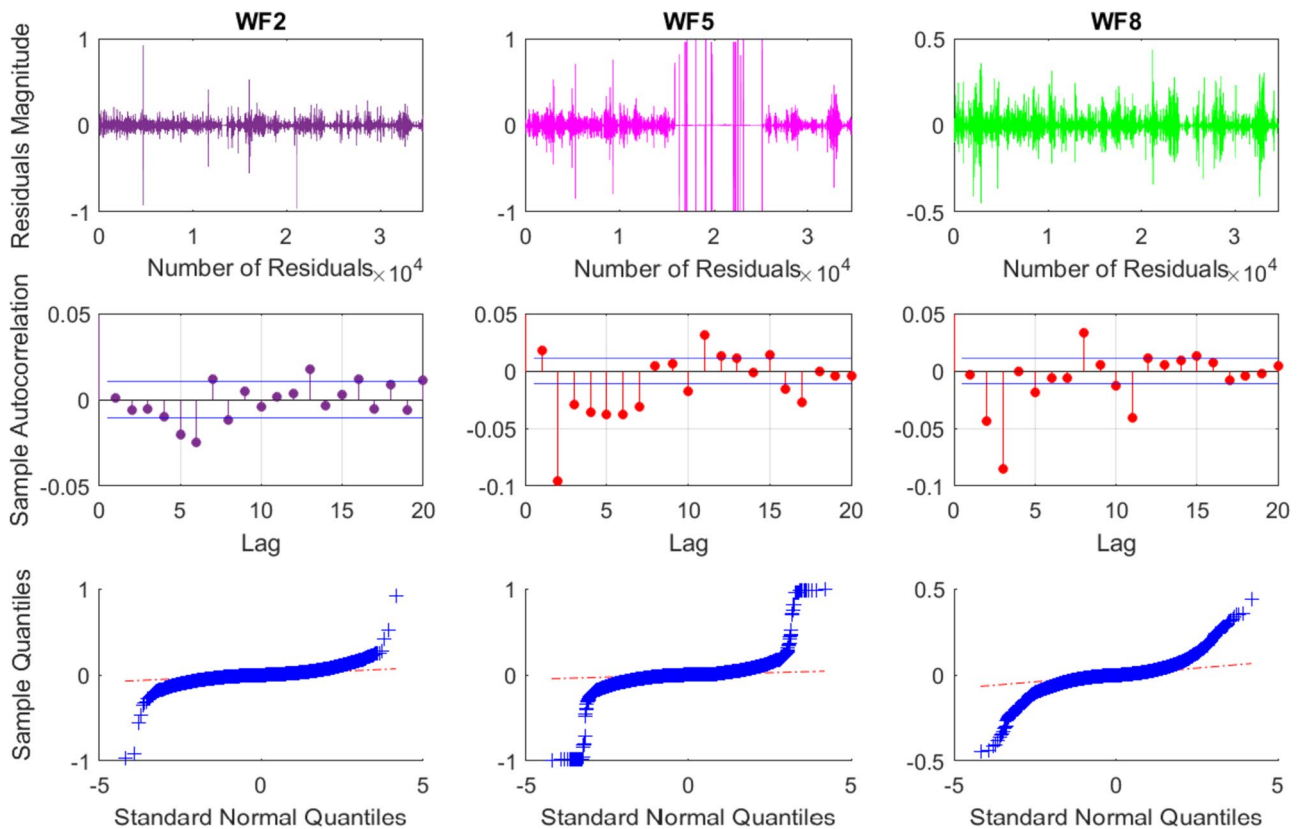

Fig. 2 Residuals, sample autocorrelation and QQ plot of three WFs for first time horizon

where, $\overline{x}_i$ and $\overline{x}_i$ are mean of X_i and X_j time series, respectively. In this paper cross-correlation function approach is used to validate the spatiotemporal correlation between different wind power time series.

Case Study

Dataset

Proposed model efficacy is validated through application on real time nearby WF datasets. It is implemented on a publicly available wind power data sets which is obtained from Australian WFs [33, 40]. Nine WFs from the four regions are selected to generate wind power scenarios. Table 2 shows the Australian Energy Market Operator Identity (AEMO ID) given to different WFs with installed capacity, latitude, and longitude. The historical wind power data of each WF is normalized according to its installed capacity for input in the scenario generation models. Figure 1 shows the location of nine WFs. The resolution of data is 5 min and 24-step ahead scenarios are generated. Among available data, 34561 data points at a resolution of 5-min are used for the training purpose.

Section of Benchmark Models

Among various time series univariate ARIMA, multivariate VAR, and VARMA models are considered benchmark models. ARIMA is a popular time series univariate model used in a variety of applications including wind power forecasting [41, 42]. VAR is used to emphasize the superiority of multivariate time series model over the univariate model. VAR is a subset of the VARMA model that does not include the moving average component [32, 33]. VARMA is used as a benchmark to demonstrate the significance of the proposed hybrid VARMA Copula model over the standard VARMA model. Machine learning models may be advantageous over time series models, in capturing the nonlinear dynamics of wind power generation. However, the performance of these models relies on careful input feature selection and hyper parameters tuning. In this paper, the ANN [10] and GAN [11, 12] models are chosen as benchmark models. ANN is suitable for prediction with labeled input datasets, while GAN excels at generating realistic data samples, such as wind power scenarios. The GAN model consists of a generator and a discriminator, and during training, both networks improve their performance through an adversarial process. GAN can produce distinct wind power scenarios that capture the intrinsic features of historical wind power data, such as ramps and spikes. Both ANN and GAN models can generate wind power scenarios considering
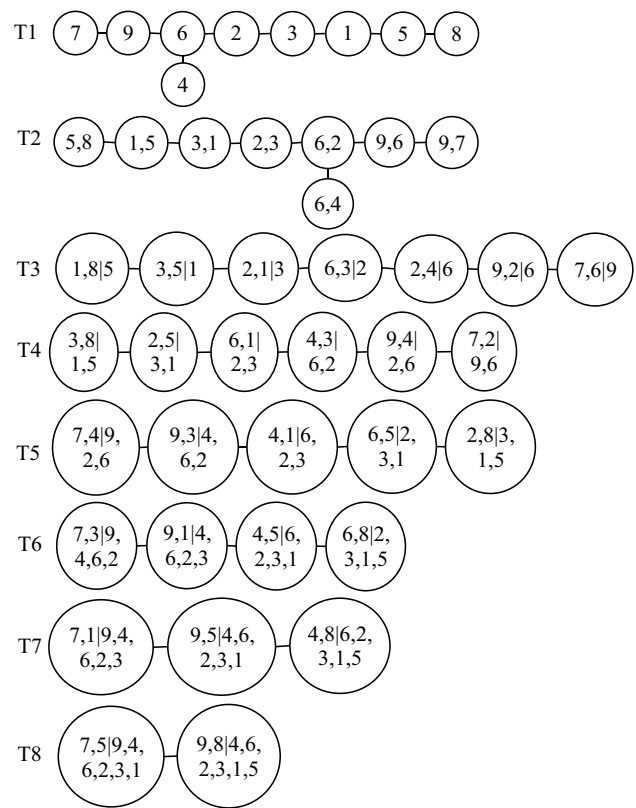

Fig. 3 Tree structure of bivariate Copulas for first time horizon

spatiotemporal correlations [10–12]. The detailed comparison of all scenario generation models is beyond the scope of the presented paper.

Simulation Platforms and Packages

All simulations have been performed on the R-studio platform version 4.1.2. The Multivariate Time Series (MTS) version 1.2.1 and Vine Copula version 2.4.5 packages are used for modelling VARMA and R-Vine copula, respectively [35, 43]. Other than these main packages, readxl, writexl, vars, forecasts, copula and RSNNS packages are used for simulation of basic functions and benchmark models. The reference manuals for these packages can be found in the Comprehensive R Archive Network (CRAN) repository [44]. All of the code and results described in this paper are publicly available at https://zenodo.org/deposit/8017860 and https://zenodo.org/deposit/8017980.

VARMA Results

This section describes the results obtained using the VARMA model. In the VARMA model, parameters are estimated through maximum likelihood approach. For reducing

Table 4 Bivariate copula selection for R-Vine structure of first time horizon

Tree	Edge E_i	Type of copula	Copula parameters		$\hat{\tau}$	Total number of bivariate copulas
			α_1	α_2		
1	5,8	BB8	6.00	0.87	0.64	8
	1,5	Rotated BB8 90 degrees	-6.00	-0.89	-0.66	
	3,1	Frank	7.17	-	0.57	
	2,3	BB8	5.59	0.7	0.50	
	6,4	Frank	-9.01	-	-0.64	
	6,2	Frank	8.49	-	0.62	
	9,6	Frank	10.23	-	0.67	
	9,7	Survival BB8	6.00	0.75	0.56	
2	1,8|5	Rotated Tawn type 1 90 degrees	-1.11	0.30	-0.05	7
	3,5|1	Frank	-3.46	-	-0.35	
	2,1|3	Survival BB8	1.19	0.99	0.09	
	6,3|2	Survival BB8	1.85	0.94	0.26	
	2,4|6	Rotated BB8 270 degrees	-6.00	-0.39	-0.28	
	9,2|6	BB8	1.47	0.99	0.20	
	7,6|9	Student's t	0.29	7.14	0.19	
3	3,8|1,5	Rotated BB8 270 degrees	-6.00	-0.30	-0.20	6
	2,5|3,1	Frank	-4.65	-	-0.43	
	6,1|2,3	Survival BB8	1.80	0.96	0.27	
	4,3|6,2	Rotated BB8 270 degrees	-6.00	-0.40	-0.28	
	9,4|2,6	Rotated BB8 90 degrees	-6.00	-0.43	-0.31	
	7,2|9,6	BB8	1.92	0.91	0.25	
4	2,8|3,1,5	Frank	-3.08	-	-0.31	5
	6,5|2,3,1	Frank	-2.47	-	-0.26	
	4,1|6,2,3	Rotated BB8 90 degrees	-2.68	-0.79	-0.30	
	9,3|4,6,2	Gaussian	0.14	-	0.09	
	7,4|9,2,6	Independence	-	-	-	
5	6,8|2,3,1,5	Rotated BB8 90 degrees	-1.90	-0.80	-0.18	4
	4,5|6,2,3,1	Student's t	0.19	9.49	0.12	
	9,1|4,6,2,3	BB8	2.57	0.77	0.27	
	7,3|9,4,6,2	BB8	1.53	0.91	0.16	
6	4,8|6,2,3,1,5	Frank	1.68	-	0.18	3
	9,5|4,6,2,3,1	Survival Clayton	0.04	-	0.02	
	7,1|9,4,6,2,3	Independence	-	-	-	
7	9,8|4,6,2,3,1,5	Independence	-	-	-	2
	7,5|9,4,6,2,3,1	Rotated Tawn type 2 270 degrees	-17.54	0	0	
8	7,8|9,4,6,2,3,1,5	Independence	-	-	-	1

computational efforts, here (1, 1) order of VARMA is used. Total of 171 parameters are estimated for VARMA (1, 1) model. These include 81 AR, 81 MA and 09 standard variance parameters. For simplicity, three representative WFs (WF2, WF5 and WF8) are selected from each zone and results for these three WFs are discussed in detail. Table 3 shows the estimated AR and MA coefficients for three representative WFs. After parameter computation, obtained residual covariance matrix is used to generate a normally distributed white noise for each WF separately.

Figure 2 depicted the diagnosis of VARMA model residuals for three denoted WFs. Because of the normalization of the input data, the residuals range from -1 to 1. The obtained residuals appear to be a noise signal with many random spikes. This implies that the obtained results should be scrutinized further to assess the quality of the VARMA model. For the residuals of the VARMA, a sample autocorrelation function (ACF) and a quantile–quantile (QQ) plot are drawn. ACF plots demonstrate the presence of correlation in the sample data with different time lags.

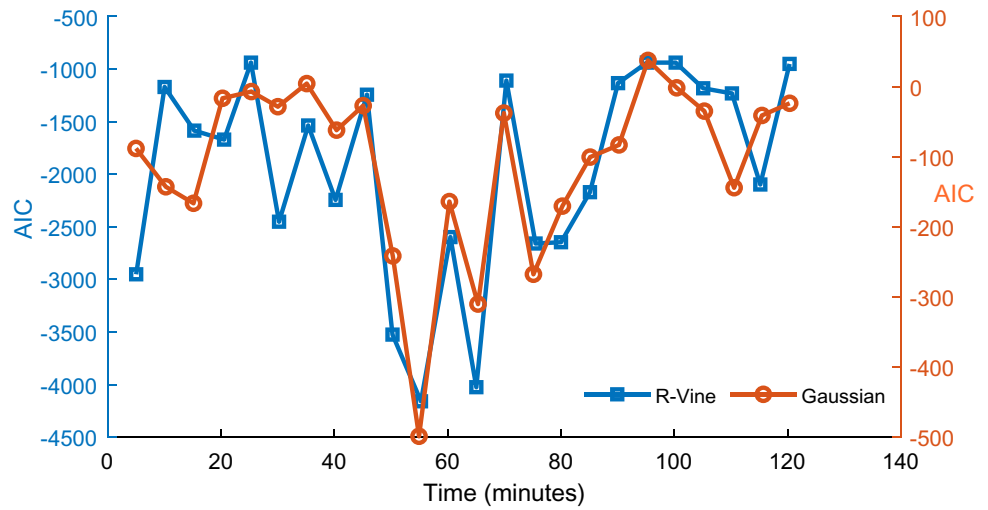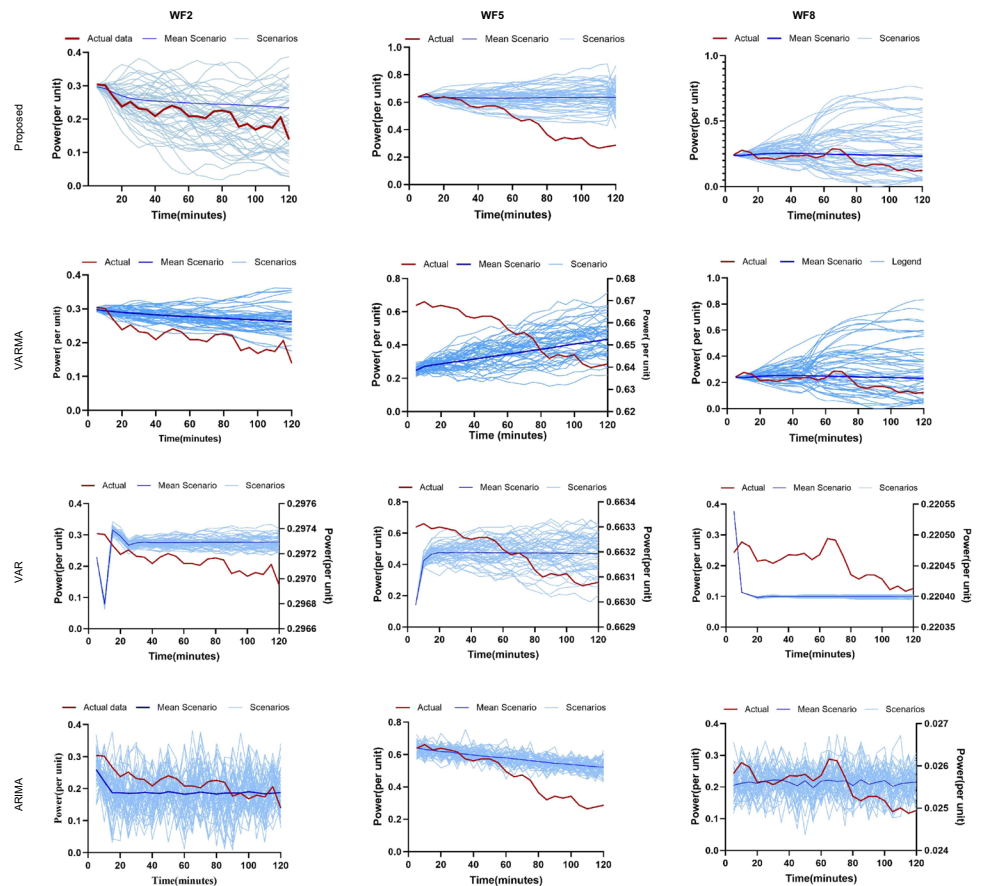**Fig. 4** Performance of R-vine and Gaussian copulas

Fig. 5 Generated scenarios for WF2, WF5 and WF8 using proposed, VARMA, VAR and ARIMA approaches. In the dual-axis sub-plots, the right y-axis depicts the generated wind power and mean scenarios, measured in per unit. "Per unit" refers to the ratio of the power output of WFs to their installed capacities in MW

Ideally, residuals should be uncorrelated. The ACF plots of WF2, WF5, and WF8 show that sample autocorrelation exceeds the limits for multiple lags. For WF2, sample correlation at lags 5 to 8, 13, and 16 exceeds the threshold limit. This high sample autocorrelation value indicates that the correlation has been left in the residuals due to model fitting errors that must be extracted using copula models.

The QQ plot is drawn between standard normal quantiles and sample residual quantiles. The QQ plot will be straight if the residual follows a normal or uniform distribution. However, the sample quantiles for WF residuals are not properly aligned with standard normal quantiles. This demonstrates that the residuals do not follow the marginal distribution and have tail asymmetry. As shown in the figure, the tail asymmetry is greater for WF5 and WF8 than for WF2.

Table 5 Energy Score Evaluation

WFs	ARIMA	VAR	VARMA	ANN	GAN	Proposed
WF1	3.10	0.31	0.32	0.57	0.91	**0.30**
WF2	0.20	0.42	0.26	1.48	1.26	**0.17**
WF3	0.75	0.70	0.69	0.78	1.44	**0.59**
WF4	0.21	0.19	0.27	0.54	1.40	**0.18**
WF5	1.69	1.15	1.13	**0.51**	0.80	0.85
WF6	1.34	1.31	1.24	0.72	**0.52**	0.98
WF7	0.92	0.30	0.29	1.19	1.14	**0.28**
WF8	1.40	1.38	0.81	0.68	**0.47**	0.54
WF9	0.78	0.70	0.64	1.09	**0.50**	0.52

Copula models must be used to model such asymmetrical dependency.

Copula Results

The estimated residuals from VARMA model are separated for each time step to model dependency using R-vine copula. Nine-dimensional R-vine copula is required to model dependency between considered WFs for each time step. Further, PCC decomposed a 9-dimensional copula into 36 bivariate copulas using the sequential approach as discussed in Algorithm 2. Eight trees graphically represent the dependency between bivariate copulas. The obtained tree structure for first-time step is shown in Fig. 3. The nodes {1 *to* 9} of the initial tree represents the nine considered WFs. These nodes are connected through the eight edges {5,8},{1,5},{3,1}, {2,3}, {6,4}, {6,2}, {9,6} and {9,7}. The copula type, number, estimated parameters, and Kendall's taus $\hat{\tau}$ for each edge of the tree are shown in Table 4. For the second tree the nodes {5,8} and {1,5} are joined by the edge {1, 8|5} because node {5} is common between node {1} and {8}, and satisfy the proximity condition. Similarly, the tree's other edges can be defined.

In the Fig. 3, the nodes of the first tree T_1 are the WFs and these are joined according to their dependency structure. It is observed in the tree T_1 that WF5 and WF8 are closer to each other than WF3. That is also observed from the Fig. 1, the geographical distance between WF5 and WF8 is smaller than their distance from WF3.

The range of $\hat{\tau}$ from -0.64 to 0.67 for tree T_1 shows the strong dependency between the WFs. After constructing the vine tree, different copulas are selected for each edge. The selection process is typically done using a sequential method mentioned in the Algorithm 2. This algorithm evaluates various candidate copulas for each edge and selects the one that best captures the dependence structure observed in the residual data. The sequential method uses statistical criteria, such as AIC to compare the goodness-of-fit of different copulas. The AIC criteria measure how well a copula fits the observed data and assess the quality

of the dependence structure captured by the copula. The sequential method iteratively evaluates different copulas for each edge, assessing their fit to the data and selecting the one that minimizes the selected criterion. This process is repeated for each edge in the vine tree until copulas have been selected for all edges. The type of copula selected for each edge in any tree also depends on the type of tail dependency presented in the residuals of WFs. For the first time step, the selected copulas are Gaussian, Frank, t, BB8, Rotated BB8 (90 and 270 degrees), Rotated tawn (type 1 and 2), and independent as shown in Table 4. Among these copulas, Frank and Gaussian are one parameter copulas.

From the Table 4, it is observed that number of bivariate copulas and value of $\hat{\tau}$ are decreasing from tree T_1 to T_8. Thus, mostly independent copulas are selected for higher-level trees and most dependences are modeled through the lower-level trees. The high proportion of independence copulas can significantly reduce the computational burden of modeling the R-vine copula. Additionally, the dependence model is an offline model. Once the model is established, enough scenarios can be extracted from it without excessive computation. It is not necessary to update the dependence model daily. In practice, it can be updated every few weeks.

The selection of R-vine copula instead of Gaussian copula to model for modeling spatial correlation between WFs is further validated by AIC information criteria in this paper. Figure 4 shows the calculated AIC over the given sample set for these two copula models for 120-steps. This figure shows that R-vine copula outperforms the gaussian copula in terms of minimum AIC. So, R-vine copula offers higher accuracy and more flexibility to model asymmetrical dependences between WFs as compared to Gaussian copula.

Scenario Results

The generated per-unit wind power scenarios using proposed approach and set of benchmarks models for denoted WFs are
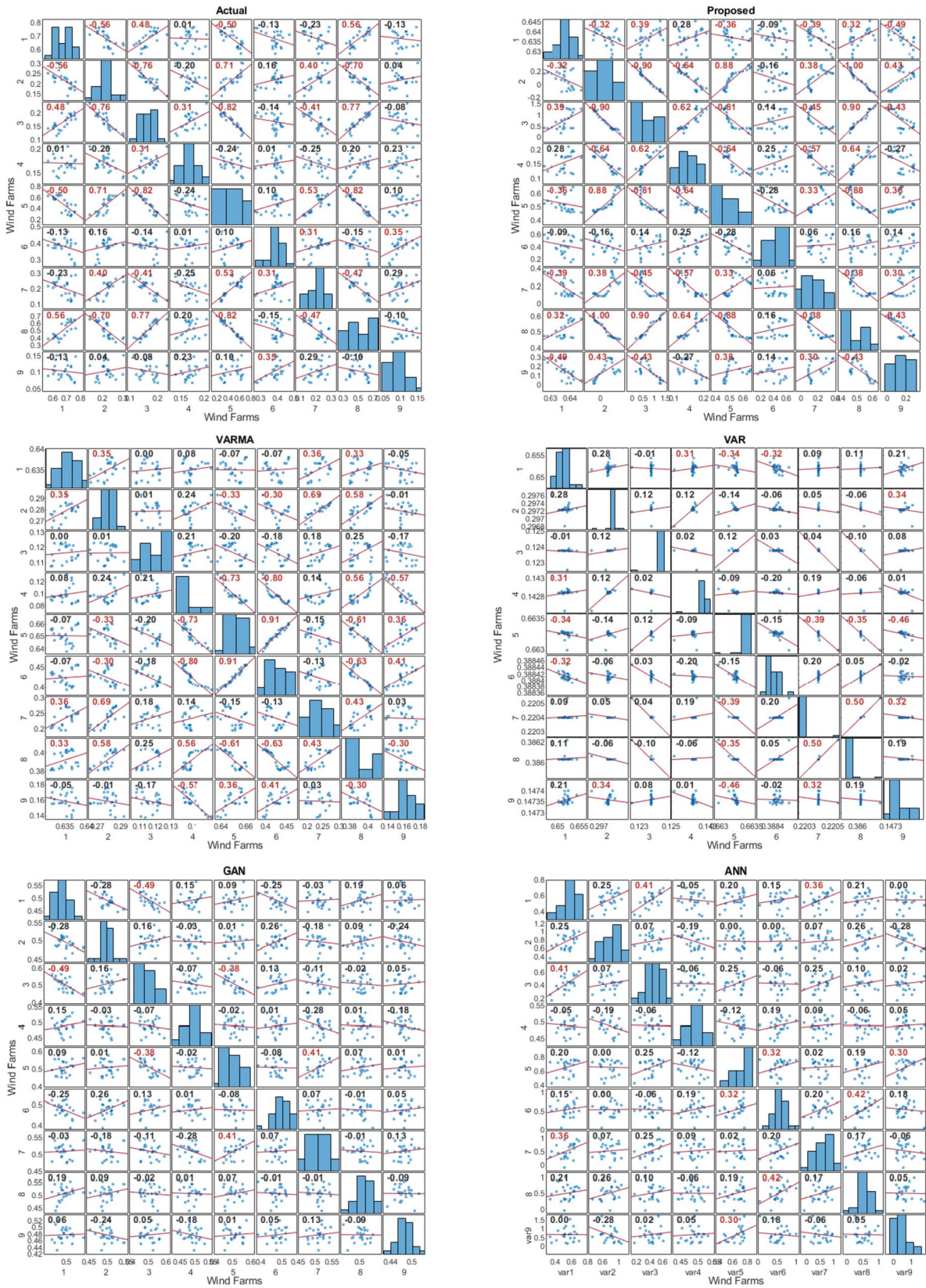
Fig. 6 Kendall's correlation for generated scenrios and actual output of considered WFs. Scenarios are generated using proposed approach, VARMA, VAR, GAN and ANN methods
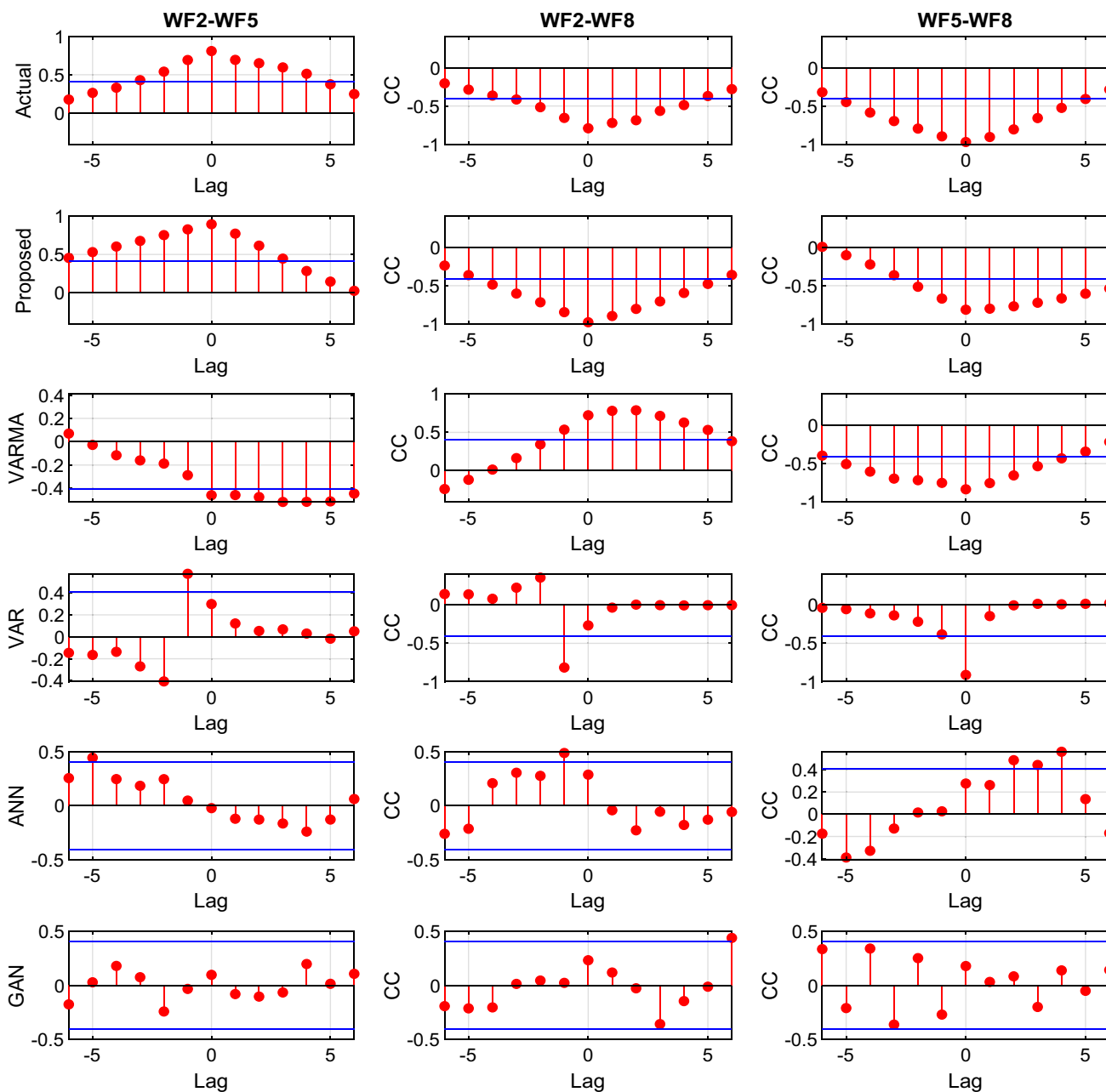
Fig. 7 CCF plots of three WFs for their actual data and mean power scenario using propoased approach, VARMA, VAR, ANN and GAN methods

shown in Fig. 5. A total of 1000 scenarios are generated for each WF and 50 scenarios are plotted to better visualize multiple parameters through YY plot. In Fig. 5 left Y-axis shows the actual value and, the right Y-axis show the scenarios & mean scenario. The significant bias between the actual values and generated scenarios using a set of benchmark models are witnessed in the scenario results depicted through YY plot. From Fig. 5, it is observed that scenarios generated using proposed distribution free approach capture the wind uncertainty with minimum bias. Although the assumed error distribution is same for all the

approaches, a significant difference is observed in generated scenarios. Because VAR Model underperform and linearizes with time and the ARIMA model shows abrupt waves and irregular fluctuations and spikes in generated scenarios.

Scenario Results Evalauation through Energy Score, Kendall's correlation and Cross-Correaltion Plots

The proposed model can generate quality scenarios, which are then evaluated using energy scores, CCF plot and

heat plot. The energy score for proposed and benchmark approaches are provided in Table 5. The table highlights the minimum energy score obtained for the proposed and benchmark approaches in bold letter. In terms of minimum energy score, the proposed method outperforms 55.56% of WFs. However, GAN and ANN outperform for 33.33% and 11% of WFs, respectively. In the context of time series approaches, VAR offers the lowest energy score for 89% of WFs when compared to the ARIMA model. This is due to the VAR model's multivariate nature. When compared to VAR and VARMA, the proposed VARMA-Copula model provides the lowest energy scores for all WFs. Because the proposed method can model both correlation and tail dependency.

The Kendall's correlation plots and cross-correlation function are used to check spatiotemporal correlation in the generated scenarios. The Kendall's correlation plot for the considered WFs is shown in Fig. 6. Kendall's correlation ranges from -1 to 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation. This correlation figure is plotted using the actual data and generated scenarios through proposed and benchmark approaches to show the spatial correlation between the WFs. From Fig. 6, it is visualized that rank correlation obtained for scenarios generated using the proposed approach is almost like the correlation obtained using real or actual data. However, it is significant for other benchmark approaches. The minimum deviation in heat map validates that the proposed approach can retain spatial correlations in the generated scenarios.

To complement the Kendall's correlation coefficient, a significance test using the "r test" is also performed on sample scenario generated using the proposed approach and benchmark approaches and the actual scenario. This test determines if the observed correlation coefficient is significantly different from zero, allowing you to assess the statistical significance of the association. The null hypothesis assumes no correlation (tau = 0), and the alternative hypothesis suggests a non-zero correlation. The result of "r test" is highlighted by the red color in the Fig. 6. It shows the strong correlation between the WFs. For example, WF1 has the strong correlation with WF2, WF3, WF5 and WF8 (Fig. 6).

The Kendall's rank correlation coefficient matrix can be used to create correlation figure, which only display the spatial correlations between the two data series over the given time period. On the other hand, CCF captures the spatial correlation between fixed and shifted (lagged) series as a function of time. The CCF for represented WFs are shown in Fig. 7. This figure shows that the cross-correlation plot for the mean scenario using the proposed VARMA-Copula approach follows a nearly identical pattern and moves in the same direction as the cross-correlation plot for the actual data. However, the CCF plot of the actual data differs significantly from the CCF plot of the WFs combinations for the VARMA, VAR, ANN, and GAN models. The proposed approach provides a two-stage modelling solution for spatial correlation between WFs. At first, spatial correlation is modelled using mutual parameters of the VARMA model. Later, spatial correlation is recovered from the residuals of the VARMA model using the R-Vine copula.

Conclusion

To address the uncertainties in various decision-making problems of power system and electricity market, a high-quality generated scenarios plays a significant role. A hybrid VARMA-Copula based distribution-free approach has been proposed in this paper to generate wind power scenarios considering spatiotemporal correlation. The proposed model can generate scenarios for multiple WFs considering spatiotemporal correlation without the assumption of the marginal distribution. The proposed model has been implemented on the nine WFs located in Australia. The superiority of the proposed scenario generation approach has been proved by comparison with a set of benchmark models. The obtained results show that the proposed approach can generate quality scenarios with minimum energy scores without loss of spatiotemporal correlation, which can easily be depicted in the Kendall's correlation and CCF plots. It emphasizes that uncertainty modelling is improved by incorporating spatial–temporal correlation between nearby WFs. The R-Vine is a flexible copula model that uses a variety of bivariate copulas to represent asymmetrical dependency. This extends the proposed method for generating quality wind power scenarios with minimum energy scores. The proposed method can be enhanced by applying adaptive and non-linear forecasting models with time-varying parameters to generate wind power scenarios. The proposed work could be extended to generate load, solar generation, and price scenarios for different power systems and electricity markets applications.

References

1. Global Wind Energy Council (2022) Global Wind Report. https://gwec.net/wp-content/uploads/2022/03/GWEC-GLOBAL-WIND-REPORT-2022.pdf. Accessed 6 June 2023

2. Hui L, Zhouyang R, Miao F, Wenyuan L, Yan X, Yunpeng J, Weiyi X (2022) A review of scenario analysis methods in planning and operation of modern power systems: Methodologies, applications, and challenges. Electr Power Syst Res 205:107722. https://doi.org/10.1016/j.epsr.2021.107722

3. Jinghua L, Zhou J, Chen B (2020) Review of wind power scenario generation methods for optimal operation of renewable energy systems. App Energ 280:115992. https://doi.org/10.1016/j.apenergy.2020.115992

4. Pierre P, Henrik M, Henrik AN, Georgios P, Bernd K (2009) From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind Energy 12:51–62. https://doi.org/10.1002/we.284

5. Morales JM, Minguez R, Conejo AJ (2010) A methodology to generate statistically dependent wind speed scenarios. App Energ 87:843–855. https://doi.org/10.1016/j.apenergy.2009.09.022

6. Kailash CS, Prerna J, Rohit B (2013) Wind power scenario generation and reduction in stochastic programming framework. Electr Power Comp Syst 41:271–285. https://doi.org/10.1080/15325008.2012.742942

7. Guzman D, Javier GA, Jose C (2016) Wind power scenario generation through state-space specifications for uncertainty analysis of wind power plants. App Energy 162:21–30. https://doi.org/10.1016/j.apenergy.2015.10.052

8. Xi-Yuan M, Yuan-Zhang S, Hua Liang F (2013) Scenario generation of wind power based on statistical uncertainty and variability. IEEE Trans Sustain Energy 4:894–904. https://doi.org/10.1109/TSTE.2013.2256807

9. Duehee L, Ross B (2017) Load and wind power scenario generation through the generalized dynamic factor model. IEEE Trans Power Syst 32:400–410. https://doi.org/10.1109/TPWRS.2016.2562718

10. Stylianos IV, Evaggelos GK, Christos KS, Anastasios GB, Joas PSC (2016) ANN-based scenario generation methodology for stochastic variables of electric power systems. Electr Power Syst Res 134:9–18. https://doi.org/10.1109/ISAP.2015.7325540

11. Yize C, Yishen W, Daniel K, Baosen Z (2018) Model-free renewable scenario generation using generative adversarial networks. IEEE Trans Power Syst 33:3265–3275. https://doi.org/10.1109/TPWRS.2018.2794541

12. Congmei J, Yongfang M, Yi C, Mingbiao Y, Songbing T (2018) Scenario generation for wind power using improved generative adversarial networks. IEEE Access 6:62193–62203. https://doi.org/10.1109/ACCESS.2018.2875936

13. Yufan Z, Qian A, Fei X, Ran H, Tianguang L (2020) Typical wind power scenario generation for multiple wind farms using conditional improved Wasserstein generative adversarial network. Inter J Elect Power Energy Syst 114:105388. https://doi.org/10.1016/j.ijepes.2019.105388

14. Jin T, Qiuwei W, Menglin Z, Wei W, Nikos H, Feng L, Theodoros K (2021) Wind power scenario generation with non-separable spatio-temporal covariance function and fluctuation-based clustering. Int J Electr Power Energy Syst 130:106955. https://doi.org/10.1016/j.ijepes.2021.106955

15. Zhe Y, Zhiping C, Giorgio C (2019) A copula-based scenario tree generation algorithm for multiperiod portfolio selection problems. Ann Oper Res 292. https://doi.org/10.1007/s10479-019-03147-9

16. Jeffrey D, Eike CB, Claudia C, Dorota K (2012) Selecting and estimating regular vine copula and application to financial returns. Comp Stat & Data Analysis 59(52–69):59. https://doi.org/10.1016/j.csda.2012.08.010

17. Tim B, Roger C (2002) Vines-a new graphical model for dependent random variables. Ann Statist 30:1031–1068. https://doi.org/10.1214/aos/1031689016

18. Olivier PF (2013) Sklar's theorem by probabilistic continuation and two consistency results. J Multivar Anal 122:271–277. https://doi.org/10.1016/j.jmva.2013.07.010

19. Jan FM, Matthias S (2017) Simulating Copulas, Series in Quantitative Finance. World Scientific Publishing Com, Singapore

20. Christian G, Anee CF (2007) Everything you always wanted to know about copula modeling but were afraid to ask. J Hydrol Eng 12(347–368):2007. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)

21. Raik B (2018) Generation of time-coupled wind power infeed scenarios using pair-copula construction. IEEE Trans Sust Energ 9:1298–1306. https://doi.org/10.1109/TSTE.2017.2782089

22. Mario AR, Kristen S (2022) Spatio-temporal probabilistic forecasting of wind power for multiple farms: A copula-based hybrid model. Int J Forecast 38:300–320. https://doi.org/10.1016/j.ijforecast.2021.05.013

23. Yibin Q, Qi L, Yuru P, Hanqing Y, Weirong C (2019) A scenario generation method based on the mixture vine copula and its application in the power system with wind/hydrogen production. Inter J Hydro Energy 44:5162–5170. https://doi.org/10.1016/j.ijhydene.2018.09.179

24. Zhao W, Weisheng W, Chun CL, Zheng W, Yunhe H (2018) Probabilistic forecast for multiple wind farms based on regular vine copulas. IEEE Trans Power Syst 33:578–589. https://doi.org/10.1109/TPWRS.2017.2690297

25. Zhao W, Weisheng W, Chun C L, Bo W (2020) Forecasted scenarios of regional wind farms based on regular vine copulas. J Mod Power Syst Clean Energy 8:77–85. https://doi.org/10.35833/MPCE.2017.000570

26. Qingyu T, Shihong M, Fuxing Y, Yaowang L, Haoran Y, Ji H, Di Z, Weichen Y (2021) Forecasting scenario generation for multiple wind farms considering time-series characteristics and spatial-temporal correlation. J Mod Power Syst Clean Energy 9:837-848. https://doi.org/10.35833/MPCE.2020.000935

27. Giovani G, Zhou R, J. L (2021) Spatial-temporal modelling of wind speed-a vine copula based approach. http://hdl.handle.net/11343/282427. Accessed 6 June 2023

28. Qingyu T, Shihong M, Fuxing Y, Weichen Y, Yujun L, Zhong Z (2023) An improved wind power uncertainty model for day-ahead robust scheduling considering spatio-temporal correlations of multiple wind farms. Int J Elect Power Energy Syst 145:108674. https://doi.org/10.1016/j.ijepes.2022.108674

29. Yanting L, Xinghao P, Yu Z (2022) Forecasting methods for wind power scenarios of multiple wind farms based on spatio-temporal dependency structure. Ren Energ 201:950–960. https://doi.org/10.1016/j.renene.2022.11.002

30. Attoti BK, Abhijit RA (2023) Time-coupled day-ahead wind power scenario generation: a combined regular vine copula and variance reduction method. Energ 265:126173. https://doi.org/10.1016/j.energy.2022.126173

31. Mengqiu F, Xiang F, Xu B, Wang T, Pan L, Liu Y, Liu J (2022) Data-Driven Load Pattern Identification Based on R-Vine Copula and Random Forest Method. IEEE Trans Indus App 58:7919–7929. https://doi.org/10.1109/TIA.2022.3200920

32. Miao H, Vijay V, Junshan Z (2015) A sparsified vector autoregressive model for short-term wind farm power forecasting. IEEE Power Energy Soc Gen Meeting 1–5. https://doi.org/10.1109/PESGM.2015.7285972

33. Jethro D, Pinson P (2016) Very-short-term probabilistic wind power forecasts by sparse vector autoregression. IEEE Trans Smart Grid 7:763–770. https://doi.org/10.1109/TSG.2015.2424078

34. Tsay RS (2014) Multivariate Time Series Analysis. Wiley, Canada

35. Rue ST, David W (2021) All-Purpose Toolkit for Analyzing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models. R package version. [Online]:Available: https://cran.r-project.org/web/packages/MTS/MTS.pdf

36. Wan C, Xu Z, Pinson P, Dong ZY, Wong KP (2014) Probabilistic forecasting of wind power generation using extreme learning

machine. IEEE Trans Power Syst 29(3):1033–1044. https://doi.org/10.1109/TPWRS.2013.2287871

37. Hu Q, Zhang S, Man Y, Xie Z (2017) Short-term wind speed or power forecasting with heteroscedastic support vector regression. IEEE Trans Sustain Energy 7(1):241–249. https://doi.org/10.1109/TSTE.2015.2480245

38. Emiliano T, Stefano M, Paul E, Bruno S (2019) A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas. Prob Eng Mech 55:1–16. https://doi.org/10.1016/j.probengmech.2018.08.001

39. Pinson P, Girard R (2012) Evaluating the quality of scenarios of short-term wind power generation. App Energy 96:12–20. https://doi.org/10.1016/j.apenergy.2011.11.004

40. Jethro D (2015) Australian Electricity Market Operator (AEMO) 5 Minute Wind Power Data. University of Strathclyde [Online]. Available: https://bit.ly/38vReAU. Accessed 16 Jan 2023

41. Kalid Y, Torbjorn T, Peiyuan C (2015) ARIMA-based frequency-decomposed modeling of wind speed time series. IEEE Trans Power Syst 31:2546–2556. https://doi.org/10.1109/TPWRS.2015.2468586

42. Sambeet M, Würsig C, Chiara B (2019) Multivariate scenario generation -an arima and copula approach. Int J M Opt 9:146–149. https://doi.org/10.7763/IJMO.2019.V9.700

43. Thomas N, et al. (2021) Statistical Inference of Vine Copulas. R package version 2.4.3 [Online]. Available: https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf Accessed 16 Jan 2023

44. Comprehensive R Archive Network (CRAN), [Online] Available: https://cran.r-project.org Accessed 16 Jan 2023

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.