processes

MDPI

Article

Multi-Scale Scenario Generation for Large Hydro–Wind–Solar Systems Using Clustering and Vine Copulas

Congtong Zhang [1], Zixuan Xu [2], Yushu Wu [2], Yan Jiang [1], Kai Chen [1] and Jianjian Shen [2,*]

[1] Yunnan Power Dispatching and Control Center, Kunming 650011, China; zcongtong@163.com (C.Z.); jiangyan4585@163.com (Y.J.); 15096632707@163.com (K.C.)
[2] Institute of Hydropower and Hydroinformatics, Dalian University of Technology, Dalian 116024, China; 32406223@mail.dlut.edu.cn (Z.X.); 2461037435@mail.dlut.edu.cn (Y.W.)
* Correspondence: shenjj@dlut.edu.cn

Abstract

Large-scale hydro–wind–solar systems require scenario sets that represent cross-resource dependence without losing station-level and intra-day characteristics. A remaining challenge is to connect cluster-level joint modeling with station allocation under sum constraints and hourly reconstruction consistent with daily totals. This study develops an integrated framework: wind and photovoltaic plants are grouped by output correlation; season-specific C-vine copulas model daily wind–solar–runoff dependence; a centered log-ratio transformation and Gaussian R-vine copula allocate cluster totals to stations; and a stratified typical-day library reconstructs hourly profiles. The framework is evaluated using 872 observed days of hourly data from a grid-connected system in Southwest China comprising 53 wind plants, 47 photovoltaic plants, and one hydropower-station runoff series; 1000 joint scenarios are generated. Against independent marginal sampling, Spearman correlation-matrix errors fell from 0.419–0.432 to 0.318–0.395 (8.7–24.2%), and Kendall errors from 0.306–0.325 to 0.236–0.298 (8.3–22.8%). Compared with historical-share resampling, station allocations were more diverse but less faithful to historical share distributions; compared with a unified typical-curve library, intra-day autocorrelation errors decreased across all resource-season cases. The resulting multi-scale scenarios can serve as generation-side inputs for stochastic scheduling and operational risk analysis.

Keywords: scenario generation; hydro–wind–solar systems; C-vine copula; compositional data; renewable energy

Check for updates

Academic Editor: Davide Papurello

Received: 14 September 2026
Revised: 20 October 2026
Accepted: 22 October 2026
Published: 26 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The increasing deployment of wind and photovoltaic (PV) generation has made the coordinated operation of renewable resources an important issue for power systems [1–7]. In hydropower-rich regions, hydro–wind–solar systems can exploit the complementary characteristics of controllable hydropower and variable wind and PV generation to improve renewable-energy utilization and operational flexibility [8–14]. Effective coordination also requires an appropriate representation of resource uncertainty [15]. Wind and PV outputs vary with meteorological conditions, while runoff exhibits pronounced seasonal variation; their joint behavior can differ between the flood and dry seasons. Moreover, stations distributed across a region need not exhibit identical output patterns. A useful scenario set must therefore describe plausible combinations of resource availability while retaining station-level variation and relevant within-day characteristics. Such generation-side information provides a basis for examining alternative operating conditions and

associated risks in large-scale complementary systems [16–21]. The modeling task is consequently broader than reproducing the separate distribution of each resource: it also concerns how uncertain daily quantities and intra-day profiles are represented together.

To characterize the uncertainty described above and support the coordinated operation of hydro–wind–solar systems, existing research has investigated forecasting, scenario generation, and operational decision-making [22,23]. Time-series and learning-based forecasting methods estimate future output or its predictive distribution from available observations [24–26]. Robust analysis and optimal control studies use uncertainty descriptions to determine operating actions [27], while recent interpretable multi-step wind-power forecasting research further connects prediction with operational applications [28]. Scenario generation has a related but distinct role: it constructs multiple plausible realizations through which the joint behavior of uncertain resources can be examined. Forecast accuracy alone does not establish whether a scenario ensemble reproduces dependence among wind power, PV generation, and runoff, or whether it retains differences among individual stations. Likewise, a scheduling or control formulation does not itself specify the statistical properties of its generation-side inputs. For the application considered here, the required outputs are joint daily resource quantities and corresponding station-level intra-day profiles. This defines a scenario-synthesis task that can support subsequent operational analysis without treating scenario generation itself as a dispatch optimization problem.

To meet these scenario-synthesis requirements, existing studies have explored data-driven generative methods and explicit statistical dependence models, together with dimension-reduction strategies for large station collections. Diffusion-based PV scenario generation [29] and conditional generative modeling of renewable-power output [30] illustrate the use of learned distributions to construct scenario ensembles. For geographically distributed wind resources, Shahirinia et al. [31] investigated dependence modeling with vine copulas, including the effects of dataset size and copula choice. Tan et al. [32] combined station clustering, a C-vine dependence model, and time-series reconstruction for high-dimensional wind–PV scenarios. Related work on equivalent representations of multiple wind farms [33] demonstrates the use of aggregation for system analysis, although its objective is dynamic interaction assessment rather than statistical scenario synthesis. For hydro–wind–solar systems, Iung et al. [34] developed a copula-based joint generation framework, Liu et al. [35] combined temporal and spatial dependence modeling to generate daily multi-resource scenarios, and Feng et al. [36] developed a two-stage generative approach combining variational autoencoders and generative adversarial networks. These studies provide important foundations for dimension reduction, dependence modeling, and coupled resource scenario generation. For large-scale hydro–wind–solar applications, a further modeling requirement is to connect joint daily resource scenarios with station-level allocation and intra-day reconstruction through a coherent generation process.

For the present large-scale hydro–wind–solar application, the specific requirement is to connect cluster-level joint generation with station-level allocation and subsequent intra-day reconstruction. This differs in emphasis from reconstructing wind–PV scenarios without runoff [32], generating daily multi-resource scenarios [35], or developing a two-stage generative model for hydro–wind–solar scenario synthesis [36]. Clustering makes joint modeling more tractable, but the resulting cluster totals do not uniquely determine the allocation to individual stations. Station shares are compositional variables: they must be non-negative and sum to one within each cluster. Once station-level daily quantities have been obtained, hourly profiles must be reconstructed in relation to those quantities rather than selected without regard to daily resource levels. The practical challenge is therefore to link these representations through explicit allocation and reconstruction rules, while evaluating the resulting joint, spatial, and temporal characteristics. The distinction addressed here lies in

this constrained daily-to-hourly workflow and its station-level outputs, rather than in the general observation that renewable resources exhibit spatiotemporal dependence.

To link these modeling stages within a unified generation process, this study proposes a multi-scale scenario-generation framework for large-scale hydro–wind–solar systems. Wind and PV stations are grouped according to their output-correlation patterns, reducing the dimension of subsequent joint modeling while retaining a structured representation of regional output dependence. Season-specific C-vine copulas then generate daily joint scenarios for wind-power clusters, PV clusters, and runoff. For station-level spatial allocation, a centered log-ratio (CLR) transformation and a Gaussian R-vine copula model the allocation shares under the cluster sum constraint. A typical daily curve library classified by season and daily quantity subsequently reconstructs station-level intra-day profiles. The contribution is the integration of these components into a modular generation process with explicit links between cluster totals, station allocations, and hourly profiles. Its distinguishing features are the inclusion of runoff in seasonal joint generation, compositional treatment of station shares, and daily-quantity-based reconstruction of intra-day profiles. The framework is evaluated using observations from a large grid-connected hydro–wind–solar system in Southwest China, with separate assessments of joint dependence, allocation characteristics, and temporal reproduction. Figure 1 presents the overall workflow.

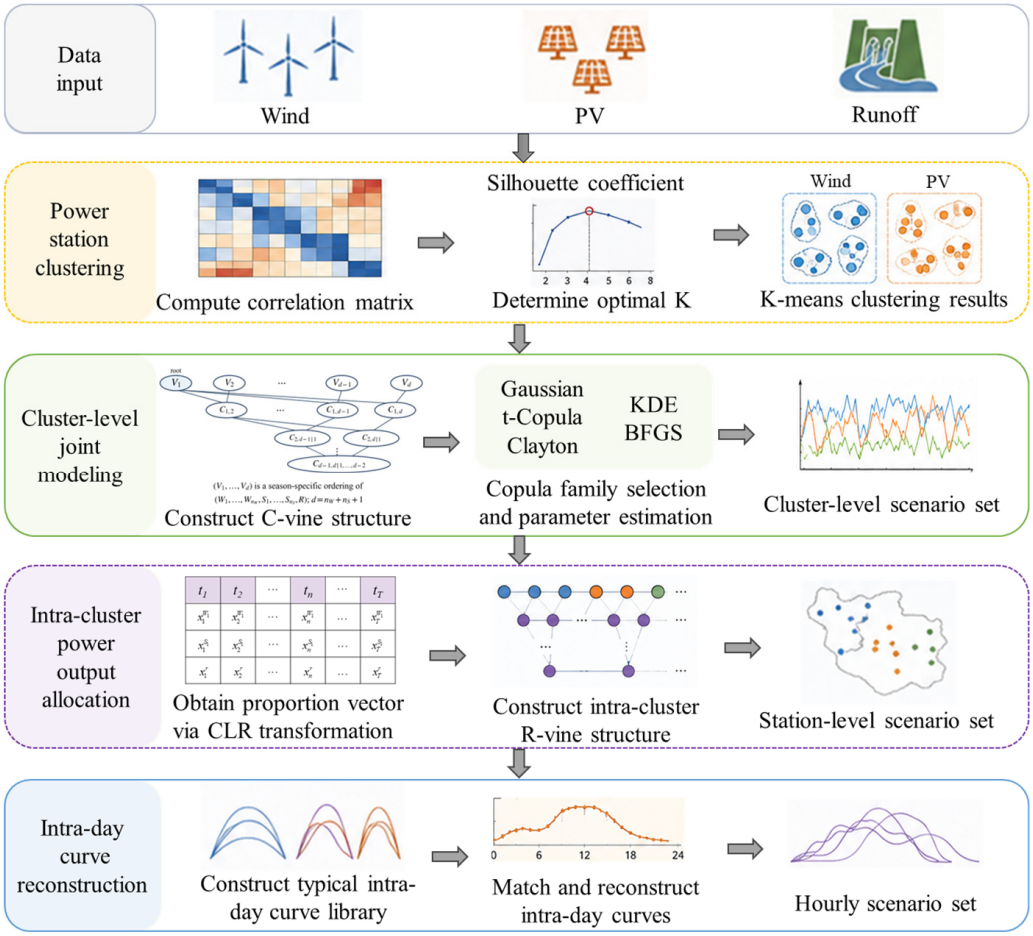


Figure 1. Framework diagram of the proposed scenario generation method.

2. Power Station Clustering Based on Correlations

With the rapid large-scale development of wind and photovoltaic (PV) power stations, the number of power stations in the same area has increased significantly [1]. If all power stations are directly modeled in a high-dimensional way, the parameter dimensions of

the Copula model will increase sharply [35], thereby significantly increasing the model complexity and the difficulty of parameter estimation. Therefore, this paper first conducts cluster analysis on the wind and PV power stations within the area, grouping stations with similar correlations into the same cluster. This reduces the dimensionality of subsequent joint modeling while effectively preserving the main spatial correlation structure within the area [37].

This paper uses the Spearman correlation coefficient to measure the degree of correlation between power stations. The Spearman correlation coefficient has good adaptability to monotonic nonlinear relationships and can accurately describe the synchronous variation characteristics of renewable energy power output [38].

The Spearman correlation coefficient between the *i-th* and *j-th* power stations is defined as:

$$\rho_{ij} = 1 - 6\sum_{t=1}^{T}\left(R_{i,t} - R_{j,t}\right)^2 / T\left(T^2 - 1\right) \quad (1)$$

In the formula: $R_{i,t}$ represents the output ranking value of power station i at time t; $R_{j,t}$ represents the output ranking value of power station j at time t.

Based on the Spearman correlation coefficient described above, an N × N correlation coefficient matrix **C** is constructed, where the *i-th* row represents the correlation between the *i-th* power station and all other power stations, and can be regarded as the correlation feature vector of that station. This vector is used as the input feature for clustering analysis in this paper.

Due to the high dimensionality of the original features, direct clustering is prone to redundant information and the curse of dimensionality. Therefore, principal component analysis (PCA) is performed on the feature matrix to project the original high-dimensional features into a low-dimensional principal component space, effectively extracting the main correlation structure in the data. In this paper, the number of principal components that achieve a cumulative variance contribution rate of 90% is selected as the reduced dimension. This retains the key correlation information to the greatest extent while significantly reducing the complexity of the feature space. The retained principal-component scores are then used as inputs to K-means clustering, which is performed separately for wind and PV stations. For both wind and PV stations, candidate cluster numbers from K = 2 to K = 8 were tested. Each K-means fit used 10 initializations and a fixed random seed of 42, and the final cluster number was selected using the average silhouette coefficient [39]. The silhouette coefficient comprehensively considers both the compactness within clusters and the separation between clusters, and is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]} \quad (2)$$

In the formula: $a(i)$ represents the average distance between sample i and samples of the same cluster; $b(i)$ represents the average distance between sample i and the nearest neighboring cluster. The silhouette coefficient ranges from -1 to 1, and the closer its value is to 1, the better the clustering effect. The average silhouette coefficient is calculated for different cluster numbers K, and the clustering result with the best value is selected as the final classification scheme.

After clustering, stations with similar output-correlation profiles are grouped together, providing a structured, lower-dimensional representation for subsequent cluster-level daily total modeling, station-level spatial disaggregation, and intra-day curve reconstruction.

3. Clustering-Level Daily Scale Scenario Generation

Multivariate Copula functions have extensive applications in economics, finance, and risk management. Through multivariate Copula functions, researchers can flexibly construct models for dependence among high-dimensional random variables without being restricted by the form of marginal distributions. This paper adopts the high-dimensional mixed Copula function C-vine conditional decomposition modeling method [40] to fit and model the correlations of wind power clusters, photovoltaic (PV) clusters, and runoff sequences within the same region, determine the binary conditional Copula equivalent form of the high-dimensional Copula joint distribution function composed of multiple energy sources, and thereby depict the joint probability structure of the hydro–wind–solar system on a daily scale. Finally, the corresponding joint distribution scenario set is generated using the inverse transform sampling method [41]. The schematic diagram of the method is shown in Figure 2.

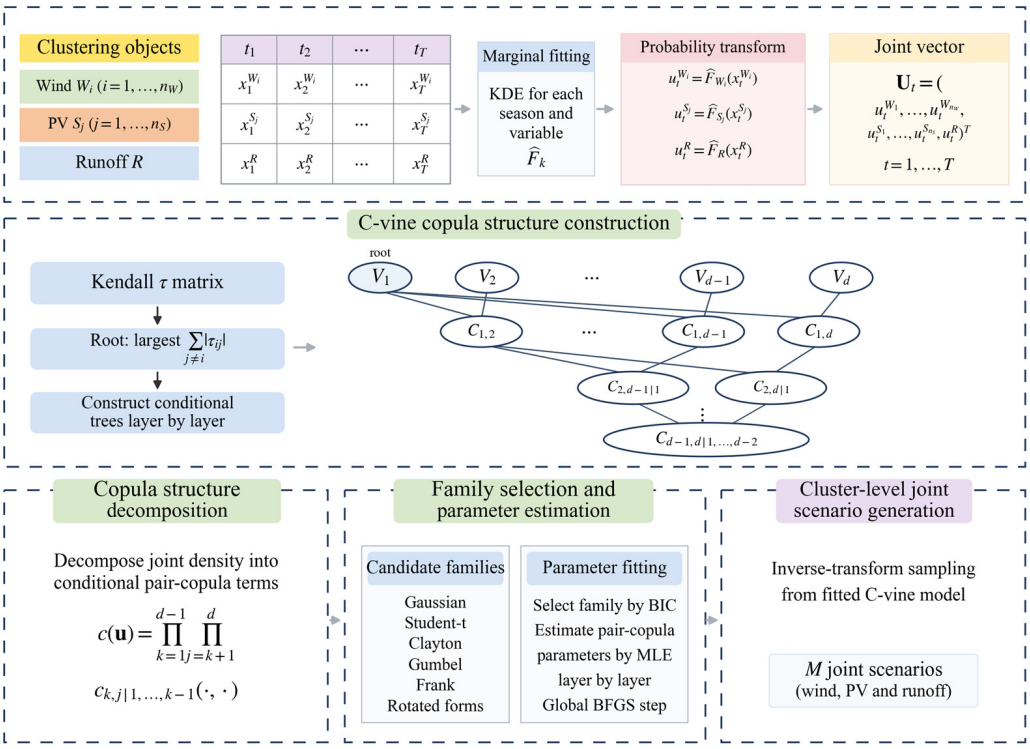


Figure 2. Principle diagram of the C-vine Copula method.

3.1. C-Vine Copula Theory

After clustering within the region, there are a total of $d = n_w + n_s + 1$ objects (n_w wind power clusters, n_s photovoltaic (PV) clusters, and one runoff sequence). The density functions of the joint distribution, marginal distributions, and conditional distributions at the daily scale are all continuous. The random vector composed of these d variables is $X = (X_1, X_2, \cdots X_d)$, and the corresponding joint distribution function is $F(x_1, x_2, \cdots x_d)$. According to Sklar's theorem, the multivariate joint distribution function can be expressed in terms of a Copula function and the marginal distribution functions of each variable as:

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) \quad (3)$$

The corresponding multivariate joint density function can be expressed as:

$$f(x_1, x_2, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) \prod_{i=1}^{d} f_i(x_i) \quad (4)$$

In the formula: $F_i(x_i)$ represents the marginal distribution function of the *i-th* variable; $C(\cdot)$ is the d-dimensional Copula function; $c_{1,2,\ldots,d}$ is the d-dimensional Copula density function; $f_i(x_i)$ is the marginal density function of the *i-th* variable.

The marginal density function of a variable can generally be obtained through parametric assumptions or non-parametric methods, while the form of a high-dimensional Copula function is difficult to determine directly. Gaussian and Student-t Copulas provide tractable multivariate models, but the Gaussian Copula has no asymptotic tail dependence and the standard Student-t Copula has symmetric upper- and lower-tail dependence [42]. These restrictions may limit their flexibility when dependence patterns differ across variable pairs; vine constructions allow different bivariate Copula families to be combined [40].

A multivariate joint Copula probability density function can be constructed as the product of multiple bivariate conditional Copula probability density functions corresponding to different conditions. A graphical structure called "vine" is introduced to describe the logical relationships after decomposing random variables. The main vine types include C-vine, D-vine, and R-vine [40,43]. Among them, the C-vine Copula method is mainly used to describe the dependence among random variables with a hierarchical structure, allowing the multivariate joint Copula density function to be decomposed into the product of several bivariate conditional Copula density functions, where each pair Copula can adopt a different Copula family.

For cluster-level joint modeling, a C-vine is adopted because its root-centered structure allows variables with stronger aggregate dependence on the remaining variables to be placed successively at the center of the decomposition. The root order is determined from the seasonal Kendall correlations, while the pair-copula family is selected for each edge. A D-vine instead follows a chain-like structure, and a general R-vine permits a broader range of tree structures; nested Archimedean copulas represent dependence through generator-based nesting rather than the pair-copula decomposition used here. Thus, the C-vine provides an explicit and interpretable dependence ordering for the proposed framework. The resulting C-vine density can be decomposed as follows:

$$c_{1,2,\ldots,d} = \prod_{j=1}^{d-1} \prod_{k=1}^{d-j} c_{j,j+k|1,\ldots,j-1}\left(F\left(x_j|x_1, \ldots, x_{j-1}\right), F\left(x_{j+k}|x_1, \ldots, x_{j-1}\right)\right) \quad (5)$$

The marginal conditional distribution $F\left(x_j|\mathbf{v}\right)$ can be expressed in terms of the partial derivatives of the Copula function as follows:

$$F\left(x_j|\mathbf{v}\right) = \frac{\partial C_{j,\mathbf{v}}\left(F_j\left(x_j\right), F_{\mathbf{v}}(\mathbf{v})\right)}{\partial F_{\mathbf{v}}(\mathbf{v})} \quad (6)$$

In the formula: c in Equation (5) denotes a conditional pair-Copula density. In Equation (6), C denotes the corresponding conditional pair-Copula distribution function, and F denotes a conditional cumulative distribution function; v_j is the *j-th* variable in the vector v; v_{-j} represents the sub-vector obtained by removing the *j-th* variable.

3.2. Determination of the C-Vine Structure

The study area consists of d variables. The daily time series of each variable is regarded as a random variable $X_i (i = 1, 2, \ldots, d)$. To construct the C-vine model that best conforms to

the actual correlation structure, this paper adopts the sequential search method based on Kendall's correlation coefficient. The specific steps are as follows:

Assume that any variable X^* among $X_1, X_2, \ldots, X_d$ is the root node of the C-vine tree, and the remaining variables are reordered as $(X_2, X_3, \ldots, X_d)$. Calculate the Kendall correlation coefficient $\hat{\tau}_{*,i}$ between X^* and each of the other variables based on historical samples.

Let $X_1, X_2, \ldots, X_d$ serve as candidate root nodes X^* in turn, and calculate the total sum of their correlation strengths:

$$\hat{S}^* = \sum_{i=1}^{d-1} |\hat{\tau}_{*,i}| \tag{7}$$

Select the variable that maximizes $\hat{S}^*$ as the root node of the first layer of the C-vine, and renumber it as X_1. The remaining variables are renumbered as $X_2, X_3, \ldots, X_d$.

For the *e-th* layer (e = 2, 3, …, $d-1$), take any variable X^* among the current remaining variables as the root node of the conditional tree, repeat the above process to calculate its $\hat{\tau}$ and $\hat{S}^*$ with the other variables, and successively determine the variable order corresponding to each level node.

Through the above steps, the selected C-vine dependency structure can be constructed. It is worth noting that due to the significant differences in the correlation of wind, solar, and hydro among different hydrological periods, this paper establishes two independent C-vine models, constructing dependency structures for the flood season and the dry season separately.

3.3. Selection of C-Vine Copula Function Form and Parameter Estimation

After the C-vine structure is determined, the function family of each bivariate conditional Copula needs to be selected, and its parameters estimated, thereby fully determining the multivariate joint distribution function. The main steps are as follows:

The selection of the Copula function family adopts information criteria to comprehensively evaluate candidate models from the aspects of goodness-of-fit and model complexity. Common criteria include AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) [44,45]:

$$\mathrm{AIC} = -2\ln L + 2k \tag{8}$$

$$\mathrm{BIC} = -2\ln L + k\ln N \tag{9}$$

In the formula: k represents the number of parameters to be estimated in the Copula function; L is the maximum likelihood value; N is the effective sample size. AIC places more emphasis on fitting accuracy, while BIC can effectively prevent overfitting when the sample size is large. Considering that the scenario generation in this paper requires both generalization ability and physical rationality, the Copula family with the smallest BIC is preferred.

The parameters of the C-vine Copula function are determined using the maximum likelihood estimation method. The historical daily scale sample matrix is $Q = [q_1, q_2, \ldots, q_N]^T \in R^{N\times d}$, where N is the sample size. The log maximum likelihood function of the C-vine Copula is constructed as follows:

$$\mathcal{L}(\theta_c) = \sum_{t=1}^{N}\sum_{j=1}^{d-1}\sum_{k=1}^{d-j} \ln\left\{c_{j,j+k|1,\ldots,j-1}\left(\hat{F}\left(x_{j,t}|\mathbf{x}_{1:j-1,t}\right), \hat{F}\left(x_{j+k,t}|\mathbf{x}_{1:j-1,t}\right); \theta_{j,j+k}\right)\right\} \tag{10}$$

In the formula: θ_c is the vector composed of all Copula parameters of C-vine; $\hat{F}(\cdot)$ is the conditional cumulative distribution function estimated based on marginal distributions.

To enhance the stability of optimization, a layer-by-layer initialization strategy is adopted to solve the initial parameters: the marginal distributions of each variable are

fitted using kernel density estimation (KDE), and the original data are transformed into pseudo-observations $u_{i,t} = \hat{F}_i(x_{i,t})$. The first-layer bivariate Copula is used to estimate the parameters for $(X_1, X_j)(j = 2, \ldots, d)$; the conditional distribution $F(x_j|x_i)$ is calculated, and conditional pseudo-observations are generated, thereby estimating the parameters of the bivariate Copula at the second layer; this process is repeated, and the parameters of the Copula at the third to the $d-1$-*th* layers are estimated successively.

Using these parameters as initial values, the BFGS algorithm [46] in the quasi-Newton method is employed to jointly optimize the log-likelihood function and obtain the final parameter estimates.

3.4. Generation of Daily Total Scenario Sets at the Clustering Level

Based on the $U(0,1)$ distribution, R groups of d-dimensional random samples are generated. Using the inverse transform sampling method, the corresponding joint scenario sets at the daily scale for wind–solar–hydro clusters are generated through the constructed C-vine Copula joint distribution models for flood and dry seasons.

4. Internal Power Allocation Within Clusters

After obtaining the joint scenarios at the cluster-level daily scale, it is necessary to allocate them reasonably to individual power stations within each cluster. However, the power output ratios of stations in the same cluster exhibit typical compositional data characteristics [47,48]: each component is non-negative and strictly sums to 1. A generic multivariate Copula fitted directly to station shares does not, by itself, enforce their sum-to-one constraint. A simple strategy would systematically lose the spatial co-fluctuation structure within the cluster and reduce the physical credibility of the scenarios.

To fundamentally remove the closure constraint and preserve the true relative dependence structure among stations, this paper introduces the centered log-ratio (CLR) transformation to preprocess the power output ratios of stations within a cluster [47]:

$$z_i = \mathrm{clr}(p_i) = \ln p_i - \frac{1}{m}\sum_{j=1}^{m} \ln p_j,\ i = 1, 2, \ldots, m \tag{11}$$

In the formula, m represents the number of power stations within a cluster; p_i is the proportion of the *i-th* power station within the cluster. After the transformation, the variable satisfies $\sum_{i=1}^{m} z_i = 0$, and its effective degrees of freedom are only $m-1$. Therefore, only the first m-1 dimensions are modeled, and the m-th dimension is recovered from the zero-sum constraint. The isometric log-ratio (ILR) transformation provides an orthonormal alternative for compositional data [49], but the present implementation uses CLR coordinates.

The number of power stations within a cluster is large. The C-vine Copula is prone to bias and is difficult to fit stably in high-dimensional or non-centralized dependence scenarios. Therefore, the R-vine structure is used instead, with the tree sequence optimized automatically via a data-driven approach [50]. To balance high-dimensional stability and computational efficiency, the Copula family is simplified, retaining only the Gaussian family. Finally, the sampled m-dimensional vector is transformed back to the proportional form through the inverse CLR transformation, strictly ensuring that the sum of the allocation proportions of each power station equals 1, thus achieving station-level spatial power disaggregation under total conservation.

To maintain a parsimonious allocation structure for large-scale, high-dimensional systems, station-share vectors are generated separately from cluster daily totals within each season. This is a modeling approximation, not a claim of strict statistical independence. Conditional allocation can improve fidelity at certain total levels, but the seasonal specifica-

tion is retained to use the full seasonal sample without fitting separate high-dimensional models for each total range. The framework remains compatible with total-conditioned allocation when required.

5. Intra-Day Curve Reconstruction

To further generate complete time-series scenarios, it is necessary to restore the intra-day output curves of each power station. Considering that the intra-day fluctuation patterns of renewable energy vary significantly with resource abundance and scarcity, this paper constructs a typical intra-day pattern library categorized by "flood/dry season—daily electricity level". First, the historical intra-day output curves are normalized by daily total electricity to extract pure pattern features. Within each season, the historical days of each station and the runoff series are divided into three quantile-based levels according to their daily aggregate values. For each level, three representative normalized 24 h curves are extracted using K-means clustering (K = 3). During scenario generation, based on the flood/dry season category and electricity level of the obtained daily scale data, a typical curve is randomly selected from the corresponding library, and linearly scaled according to the actual daily total electricity to form a complete intra-day station-level output sequence.

6. Scenario Evaluation System

To systematically evaluate the effectiveness of the generated hydro–wind–solar joint scenario set, this paper constructs an evaluation system covering three dimensions: multi-energy correlation, spatial correlation, and fidelity of temporal fluctuation characteristics. These dimensions correspond to the modeling quality of the scenarios in three key aspects: multi-energy joint dependence, inter-station spatial structure, and intra-day temporal patterns [51,52].

6.1. Hydro–Wind–Solar Correlation Test

The Spearman correlation coefficient matrix error and Kendall correlation coefficient matrix error between historical data and the generated scenarios are calculated to evaluate the correlation of the generated runoff, wind power, and photovoltaic power output scenarios. The formula is as follows:

$$\text{Spearman RMSE} = \sqrt{\frac{1}{d^2}\sum_{i=1}^{d}\sum_{j=1}^{d}\left(R_{ij}^{\text{S, scen}} - R_{ij}^{\text{S, hist}}\right)^2} \tag{12}$$

$$\text{Kendall RMSE} = \sqrt{\frac{1}{d^2}\sum_{i=1}^{d}\sum_{j=1}^{d}\left(R_{ij}^{\text{K, scen}} - R_{ij}^{\text{K, hist}}\right)^2} \tag{13}$$

In the formula: d is the total number of objects being evaluated, and R_{ij} represents the rank correlation coefficient between objects i and j. The above indicators measure the overall deviation of the system's joint correlation structure in the form of root mean square error (RMSE). A smaller value indicates a more accurate reconstruction of the historical dependency structure by the generated scenarios.

6.2. Intra-Cluster Output Allocation Rationality Test

To evaluate the rationality of the intra-cluster output allocation scheme, this paper constructs an evaluation system from two dimensions: distribution fidelity and scenario diversity.

Distribution fidelity is evaluated using station shares of cluster-level daily output. For each station, the Wasserstein distance [53] between its generated and historical share

distributions is calculated, and the distances are then averaged across stations and clusters of the same energy type. A smaller value indicates closer agreement between the marginal station-share distributions.

$$D_{\mathrm{W}} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{m_k}\sum_{j=1}^{m_k}\int_0^1 \left|F_{k,j}^{s}(p) - F_{k,j}^{h}(p)\right|\mathrm{d}p \tag{14}$$

In the formula: K represents the number of clusters for the energy type being evaluated; m_k represents the number of stations in cluster k; p represents a station's share of its cluster-level daily output; ${F_{k,j}}^{s}(p)$ and ${F_{k,j}}^{h}(p)$ represent the empirical cumulative distribution functions of the generated and historical shares of station j in cluster k.

Scenario diversity is quantified by the ratio of the average Euclidean distance among generated scenarios to the average Euclidean distance among historical scenarios and is defined as the scenario diversity ratio.

$$R_{\mathrm{div}} = \frac{1}{N_s(N_s-1)}\sum_{i=1}^{N_s}\sum_{j=1|j\neq i|}^{N_s}\left\|P_i^s - P_j^s\right\|_2 \Big/ \frac{1}{N_h(N_h-1)}\sum_{i=1}^{N_h}\sum_{j=1|j\neq i|}^{N_h}\left\|P_i^h - P_j^h\right\|_2 \tag{15}$$

In the formula: N_h represents the number of historical data scenarios; P_i^h is the proportion vector of the *i-th* historical scenario.

6.3. Intra-Day Curve Reconstruction Fidelity Test

To evaluate the fidelity of the intra-day output curve reconstruction scheme, this paper constructs an evaluation system from two dimensions: temporal dynamic fidelity and peak-time error.

Temporal dynamic fidelity is measured by the autocorrelation coefficient error. The autocorrelation coefficient error is the absolute error of the first-order lag autocorrelation coefficient between the generated scenarios and the historical data. This indicator reflects the consistency of the evolution pattern of adjacent hourly output sequences. A smaller value indicates that the temporal smoothness and fluctuation pattern of the generated curve are closer to the historical reality. The formula is as follows:

$$E_{\mathrm{acf}} = \left|\frac{1}{N_s}\sum_{i=1}^{N_s}\rho_i^s - \frac{1}{N_h}\sum_{j=1}^{N_h}\rho_j^h\right| \tag{16}$$

In the formula: ρ_i^s represents the Pearson autocorrelation coefficient of the adjacent-hour output sequence of the *i-th* generated scenario; ρ_j^h represents the corresponding autocorrelation coefficient of the *j-th* historical scenario.

The peak-time error (h) is calculated as the absolute difference between the mean daily peak-hour indices of generated and historical curves. It therefore summarizes the difference in their central peak timing rather than matching individual days. Peak hour is inherently cyclic, and circular statistics are relevant when peak times straddle the midnight boundary [54]; the present arithmetic-mean statistic does not account for that wrap-around.

$$E_{\mathrm{peak}} = \left|\frac{1}{N_s}\sum_{i=1}^{N_s}t_i^s - \frac{1}{N_h}\sum_{j=1}^{N_h}t_j^h\right| \tag{17}$$

In the formula: t_i^s represents the time corresponding to the maximum output of the *i-th* generated scenario; t_j^h represents the time corresponding to the maximum output of the *j-th* historical scenario.

7. Case Study Analysis

This paper takes a certain area in Southwest China as the study area and collects historical output and runoff data from 2023 to 2025 for 47 photovoltaic (PV) power stations, 53 wind power stations, and the head hydropower station of a cascade hydropower station group in this region. According to local hydrological patterns, the period from June to October is defined as the flood season, and the remaining months as the dry season. Based on this, two independent joint scenario generation models are constructed for the flood season and the dry season, respectively. For renewable-energy bases whose output is delivered through LCC-HVDC links, interactions between converters and renewable plants can lead to temporary overvoltage during disturbances [55]. The present case study focuses on generation-side hydro–wind–solar uncertainty rather than a specific export corridor or converter transients. Its generated scenarios could provide generation-side inputs for future analyses that also account for power-transfer constraints and transmission-system dynamics.

Figure 3 illustrates how joint scenarios represent variations in hydro–wind–solar resource levels and temporal profiles while accounting for cross-resource dependence. These scenarios can provide generation-side uncertainty inputs for subsequent scheduling and risk analyses.



Figure 3. Potential applications of hydro–wind–solar joint scenarios.

7.1. Clustering Results and Dimensionality Reduction Effect

To reduce the complexity of high-dimensional joint modeling, cluster analysis based on Spearman correlation is performed on the 47 PV power stations and 53 wind power stations in the region. The results show that the average pairwise correlation coefficient among PV power stations is 0.891, while that among wind power stations is 0.636. On this basis, the performance of different cluster numbers K is evaluated using the silhouette coefficient. As shown in Figures 4 and 5, the optimal cluster number for PV power stations is K = 2, and for wind power stations, K = 3.

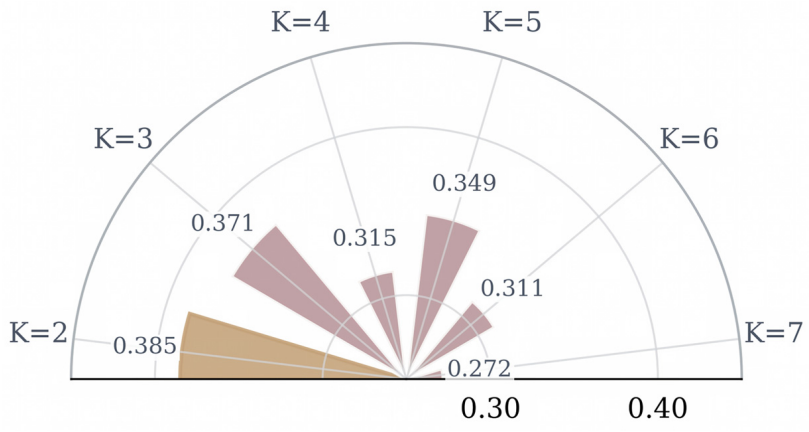


Figure 4. Silhouette coefficient corresponding to different cluster numbers K for PV power stations.

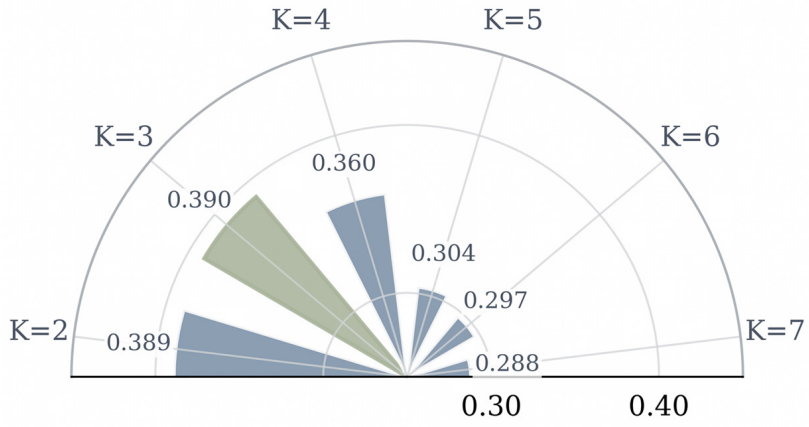


Figure 5. Silhouette coefficient corresponding to different cluster numbers K for wind power stations.

The clustering results show that the two clusters of PV power stations contain 14 and 33 stations, respectively, with intra-cluster average correlation coefficients of 0.836 and 0.915, demonstrating very strong synergistic output characteristics. The wind power clustering results include three clusters, containing 21, 16, and 16 stations, respectively, with intra-cluster average correlation coefficients of 0.759, 0.697, and 0.645, respectively.

This section compresses the original 100-dimensional renewable energy system into 5 cluster units, laying a low-dimensional, structured data foundation for subsequent joint modeling.

For the present case covering 101 objects, the complete workflow for generating 1000 joint scenarios for each season required approximately 70 s on an AMD Ryzen 7 7745HX processor (TSMC (Taiwan Semiconductor Manufacturing Company), Hsinchu, Taiwan) with 16 GB RAM. The framework can accommodate additional stations by updating the clustering, intra-cluster allocation models, and station-specific curve libraries. Its hierarchical structure retains station-level outputs without placing all station variables in a single daily joint-dependence model.

7.2. Hydro–Wind–Solar Joint Scene Generation

After clustering the renewable-energy stations and reducing the feature dimensionality, scenario generation proceeds in three stages. First, season-specific C-vine Copula models are constructed for the flood and dry seasons to generate cluster-level daily joint scenarios of hydro–wind–solar systems. Second, the centered log-ratio (CLR) transformation combined with a Gaussian R-vine Copula is used to allocate the daily total output of each cluster to its individual stations. Third, complete station-level intra-day output series are reconstructed

through pattern matching and linear scaling using a typical intra-day curve library classified by season and daily electricity level. Figures 6–8 present selected representative PV, wind-power, and runoff curves from the respective flood- and dry-season typical curve libraries. Finally, a high-dimensional joint scenario set containing 1000 scenarios and covering 101 objects is generated.

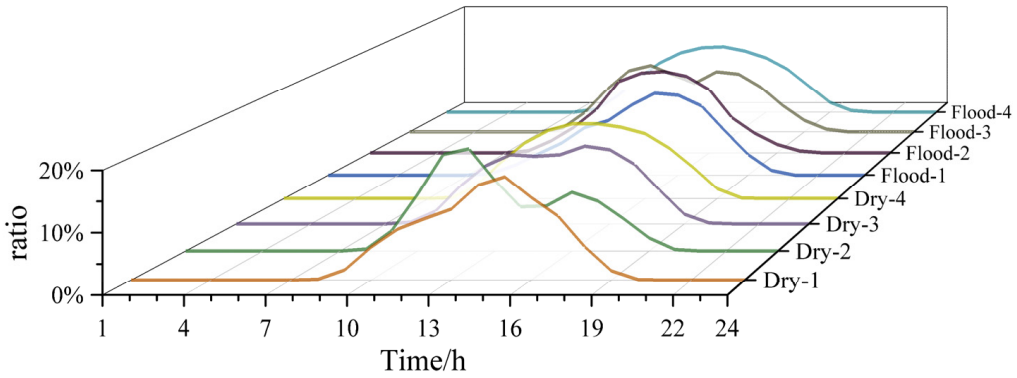


Figure 6. Typical daily curve database of PV power station during flood and dry season.

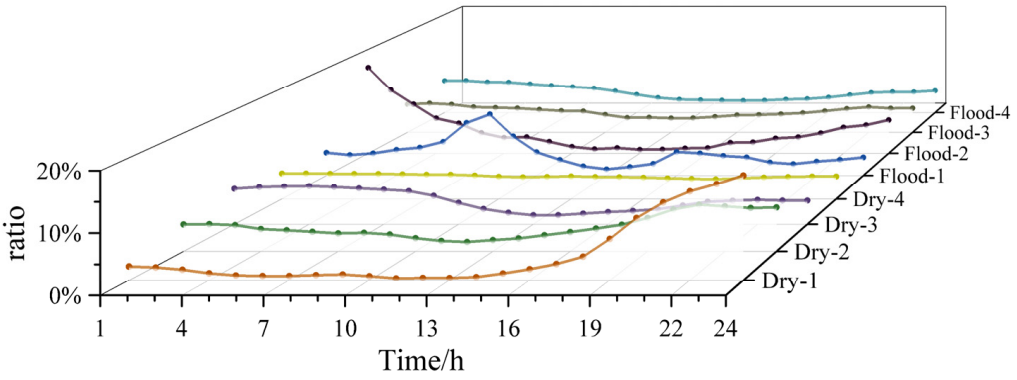


Figure 7. Typical daily curve database of wind power station during flood and dry season.

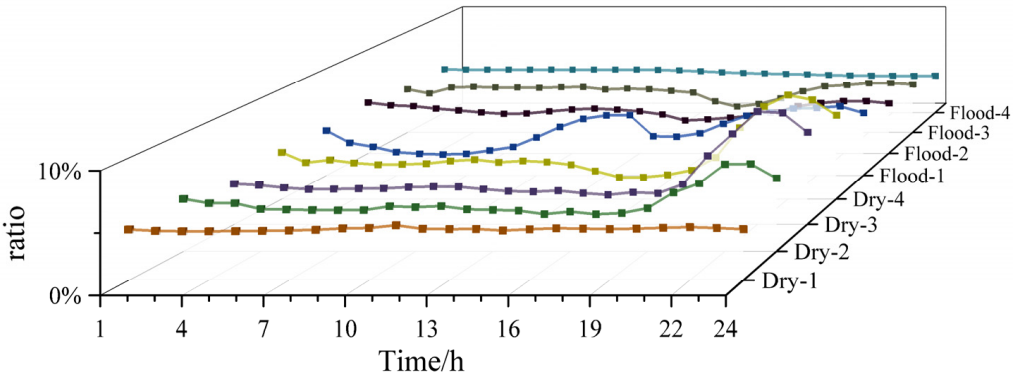


Figure 8. Typical daily curve database of runoff during flood and dry season.

7.3. Scene Assessment

To comprehensively assess the effectiveness of the proposed scenario generation framework, this paper conducts verification from three aspects: hydro–wind–solar correlation, station-level spatial allocation structure, and the fidelity of intra-day output patterns.

7.3.1. Hydro–Wind–Solar Correlation Test

To verify the effectiveness of the proposed model (the main model) for daily scale multi-energy joint dependency modeling, this paper constructs a control scheme (Model

I): at the cluster-level daily scale generation stage, the C-vine Copula joint modeling is removed and replaced with independent sampling based on the historical marginal distributions of each cluster; all other steps are exactly the same as the main scheme. By comparing the statistical characteristics of the generated scenarios of the two schemes, the contribution of the proposed model in terms of hydro–wind–solar correlation is tested.

The Spearman correlation coefficient matrix error and Kendall correlation coefficient matrix error between historical data and generated scenarios are calculated. The results are shown in Figure 9 and Table 1. During the flood season, the Spearman and Kendall errors of the main model are 0.395 and 0.298, respectively, which are better than those of Model I (0.432 and 0.325); during the dry season, the superiority is more significant, with the two errors decreasing from 0.419 and 0.306 to 0.318 and 0.236, respectively. These lower errors indicate that the generated daily scenarios retain more of the observed co-variation among wind power, PV power and runoff than independent sampling. This is relevant when examining conditions in which the three resources fluctuate together, although the remaining errors show that their historical dependence is not reproduced exactly.

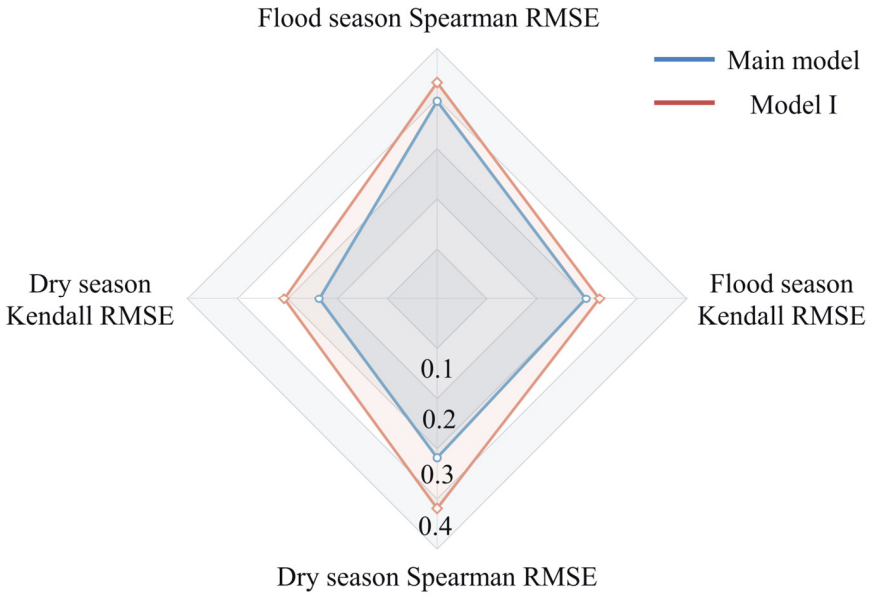


Figure 9. Hydro–wind–solar correlation test results.

Table 1. Hydro–wind–solar correlation test results.

	Model	Spearman RMSE	Kendall RMSE
Flood season	Main model	0.39	0.30
	Model I	0.43	0.33
Dry season	Main model	0.32	0.24
	Model I	0.42	0.31

To further assess the choice of the daily scale dependence model, the C-vine Copula was compared with a D-vine Copula and a multivariate Gaussian Copula using the same six variables—three wind-power clusters, two PV clusters, and runoff—and season-specific historical data. Each model generated 1000 joint scenarios in each of ten repeated simulations. Table 2 reports the mean Spearman and Kendall correlation errors. This comparison evaluates dependence reproduction at the cluster-level daily scale, before station-level allocation and intra-day reconstruction.

Table 2. Comparison of cluster-level daily dependence models.

Season	Copula Model	Mean Spearman Error	Mean Kendall Error
Flood	C-vine	0.0321	0.0233
Flood	D-vine	0.0343	0.0249
Flood	Gaussian	0.0352	0.0245
Dry	C-vine	0.0372	0.0278
Dry	D-vine	0.0341	0.0248
Dry	Gaussian	0.0358	0.0247

In the flood season, the C-vine Copula achieves the lowest mean Spearman and Kendall correlation errors among the three models. In the dry season, its errors are close to those of the D-vine and Gaussian Copulas, indicating comparable dependence reproduction. Overall, the C-vine provides competitive correlation fidelity across both seasons, while its root-centered structure offers an explicit and interpretable dependence ordering for the proposed framework.

To supplement the correlation assessment, joint low- and high-resource states were examined using the historical seasonal 10th and 90th percentiles of each resource's daily quantity, respectively. Daily wind and PV quantities were aggregated across their respective clusters, while runoff was considered separately. The same historical thresholds were applied to the main model and Model I. Joint occurrence frequencies were calculated for the wind–PV, wind–runoff, and PV–runoff combinations. For each season and resource state, the absolute differences between generated and historical frequencies were averaged across the three resource pairs. Table 3 presents the resulting mean absolute frequency errors.

Table 3. Mean absolute errors in joint low- and high-resource occurrence frequencies.

Season	Resource State	Main Model	Model I
Flood	Low	0.690	0.190
Flood	High	0.590	3.890
Dry	Low	0.340	0.979
Dry	High	0.336	1.331

As shown in Table 3, the main model achieves smaller occurrence-frequency errors in three of the four aggregated comparisons. In the dry season, the errors decrease from 0.979 to 0.340 percentage points for joint low-resource states and from 1.331 to 0.336 percentage points for joint high-resource states. The flood-season high-resource error also decreases substantially, from 3.890 to 0.590 percentage points. In contrast, the flood-season low-resource error increases from 0.190 to 0.690 percentage points. These results indicate closer reproduction of historical joint scarcity and abundance in most evaluated categories, while showing that the improvement varies with season and resource level.

7.3.2. Intra-Cluster Output Allocation Rationality Test

To assess the proposed intra-cluster output allocation method (the main model), Model II is constructed as a historical-share resampling benchmark. For each generated daily scenario, a complete station-share vector is randomly selected from a historical day in the same cluster and season and applied to the generated cluster-level daily total. The selected vector is fixed for that daily allocation, but different scenarios may use different historical vectors. All other steps remain the same as in the main model. The comparison therefore examines the effect of generating new share patterns through the proposed method rather than directly reusing observed patterns.

The Wasserstein distances and scenario diversity ratios of the two allocation schemes are reported in Tables 4 and 5 for the flood and dry seasons, respectively. Model II resamples complete historical station-share vectors, allowing its allocations to retain observed share patterns; accordingly, it achieves smaller Wasserstein distances in all four comparisons. The main model generates new allocation patterns but has larger Wasserstein distances, ranging from 0.032 to 0.047 for wind power and from 0.032 to 0.035 for PV power. These results indicate a greater difference from the historical marginal share distributions than for Model II. The Wasserstein distance should therefore be interpreted together with scenario diversity, rather than described on its own as evidence of superior distribution fidelity.

Table 4. Results of the intra-cluster output allocation rationality test in the flood season.

Energy Type	Model	Wasserstein Distance	Scenario Diversity Ratio
Wind power	Main model	0.047	3.81
	Model II	0.006	1.22
PV	Main model	0.035	4.72
	Model II	0.004	1.20

Table 5. Results of the intra-cluster output allocation rationality test in the dry season.

Energy Type	Model	Wasserstein Distance	Scenario Diversity Ratio
Wind power	Main model	0.032	4.74
	Model II	0.002	1.24
PV	Main model	0.032	4.02
	Model II	0.005	1.24

In the flood season, the main model achieves diversity ratios of 3.81 for wind power and 4.72 for PV power, compared with 1.22 and 1.20 for Model II. In the dry season, the corresponding ratios are 4.74 and 4.02 for the main model, whereas Model II remains at 1.24 for both resources. These results show that the main model produces a wider range of intra-cluster allocation patterns than Model II. Thus, scenarios with similar cluster-level daily totals can exhibit station-level contributions beyond the share vectors directly resampled from historical days. Because the historical diversity ratio is 1 by definition, its higher ratios indicate greater dispersion rather than necessarily better fidelity. Together with the larger Wasserstein distances, this represents a trade-off between scenario diversity and agreement with historical share distributions. Correlation errors and scenario diversity describe different properties: the former assess cross-resource dependence, whereas the latter reflects the range of station-level conditions available to downstream analyses. For example, learning-based DER coordination has been evaluated under multiple load scenarios [56], illustrating why exposure to varied operating conditions can matter. The diversity ratio is therefore interpreted together with distribution fidelity, rather than as a measure of downstream control performance.

7.3.3. Intra-Day Curve Reconstruction Fidelity Test

To verify the effectiveness of the proposed station-level intra-day curve modeling method (the main model), this paper constructs a control scheme (Model III): a unified typical intra-day pattern library is used for intra-day curve allocation, while all other steps are exactly the same as the main scheme. By comparing the statistical characteristics of the two schemes on the station-level intra-day output curves, the ability of the main scheme to characterize the intra-day curves of power stations is tested.

Tables 6 and 7 and Figures 10 and 11 present the intra-day time-series reconstruction results of the two methods in the flood and dry seasons. In terms of autocorrelation

error, the main model outperforms Model III for all resource types in both seasons. In the flood season, the first-order autocorrelation errors of wind power, PV power, and runoff decrease from 0.258, 0.126, and 0.623 to 0.108, 0.079, and 0.115, respectively. In the dry season, the corresponding errors decrease from 0.256, 0.107, and 0.584 to 0.140, 0.060, and 0.224. In particular, the main model provides the most pronounced improvement for runoff, reducing the error by approximately 81.5% and 61.6% in the flood and dry seasons, respectively. Because first-order autocorrelation describes dependence between adjacent hours, the lower errors indicate that the generated profiles more closely reproduce historical hour-to-hour persistence. Longer-lag dependence and the distribution of hourly changes are assessed separately in Table 8.

Table 6. Results of the intra-day time-series scenario test in the flood season.

Energy Type	Model	Autocorrelation Error	Peak-Time Error (h)
Wind power	Main model	0.108	1.885
	Model III	0.258	1.047
PV	Main model	0.079	0.591
	Model III	0.126	0.677
Runoff	Main model	0.115	5.364
	Model III	0.623	5.488

Table 7. Results of the intra-day time-series scenario test in the dry season.

Energy Type	Model	Autocorrelation Error	Peak-Time Error (h)
Wind power	Main model	0.140	1.829
	Model III	0.256	1.180
PV	Main model	0.060	0.539
	Model III	0.107	0.664
Runoff	Main model	0.224	0.539
	Model III	0.584	4.607

Table 8. Multi-lag autocorrelation errors and hourly ramp distribution distances in the flood and dry seasons.

Season	Energy Type	Model	Lag 1	Lag 2	Lag 4	Lag 8	Ramp Distance
Flood	Wind power	Main model	0.108	0.233	0.266	0.095	3.106
Flood	Wind power	Model III	0.258	0.079	0.122	0.113	5.376
Flood	PV	Main model	0.079	0.080	0.026	0.087	1.857
Flood	PV	Model III	0.126	0.102	0.057	0.122	1.794
Flood	Runoff	Main model	0.115	0.218	0.279	0.288	123.533
Flood	Runoff	Model III	0.623	0.448	0.226	0.004	1155.930
Dry	Wind power	Main model	0.140	0.248	0.303	0.133	4.195
Dry	Wind power	Model III	0.256	0.078	0.121	0.129	15.392
Dry	PV	Main model	0.060	0.059	0.019	0.062	1.463
Dry	PV	Model III	0.107	0.084	0.030	0.095	2.086
Dry	Runoff	Main model	0.224	0.303	0.121	0.038	14.706
Dry	Runoff	Model III	0.584	0.345	0.041	0.087	303.535

To further assess intra-day temporal fidelity, autocorrelation errors at lags of 1, 2, 4 and 8 h and the Wasserstein distances between hourly ramp distributions were calculated (Table 8). The main model achieves lower autocorrelation errors in 16 of the 24 season–resource–lag comparisons and lower ramp-distribution distances in five of the six season–resource comparisons. Although it does not outperform Model III in every individual comparison, these results support its better overall representation of intra-day temporal behavior.

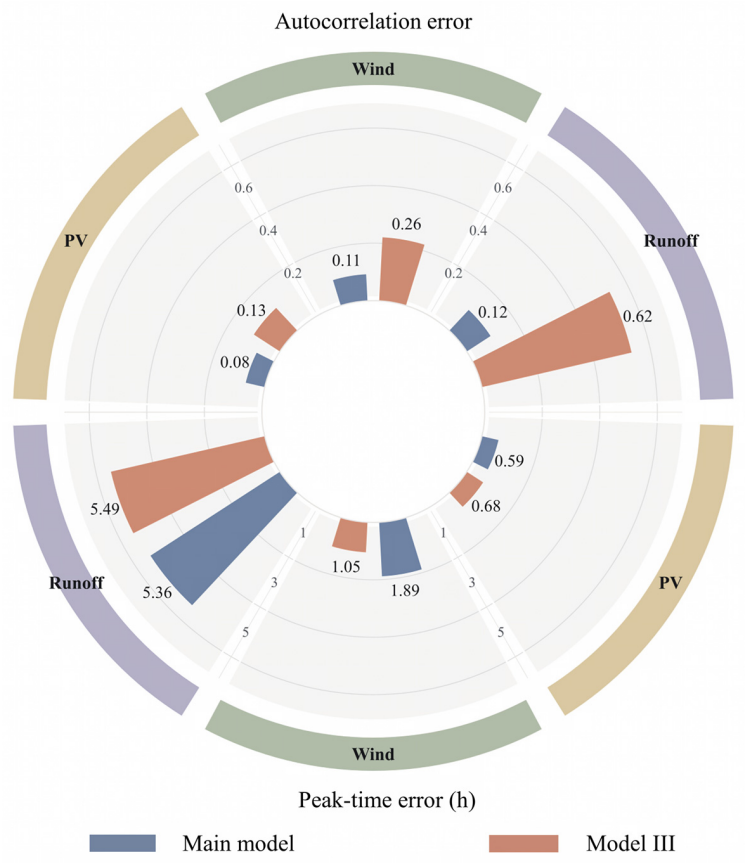


Figure 10. Intra-day temporal fidelity evaluation in the flood season.

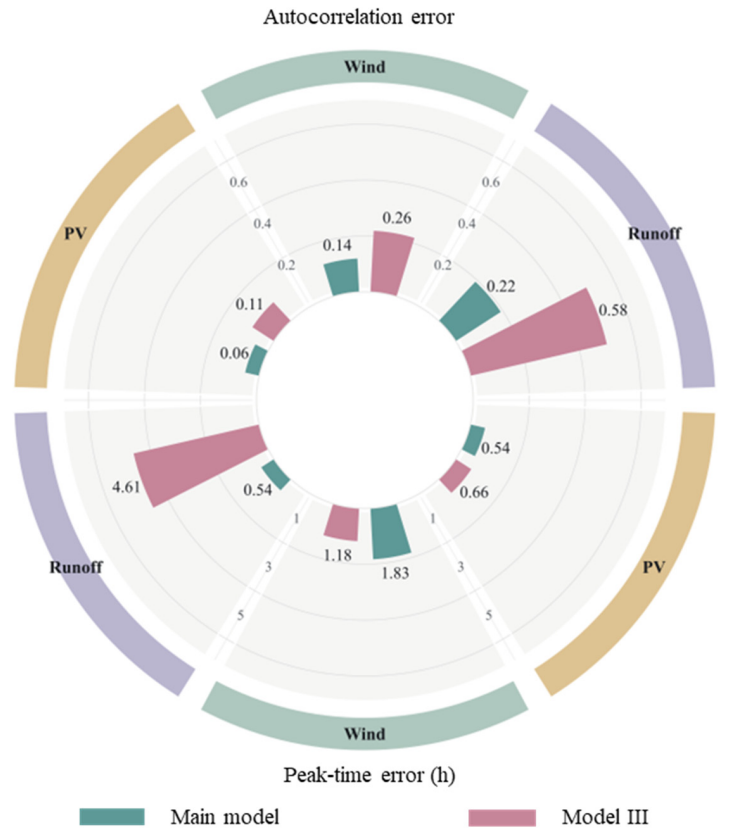


Figure 11. Intra-day temporal fidelity evaluation in the dry season.

In terms of peak-time error, the main model shows a stable advantage in reproducing PV peak positions. The PV peak-time errors decrease from 0.677 h and 0.664 h to 0.591 h and 0.539 h in the flood and dry seasons, respectively. For runoff, the main model substantially reduces the dry-season peak-time error from 4.607 h to 0.539 h. In the flood season, the peak-time errors of the main model and Model III are 5.364 h and 5.488 h, respectively, indicating comparable performance. For wind power, however, the main model has higher peak-time errors than Model III in both seasons (1.885 versus 1.047 h in the flood season and 1.829 versus 1.180 h in the dry season). This indicates a remaining limitation in wind peak-time reconstruction.

Taken together, these results indicate that the hierarchical curve library generally improves the representation of within-day persistence and hourly changes, rather than only reducing a single autocorrelation statistic. Peak-time accuracy is more resource-dependent, as shown by the different results for PV power, runoff and wind power.

A five-repeat stratified 80/20 holdout check was conducted to examine the sensitivity of the runoff curve library to the number of typical curves. Increasing K from 2 to 3 reduced the mean nearest-centroid error on held-out normalized profiles from 0.00666 to 0.00603 in the flood season and from 0.01080 to 0.01014 in the dry season; K = 4–5 produced smaller additional improvements. These results support (K = 3) as a parsimonious choice, with no deterioration in holdout curve coverage observed at the selected value.

8. Conclusions

To address high-dimensional uncertainty scenario generation in large-scale hydro–wind–solar systems, this study proposes a multi-scale correlation-preserving scenario synthesis framework based on hierarchical dependency clustering and validates it using a hydro–wind–solar system in Southwest China. The main conclusions are as follows:

(1) Based on Spearman correlations, this paper clusters wind and PV power stations, compressing the original 100-dimensional renewable energy system into 5 cluster units. While effectively preserving the main spatial correlation structure, this significantly reduces the complexity of subsequent joint modeling, providing a structured, low-dimensional data foundation for high-dimensional hydro–wind–solar scenario generation.

(2) Season-specific C-vine copulas are used to model the daily joint dependence among wind-power clusters, PV clusters and runoff. Compared with independent sampling, the main model reduces the Spearman and Kendall correlation-matrix errors to 0.395 and 0.298 in the flood season, and 0.318 and 0.236 in the dry season, respectively. These results indicate improved, although not exact, reproduction of the historical cross-resource dependence structure.

(3) This paper introduces the CLR transformation to handle the closure constraint of compositional data and combines it with a Gaussian R-vine Copula to achieve intra-cluster spatial allocation. Compared with historical share-vector resampling, the proposed method generates more diverse allocation patterns. The larger Wasserstein distances indicate a trade-off between scenario diversity and agreement with historical share distributions, and the two indicators are therefore considered jointly when assessing the allocation results.

(4) This paper constructs a typical intra-day curve library categorized by “flood/dry season—daily electricity level” and restores intra-day station-level time-series scenarios through curve matching and linear scaling. The results show that the proposed method yields lower first-order autocorrelation errors for all three resources and lower peak-time errors for PV and dry-season runoff than the unified typical curve method, while wind peak-time errors remain higher.

(5) Compared with independent sampling, the proposed framework reduces the seasonal Spearman and Kendall correlation-matrix errors by 8.7–24.2% and 8.3–22.8%, respec-

tively. Preserving these relationships helps the scenario set represent plausible combinations of wind power, PV generation, and runoff, which are relevant to downstream scheduling and risk analysis. Recent typhoon-resilience research considers weather-affected operating conditions and distributed-resource flexibility [57]; integrating the generated scenarios with hazard and network models could therefore be explored in future resilience studies.

The proposed framework achieves multi-scale scenario synthesis, progressing from cluster-level daily joint scenarios to station-level intra-day time-series scenarios. It provides high-fidelity scenario inputs for stochastic scheduling, operational risk analysis, and flexible-resource optimization in large-scale hydro–wind–solar systems.

The proposed framework can be applied to other hydro–wind–solar regions when suitable historical station-output and runoff data are available. The overall modeling procedure remains the same as that presented in this study, while station clustering, seasonal classification, marginal distributions, vine-copula models, intra-cluster allocation, and the typical daily curve library should be established using local data. In particular, the flood- and dry-season definitions should reflect local hydrological conditions. Thus, applying the framework elsewhere follows the same modeling steps rather than directly reusing the models and parameters fitted for Southwest China. Its performance should be evaluated according to the resource and climatic conditions of the target region.

Author Contributions: Conceptualization, C.Z. and Y.J.; methodology, C.Z., Z.X. and Y.W.; software, C.Z. and Z.X.; validation, Z.X., Y.W. and K.C.; formal analysis, C.Z., Y.W. and J.S.; investigation, C.Z. and Y.J.; resources, K.C.; data curation, Z.X. and Y.J.; writing—original draft preparation, C.Z.; writing—review and editing, Y.W., Z.X., Y.J., K.C. and J.S.; visualization, C.Z. and Y.W.; supervision, J.S. and K.C.; project administration, Y.J.; funding acquisition, Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: This achievement was funded by the Science and Technology Project of Yunnan Electric Power Grid Co. Ltd. (grant no. 05000020250303010500066) and National Natural Science Foundation of China (grant no. U23A20667).

Data Availability Statement: The original data used in this study cannot be publicly shared due to confidentiality restrictions. However, the processed data, analysis results, and related derived datasets can be provided upon request. We are happy to share these when needed.

Conflicts of Interest: The authors declare that this study received funding from Yunnan Electric Power Grid. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

1. Wang, Y.; Wang, R.; Tanaka, K.; Ciais, P.; Peñuelas, J.; Balkanski, Y.; Sardans, J.; Hauglustaine, D.; Liu, W.; Xing, X.; et al. Accelerating the energy transition towards photovoltaic and wind in China. *Nature* **2023**, *619*, 761–767. [CrossRef]
2. Zhu, Y.; Chen, S.; Huang, W.; Wang, L.; Ma, G. Complementary operational research for a hydro-wind-solar hybrid power system on the upper Jinsha River. *J. Renew. Sustain. Energy* **2018**, *10*, 043309. [CrossRef]
3. Li, H.; Zhao, L.; Yang, Z.; Liu, P.; Ming, B.; Liu, B. Mitigating uncertainties: Robust stochastic optimal operation of inter-regional hydro-wind-solar hybrid systems. *Renew. Energy* **2026**, *261*, 125230. [CrossRef]
4. Shen, J.; Wang, Y.; Lin, M.; Cheng, C.; Kazak, J.K.; Wang, J.; Guo, X.; Li, X.; Zhou, B.; Ge, L. Quantifying the impact of extreme weather on China's hydropower–wind–solar renewable energy system. *Nat. Water* **2025**, *3*, 415–429. [CrossRef]
5. Yuan, K.; Zhang, T.; Xie, X.; Du, S.; Xue, X.; Abdul-Manan, A.F.N.; Huang, Z. Exploration of low-cost green transition opportunities for China's power system under dual carbon goals. *J. Clean. Prod.* **2023**, *414*, 137590. [CrossRef]
6. Karimi, H.; Jadid, S. Sustainable and reliability based coalition forming model for smart multi-microgrid systems considering battery and water storage systems. *J. Energy Storage* **2024**, *102*, 114050. [CrossRef]
7. Karimi, H.; Heydarian-Forushani, E.; Mehri, S.; Salman, M.; Ghandour, R.; Zayyani, H. A bi-level multi-energy scheduling model for virtual power plants in competitive market environment. *Electr. Power Syst. Res.* **2026**, *254*, 112612. [CrossRef]

8. Deng, X.; Lv, T. Power system planning with increasing variable renewable energy: A review of optimization models. *J. Clean. Prod.* **2020**, *246*, 118962. [CrossRef]
9. Ma, C.; Liu, L. Optimal capacity configuration of hydro-wind-PV hybrid system and its coordinative operation rules considering the UHV transmission and reservoir operation requirements. *Renew. Energy* **2022**, *198*, 637–653. [CrossRef]
10. Cheng, Q.; Liu, P.; Xia, J.; Ming, B.; Cheng, L.; Chen, J.; Xie, K.; Liu, Z.; Li, X. Contribution of complementary operation in adapting to climate change impacts on a large-scale wind–solar–hydro system: A case study in the Yalong River Basin, China. *Appl. Energy* **2022**, *325*, 119809. [CrossRef]
11. Tian, Y.; Chang, J.; Wang, Y.; Wang, X.; Meng, X.; Guo, A. The capacity planning method for a hydro-wind-PV-battery complementary system considering the characteristics of multi-energy integration into power grid. *J. Clean. Prod.* **2024**, *446*, 141292. [CrossRef]
12. Sun, J.; Su, C.; Jurasz, J. Risk-aware day-ahead and real-time scheduling of wind-solar-pumped hydro storage hybrid energy systems based on two-stage stochastic programming. *Sustain. Energy Grids Netw.* **2026**, *46*, 102311. [CrossRef]
13. An, T.; Xu, W.; Hu, R.; Gao, D.; Cheng, C.; Gao, Y.; Yang, J. Integrated Energy System in the Context of Carbon Neutrality: A Review of Typical Structures and Key Technologies. *Processes* **2026**, *14*, 1711. [CrossRef]
14. Su, C.; Li, L.; Zhang, T.; Sui, Q.; Yang, Y. A Real-Time Scheduling Framework of Cascade Hydropower-Photovoltaic Power Complementary Systems Based on Model Predictive Control. *Appl. Energy* **2025**, *392*, 126023. [CrossRef]
15. Tamimi, M. A Linearized Optimal Power Flow Framework for Locational Marginal Pricing and Risk Management in Wind-Integrated Power Systems with Disaggregated Flexible and Contingency Reserve Modeling. *Power Control Data Process. Syst.* **2026**, *3*, 37–45. [CrossRef]
16. Akinyele, D.; Olabode, O. Recent Advances in Green and Low-Carbon Energy Resources: Navigating the Climate-Friendly Microgrids for Decarbonized Power Generation. *Processes* **2025**, *13*, 3028. [CrossRef]
17. Li, W.; Dong, F.; Liu, J.; Wang, P.; Zhao, X. A three-stage decision-making study on capacity configuration of hydropower-wind-photovoltaic-storage complementary systems considering uncertainty. *Energy* **2024**, *313*, 134007. [CrossRef]
18. Guo, Y.; Ming, B.; Huang, Q.; Liu, P.; Wang, Y.; Fang, W.; Zhang, W. Evaluating effects of battery storage on day-ahead generation scheduling of large hydro–wind–photovoltaic complementary systems. *Appl. Energy* **2022**, *324*, 119781. [CrossRef]
19. Liu, Z.; He, X. A renewable tango for drought-stricken power grids. *Nat. Water* **2026**, *4*, 545–547. [CrossRef]
20. Zhang, L.; Xie, J.; Chen, X.; Zhan, Y.; Zhou, L. Cooperative Game-Based Synergistic Gains Allocation Methods for Wind-Solar-Hydro Hybrid Generation System with Cascade Hydropower. *Energies* **2020**, *13*, 3890. [CrossRef]
21. Gu, Z.; Pan, G.; Gu, W.; Zhou, S.; Wu, Z.; Lu, S. Assessment and Prospect of Region Joint Electrolytic Hydrogen Systems Considering Multiple Energy Sources: Wind, Solar, Hydro and Thermal Power. *IEEE Trans. Ind. Appl.* **2023**, *59*, 5269–5282. [CrossRef]
22. Shademan, M.; Karimi, H.; Jadid, S. Safe resource management of non-cooperative microgrids based on deep reinforcement learning. *Eng. Appl. Artif. Intell.* **2023**, *126*, 106865. [CrossRef]
23. Najafi, A.; Vanak, A. An Accurate and Computationally Efficient Unit Commitment Formulation for Renewable-Integrated Power Systems. *Power Control Data Process. Syst.* **2026**, *3*, 46–56. [CrossRef]
24. Moon, J. Learning from Scarcity: A Review of Deep Learning Strategies for Cold-Start Energy Time-Series Forecasting. *Comput. Model. Eng. Sci.* **2026**, *146*, 2. [CrossRef]
25. Mollasalehi, A.; Farhadi, A. Solar and Wind Power Forecasting: A Comparative Review of LSTM, Random Forest, and XGBoost Models. *arXiv* **2025**, arXiv:2509.24059. [CrossRef]
26. Ferdaus, M.M.; Dam, T.; Sarkar, M.R.; Uddin, M.; Anavatti, S.G. Foundation models for clean energy forecasting: A comprehensive review. *Renew. Sustain. Energy Rev.* **2026**, *226*, 116452. [CrossRef]
27. Ye, S.; Shi, G.; Yang, X.; Han, Y.; Chen, S.; Jiang, D.; Zhang, Y.; Liu, X. Robust Analysis and Optimal Control of Flexible Interconnected Microgrids Considering Wind and Solar Uncertainty. *Processes* **2026**, *14*, 1679. [CrossRef]
28. Wan, A.; He, J.; Zhou, X.; Al-Bukhaiti, K.; Zhu, P. TSBMNet-XAI interpretable multi-step offshore wind power forecasting for enhanced energy storage integration. *Comput. Electr. Eng.* **2026**, *136*, 111230. [CrossRef]
29. Yang, Z.; Yan, J.; Hong, Y.; Yang, J.; Wang, K.; Li, Y. Scenario Generation of PV Power Based on DDPM. In *Proceedings of the 2023 IEEE 7th Conference on Energy Internet and Energy System Integration (EI2), Hangzhou, China, 20–22 October 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 2062–2066. [CrossRef]
30. Zhou, Y.; Wang, K.; Huang, X.; Xu, Y.; Wan, X.; Li, W. A Renewable Power Scenario Generation Method Based on Wasserstein Distance for Conditional Generative Adversarial Networks. In *Proceedings of the 2025 10th Asia Conference on Power and Electrical Engineering (ACPEE), Beijing, China, 22–24 March 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 395–399. [CrossRef]
31. Shahirinia, A.; Farahmandfar, Z.; Bina, M.T.; Hafezi, H.; Tanoe, V. Multi-Site Wind Farms Dependence Structure Using Vine Copulas: Impacts of Dataset Sizes and Employed Copulas. *IEEE Access* **2025**, *13*, 79596–79608. [CrossRef]
32. Tan, J.; Zhang, J.; Liu, H.; Lan, B. A high dimensional uncertain scenario generating method for wind power and photovoltaic considering spatiotemporal correlation. *Energy* **2025**, *340*, 139224. [CrossRef]

33. Han, J.; Liu, C.; Dong, X. A two-degree-of-freedom elastic energy equivalent system for evaluating interactions in power systems with multiple DFIG-based wind farms. *IEEE Trans. Sustain. Energy* **2026**, *17*, 2066–2080. [CrossRef]
34. Iung, A.M.; Oliveira, F.L.C.; Pereira, G.A.d.A.; Marcato, A.L.M. A copula-based framework for joint hydro-wind-solar scenario generation. In *Proceedings of the 2025 60th International Universities Power Engineering Conference (UPEC), London, UK, 2–5 September 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–7. [CrossRef]
35. Liu, Z.; Mo, L.; Zhang, M.; Kang, J.; Liu, W.; Sun, X.; Xiao, W. High-dimensional scenario generation method joint-driven by multiple correlations for hydro-wind-photovoltaic. *Appl. Energy* **2025**, *400*, 126551. [CrossRef]
36. Feng, Z.; Wang, X.; Niu, W.; Li, J.; Zhang, J.; Bai, T. Two-stage scenario generation of hydro-wind-solar complementary system based on improved variational autoencoder and generative adversarial networks model. *Renew. Energy* **2026**, *262*, 125358. [CrossRef]
37. Janse van Vuuren, C.Y.; Vermeulen, H.J. Clustered Wind Resource Domains for the South African Renewable Energy Development Zones. In *Proceedings of the 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), Bloemfontein, South Africa, 28–30 January 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 616–623. [CrossRef]
38. Halili, M.; Muda, V.; Mitrushi, D.; Veshaj, F.; Tabaku, A. The complementarity of wind and solar energy in Karabrun, Vlora, Albania. *Knowl. Int. J.* **2024**, *66*, 313–317. Available online: https://ojs.ikm.mk/index.php/kij/article/view/7031 (accessed on 21 October 2026).
39. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
40. Aas, K.; Czado, C.; Frigessi, A.; Bakken, H. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **2009**, *44*, 182–198. [CrossRef]
41. Hanebeck, A.; Şahin, Ö.; Havlíčková, P.; Czado, C. Sampling from Conditional Distributions of Simplified Vines. *Stat. Comput.* **2025**, *35*, 128. [CrossRef]
42. Demarta, S.; McNeil, A.J. The t copula and related copulas. *Int. Stat. Rev.* **2005**, *73*, 111–129. [CrossRef]
43. Czado, C. Vine copula based structural equation models. *Comput. Stat. Data Anal.* **2025**, *203*, 108076. [CrossRef]
44. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]
45. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
46. Nocedal, J.; Wright, S.J. *Numerical Optimization*, 2nd ed.; Springer: New York, NY, USA, 2006. [CrossRef]
47. Aitchison, J. The statistical analysis of compositional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1982**, *44*, 139–160. [CrossRef]
48. Zhang, K.; Yin, K.; Yang, W. Predicting bioenergy power generation structure using a newly developed grey compositional data model: A case study in China. *Renew. Energy* **2022**, *198*, 695–711. [CrossRef]
49. Egozcue, J.J.; Pawlowsky-Glahn, V.; Mateu-Figueras, G.; Barceló-Vidal, C. Isometric logratio transformations for compositional data analysis. *Math. Geol.* **2003**, *35*, 279–300. [CrossRef]
50. Dißmann, J.; Brechmann, E.C.; Czado, C.; Kurowicka, D. Selecting and estimating regular vine copulae and application to financial returns. *Comput. Stat. Data Anal.* **2013**, *59*, 52–69. [CrossRef]
51. Pinson, P.; Girard, R. Evaluating the quality of scenarios of short-term wind power generation. *Appl. Energy* **2012**, *96*, 12–20. [CrossRef]
52. Gneiting, T.; Stanberry, L.I.; Grimit, E.P.; Held, L.; Johnson, N.A. Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *TEST* **2008**, *17*, 211–235. [CrossRef]
53. Panaretos, V.M.; Zemel, Y. Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **2019**, *6*, 405–431. [CrossRef]
54. Mardia, K.V.; Jupp, P.E. *Directional Statistics*; Wiley: Chichester, UK, 1999. [CrossRef]
55. Liu, X.; Huang, J.; Zheng, D.; Xin, H.; Bi, T. Analysis and Suppression for Temporary Overvoltage Considering Dynamic Interactions Between LCC-HVDC and Renewable Energy Plants. *IEEE Trans. Sustain. Energy* **2025**, *16*, 1849–1859. [CrossRef]
56. Imtiaz, S.; Yang, L.; Naz, M.N.; Munir, H.M.; Altaf, M.W.; Mohammed, M.S.A. Enhanced TD3 multi-agent reinforcement learning approach for distributed energy resources coordination in active distribution network. *Appl. Soft Comput.* **2026**, *199*, 115348. [CrossRef]
57. Wang, X.; Ke, J.; Yang, H.; Gong, K.; Zhang, K.; Shahidehpour, M.; Jiang, C.; Huang, W. Aggregated Flexibility of DERs for Enhancing the Power Grid Resilience Considering Frequency Stability Constraints in Typhoons. *IEEE Trans. Ind. Appl.* **2026**, *62*, 6043–6055. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.