

Article

Cross-Correlated Scenario Generation for Renewable-Rich Power Systems Using Implicit Generative Models

Dhaval Dalal ¹, Muhammad Bilal ¹ , Hritik Shah ¹, Anwarul Islam Sifat ¹ , Anamitra Pal ^{1,*} 
and Philip Augustin ²

¹ School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85281, USA

² Salt River Project (SRP), 6504 East Thomas Road, Scottsdale, AZ 85251, USA

* Correspondence: anamitra.pal@asu.edu

Abstract: Generation of realistic scenarios is an important prerequisite for analyzing the reliability of renewable-rich power systems. This paper satisfies this need by presenting an end-to-end model-free approach for creating representative power system scenarios on a seasonal basis. A conditional recurrent generative adversarial network serves as the main engine for scenario generation. Compared to prior scenario generation models that treated the variables independently or focused on short-term forecasting, the proposed implicit generative model effectively captures the cross-correlations that exist between the variables considering long-term planning. The validity of the scenarios generated using the proposed approach is demonstrated through extensive statistical evaluation and investigation of end-application results. It is shown that analysis of abnormal scenarios, which is more critical for power system resource planning, benefits the most from cross-correlated scenario generation.

Keywords: dynamic time warping; generative adversarial network; power system planning; renewable energy; scenario generation



Citation: Dalal, D.; Bilal, M.; Shah, H.; Sifat, A.I.; Pal, A.; Augustin, P. Cross-Correlated Scenario Generation for Renewable-Rich Power Systems Using Implicit Generative Models. *Energies* **2023**, *16*, 1636. <https://doi.org/10.3390/en16041636>

Academic Editor: Germano Lambert-Torres

Received: 10 January 2023

Revised: 24 January 2023

Accepted: 2 February 2023

Published: 7 February 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Reliability planning for the transmission system of the electric power system is essential in keeping the grid operational in times of high uncertainty/variability. Traditionally, this exercise involved managing only one variable—load. The scenarios used to evaluate the system operation and resilience were worst-case loading scenarios derived from historical data with some growth projections. However, globally coordinated initiatives for carbon emission reduction have led to increased emphasis on planning integrated energy systems that feature rapid growth in renewable generation (RG), energy efficiency, and high electrification rates [1–3]. Particularly, due to the increasing proliferation of large RG sources, power system resource planning studies must include additional variables, viz., solar and wind generation. The variability associated with these new variables, which depends on weather conditions, such as solar irradiation and/or wind speed, makes reliability evaluation of renewable-rich power systems a more complex and challenging problem [4–7]. To address this problem, power system planners create *synthetic scenarios* that are aimed at capturing actual system conditions [8–10]. Two strategies that have been extensively used for creating synthetic scenarios are: *classical techniques*, which try to fit a model onto the distribution and then attempt to generate scenarios from the fitted model, and *machine learning approaches*, which learn the distribution from large amounts of historical data and are then able to produce similar scenarios. A brief overview of these two strategies is provided below.

The classical techniques typically rely on probabilistic modeling to generate new scenarios. These include methods that employ Latin hypercube sampling (LHS) [11], generalized dynamic factor model (GDFM) [12], generalized auto-regressive score (GAS) models [13], vine copula methods [14], principal component analysis (PCA) [15], and generalized Gaussian mixture models (GGMM) [16], amongst others. However, despite their

complexity, these models cannot fully capture all the correlations between the variables. Consequently, many *classical* studies still focus on one variable at a time [17]. As the power systems become more complex, it will become increasingly hard to extract models that capture all of the system's characteristics using probabilistic methods.

Machine learning models offer flexibility and versatility when generating new scenarios. Particularly, neural-network-based approaches eliminate the need for the extraction of relevant features from the available data. In [18,19], a simple single-layered neural network and radial basis function network (RBFN) were used to forecast wind power ramp-up events and distributions, respectively. However, complex tasks, such as multivariate scenario generation that is being considered in this paper, would require more complex (deeper) architectures.

In recent times, generative adversarial networks (GANs) [20] have emerged as a popular deep-learning algorithm for scenario generation. This implicit generative model is capable of transforming raw noise into meaningful information. Therefore, it can work on a variety of datasets, such as two-dimensional images and one-dimensional time-series data. Furthermore, it can generate samples that replicate the ones available in the data and other, more varied samples *not present* in the original dataset. The performance of a stand-alone GAN has been improved for scenario generation applications by using a hybrid model strategy, tweaking the error function, and/or adding appropriate conditions. For example, recurrent neural networks (RNN) with long short-term memory (LSTM) and reinforcement learning algorithms were added to the GAN model to produce wind power generation scenarios in [21]. This hybrid model strategy was tested on two case studies, and it created varied and believable scenarios in both of them. Similarly, a Wasserstein distance-based error function was embedded into convolutional GANs to improve performance in [22]. This approach was extended to condition-based solar and wind power scenario generation in [23,24]. These models accurately predicted wind ramp events and peak values but treated the variables (namely, solar and wind), independently.

Since the power system is a complex network of interconnected generation sites and electricity consumers, both residential and large-scale, the relationship between the different modes of power production and the nuances of load demand (i.e., their correlations) must be systematically considered. In line with this realization, a convolutional GAN with an LSTM-based sequence encoder was proposed in [25] to perform day-ahead forecasting of correlated photovoltaic (PV) and wind production sequences from meteorological data. In [26], correlated scenarios were generated to determine the most cost-effective generation procedure for optimizing a large-scale hydro–wind–solar hybrid system. Correlated GANs were also used in the cost-optimal scheduling of a battery energy storage system (BESS) to increase the BESS-PV system's incentive revenue [27]. Although GAN-based architectures have been applied to generate correlated scenarios for power systems, the *cross-correlation between RG and load* has not been well-explored. In addition, the scenario-generation techniques developed in [25–27] only focused on generating short-term forecasting scenarios.

To better facilitate long-term reliability planning, there is a genuine need to capture the cross-correlation present between the variables (RG and loads) while creating representative scenarios. The interdependence between these variables occurs naturally in the historical data. However, if the variables are treated independently during the scenario generation process, there is a risk of losing this interdependence and generating less meaningful scenarios. Particularly, under abnormal conditions, ignoring these cross-correlations and using the independent scenario generation approach can result in grossly misleading outcomes (see Section 4). At the same time, note that incorporating cross-correlation during multivariate scenario generation is a more challenging task. To accomplish this task, a sophisticated implicit generative model is proposed, as explained below.

1.1. Major Contributions

In this paper, a conditional recurrent GAN is proposed to generate cross-correlated scenarios on a seasonal basis. The labeling of the historical data for GAN training is an

important aspect of the methodology. In the presence of multiple variables, the determination of normal/abnormal days is not straightforward. Instead of relying on normal and abnormal labels assigned based on visual inspection, a data-driven technique is developed to create seasonal labels. Then, based on the labels, normal and abnormal day assignments are made for each season. Along with cross-correlation, this approach also captures the temporal correlations present in the time-series data of each variable.

For the generation of statistically similar but distinct correlated scenarios, the conditional recurrent GAN is modeled with the use of RNN-LSTM. An RNN-LSTM incorporated GAN is able to better process and reproduce the long-term modalities and temporal aspects in time-series data compared to a conventional GAN which does not consider these properties. Exploiting the temporal modeling capabilities of RNN-LSTMs along with the latent conditional feature modeling power of label-incorporated GANs helps enhance the relevance of the generated scenarios for different end-applications.

An extensive validation of the proposed approach is also provided in which correlated scenarios are compared against uncorrelated scenarios for an actual power system application, namely, optimal power flow (OPF). For normal conditions, uncorrelated synthesis of scenarios has a performance similar to correlated scenarios. However, during abnormal conditions, the results obtained using correlated scenario generation are more realistic than those obtained using uncorrelated scenario generation.

In summary, the novel contributions of this paper are as follows:

- Creation of a fine-tuned cross-correlated conditional recurrent GAN (C²RGAN) for multivariate scenario generation. This implicit generative model is scalable and yields relevant abnormal scenarios to augment limited historical data.
- Formulation of a data-driven labeling process for historical data to eliminate the subjectivity associated with manual labeling.
- Demonstration of the validity of correlated scenario generation for the power system OPF application in terms of cost and voltage angle distribution.

1.2. Paper Organization and Key Terms

Some of the salient terms used in this paper are explained here to provide the appropriate context.

- *Normal day* refers to a day that follows the typical seasonal pattern.
- *Abnormal day* refers to a day where one of the variables (RG and/or load) deviates significantly from the typical seasonal pattern. This is different from an abnormal operating condition/event that typically refers to line faults and sudden or unexpected load-shed/generator shut-down.
- *Scenario generation* refers to creation of *representative scenarios* for long-term resource planning. This is different from scenario forecasting, which is typically used for short-term day-ahead planning.
- *Cross-correlated scenarios*, one of the main contributions of this paper, refer to those *representative scenarios* that capture the inherent correlations between the variables. Implicit generative models are employed to extract these correlations.

The rest of the paper is structured as follows. Section 2 presents relevant insights drawn from the data-driven label assignment of the historical data used for the analysis conducted here. Section 3 provides a detailed look into the GAN architecture, selection, design, training, and implementation. Section 4 delves into an extensive analysis of the results obtained using the proposed method and their comparison with the uncorrelated scenario generation results. The conclusion is provided in Section 5.

2. Data-Driven Label Assignment

A two-variable (load and solar generation) dataset was employed in this research. The primary requirement is to generate cross-correlated labels for normal and abnormal days in the dataset so that the C²RGAN can be trained conditionally. Additionally, it

was determined that the seasonal variations in load and solar generation are significant enough to warrant the breakdown of the data by seasons first and then classify normal and abnormal days within each season. This methodology creates more homogeneity within each labeled dataset and allows the GAN to be trained better. For example, a normal winter day is sufficiently different from a normal summer day as both load and solar generation are significantly lower for the former in comparison to the latter.

2.1. Seasonal Classification

Rather than relying on a calendar-based approach to classify seasons, seasonal breakdowns are identified based on the available historical data. This enables us to capture spatial determinants of seasonal variations, such as geographic insolation and local weather patterns as well as geographic load patterns (e.g., heavy loads in winter for the colder climates and heavy loads in summer for the hotter climates). Moving daily averages of load and solar generation are plotted first, and the plot is then partitioned based on the pattern transition in both load and solar generation. Figure 1 depicts the resultant partitioning for the available historical data. The non-summer/non-winter days are classified as shoulder days. The characteristics of each season are captured in Table 1. Note that Shoulder A and Shoulder B are combined for training the GAN as they represent very similar (average) characteristics despite different slope polarities.

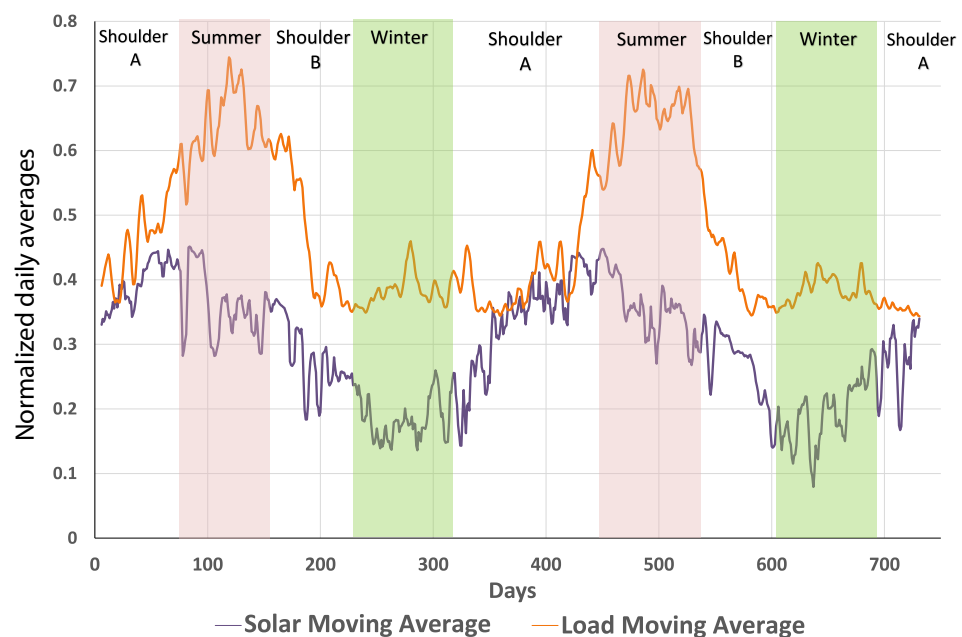


Figure 1. Seasonal classification of two years of historical data.

Table 1. Seasonal patterns.

Season	Load Pattern	Solar Pattern
Shoulder A	Low (increasing at the end)	Increasing to peak
Summer	High (peaking in the middle)	Decreasing
Shoulder B	Decreasing	Decreasing
Winter	Flat and low	Flat and low

2.2. Normal and Abnormal Days Classification

From a power system reliability planning perspective, understanding abnormal conditions is much more critical than normal conditions. Abnormal conditions require special attention, as any mismatch or sudden change in solar generation and load patterns could

impact the net load significantly. If the generated abnormal scenarios have a good correlation to the actual historical abnormal scenarios, they can enhance the analysis and understanding of such scenarios. Additionally, since the historical data does not have many abnormal scenarios, this type of scenario generation helps evaluate the system under such conditions by generating additional scenarios that are *similar but distinct*.

Normal and abnormal classification must be based on a metric that can provide clear and consistent differentiation between normal and abnormal days. In many cases, the choices of normal and abnormal days are based on ad hoc decisions. Instead, the methodology employed in this work uses the distance of any given day from a reference normal day to decide on the labeling of that day. While this methodology is universally applicable, it was determined that the application to the seasonally partitioned data was more appropriate.

To implement this methodology, a *reference normal day* must be identified for each season. The representative scenario generation methodology developed in [28] generates median representations for all seasons. The *reference normal day* is selected from the seasonal cluster by identifying the day with the shortest distance to the median seasonal representation. Since the daily profiles are time-series representations, dynamic time warping (DTW) is employed as it provides a better measure of (dis)similarity between days than the Euclidean distance measure [29,30]. DTW computes the best alignment between two time series by identifying the path with the minimum time-normalized distance between them. This is given by (1).

$$P^* = \arg \min_P \left[\frac{\sum_{s=1}^k d(p_s)}{k} \right] \quad (1)$$

where $d(p_s)$ is the distance between time-series points i_s and j_s , k is the length of the warping path, and P is the warping function. For univariate time-series data (e.g., hourly solar profile per day), the time series $i_s, j_s \in \mathbb{R}^{24 \times 1}$, and the DTW operation identifies the smallest distance by permuting through the different paths from hour 1 to hour 24. When considering multivariate time-series data, each day is represented by a matrix $d_i \in \mathbb{R}^{24 \times m}$ (where m is the number of variables), and the DTW operation is performed for each variable. DTW is particularly important for the multivariate case as it can accurately capture the *cross-correlations* between variables (such as load and solar generation), and it identifies similarities between patterns even if they are time-displaced.

For n days in a given seasonal dataset, DTW creates a symmetrical $\mathbb{R}^{n \times n}$ matrix that has the DTW distances between each pair of days. Next, the distance of each day in a season to its *reference normal day* is computed, and the sorted distances are plotted. The slope change can be detected by taking the second derivative of the DTW distance plots. The data points beyond this change point represent the abnormal days as they have farther and faster-growing distances from the reference normal day. This is illustrated in Figure 2. Applying the criterion outlined above, the summer season is split into 143 normal days and 39 abnormal days (from 2 years of data). A similar classification was obtained for the winter (146 normal and 36 abnormal days) and shoulder (327 normal and 40 abnormal days) seasons.

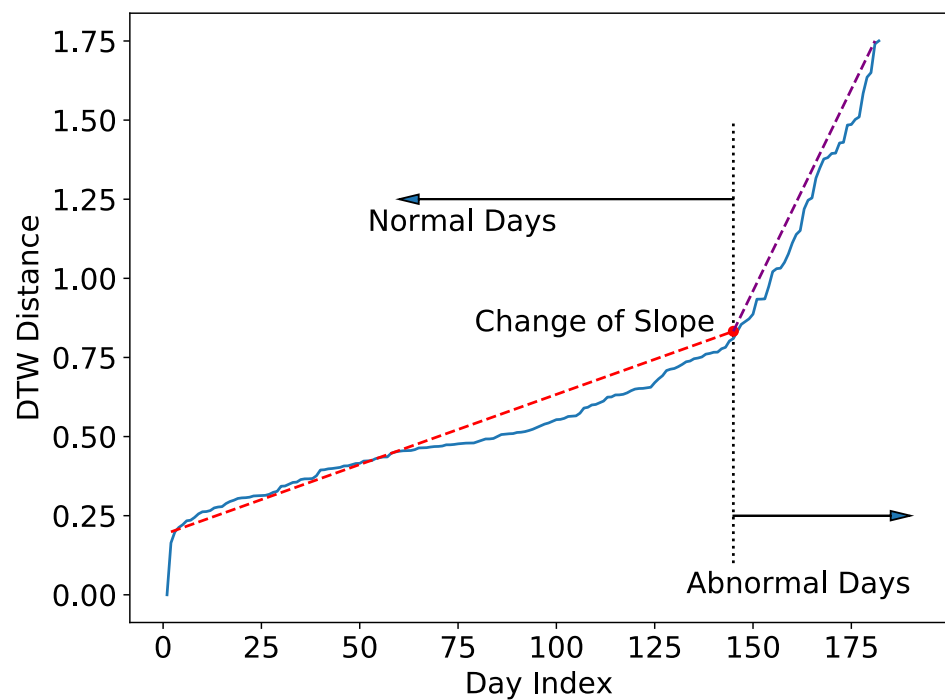


Figure 2. Summer season distances to reference normal day and normal/abnormal classification. Raw data (blue), fitted data for normal days (red), fitted data for abnormal days (purple).

3. Proposed Implicit Generative Model and Its Implementation

GANs are composed of two neural networks battling against one another (see Figure 3). The first neural network is called the generator, which aims to generate the synthesized samples. The second neural network is called the discriminator (or the critic). The discriminator's job is to differentiate between the real and the generated samples. The main objective of a GAN is to learn the distribution of a real dataset and map it to a separate latent space, from which more samples, similar to the original dataset, can be synthesized.

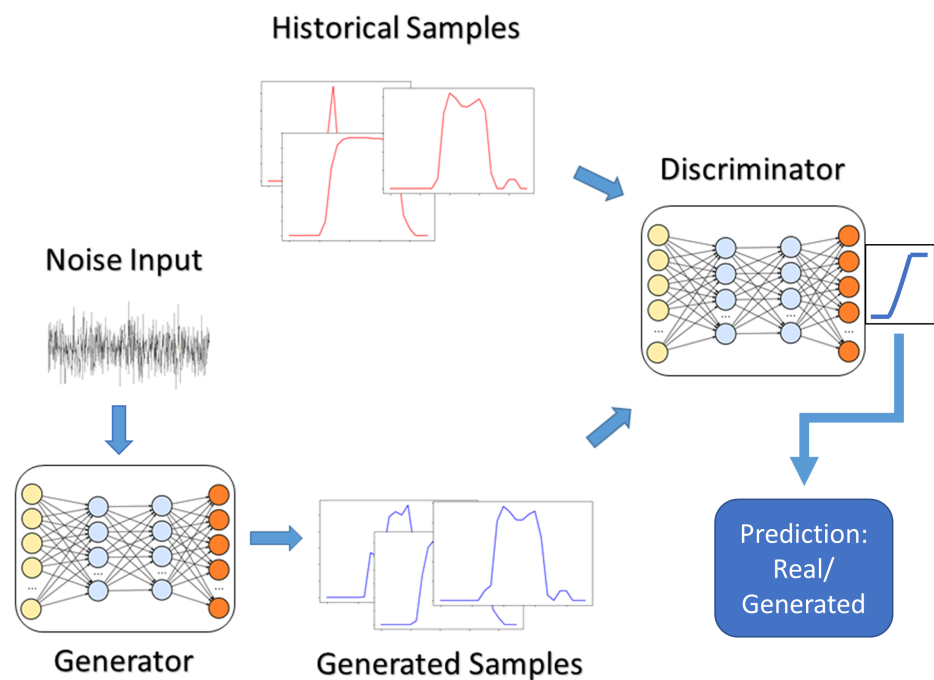


Figure 3. Architecture of generative adversarial networks (GANs).

Let us have a dataset, X , with samples x_i^t for time $t \in T$, and with dimensions i , whose distribution, $\mathbb{P}_x$, is to be learned by the generative model. Noise vector inputs, z , are sampled from a latent space, $\mathbb{P}_z$, and the multi-layer perceptrons within the generator, $G(z, \theta_g)$, are trained to map $\mathbb{P}_z$ to $\mathbb{P}_x$, without explicitly training on $\mathbb{P}_x$. This is accomplished by the generator producing samples as close to the real data's distribution as possible (denoted by $\mathbb{P}_x$). In contrast, the discriminator, $D(x, \theta_d)$, tries to distinguish the real samples from the generated ones and forces the generator to perform better. As the training progresses, the generator becomes better at producing realistic-looking samples, while the discriminator gets better at distinguishing generated samples from the real ones. The losses of the generator and the discriminator are expressed as,

$$L_G = -\mathbb{E}_Z[\log(D(G(z)))] \quad (2)$$

$$L_D = -\mathbb{E}_X[\log(D(x))] + \mathbb{E}_Z[\log(D(G(z)))] \quad (3)$$

The training of the generator and the discriminator can be summarised as a two-player mini-max game with the value function $V(G, D)$,

$$\min_G \max_D V(G, D) = \mathbb{E}_X[\log(D(x))] + \mathbb{E}_Z[\log(1 - D(G(z)))] \quad (4)$$

3.1. Proposed Conditional Recurrent GAN

GANs can be trained conditionally by incorporating labels in the training dataset, allowing the generator the ability to generate samples based on a certain event or condition. The label, y , can be any auxiliary information that can be appended to the real samples, x . The generator will then learn to associate a certain class of data with its associated label. After training has been finished, the generator can be forced to produce only a certain class of samples by appending the corresponding label, y , to all the noise vectors. The value function of the conditional GANs, conditioned on the label y , can be written as,

$$\min_G \max_D V(G, D) = \mathbb{E}_X[\log(D(x|y))] - \mathbb{E}_Z[\log(D(G(z|y)))] \quad (5)$$

Since the available historical data was a multivariate *time series*, it was necessary to include recurrent layers in both the generator and the discriminator. Recurrent layers in the generator model retain the time-series long-term modulations and help generate sequences that capture the fluctuations of the real data. In the discriminator, the recurrent layers help identify the sequential data better. The recurrent model of choice was RNN-LSTM, making the proposed machine learning model a conditional recurrent GAN. Note that the LSTM layer ensures that the recurrent GAN is properly trained to capture both short-term (daily) and long-term (seasonal) patterns in the time-series data. Furthermore, it leads to the generation of more homogeneous and valid training data for the GANs, which eventually leads to more consistent generated scenarios as the output of the GAN. The generator and the discriminator models consist of three stacked LSTM layers, along with a linear output layer. The hyperparameters of the models were tuned by comparing the observed outputs to the expected results. To optimize the discriminator output, it was trained thrice as much as the generator to maintain the best estimation ratio between the data density and the model density [31]. The model details are given in Table 2.

Table 2. Design details of cross-correlated conditional recurrent GAN (C²RGAN) model.

	Layer Type	Input Size	Output Size
Generator	LSTM	259	128
	LSTM	128	128
	LSTM	128	128
	Fully connected	128	24×2
Discriminator	LSTM	27×2	128
	LSTM	128	128
	LSTM	128	128
	Fully connected	128	1

3.2. Overall Implementation

The proposed approach of systematic model-free data segmentation and scenario generation using implicit generative models has been summarized in Figure 4. First, the historical data is preprocessed by normalizing the different variables to their peak values and creating daily profiles. Next, the data is segmented in preparation for GAN training. The available data is classified by season, and a representative day is selected for each season. Finally, the normal/abnormal classification is performed for each season, leading to the generation of six datasets (three seasons and normal/abnormal for each season). The next phase involves training the generator and discriminator using the labeled correlated datasets. The hyperparameters are tuned, and the loss functions are monitored to achieve an equilibrium that indicates a fully trained GAN model. In the next phase, the GAN model is fed labeled noise to generate similar but distinct scenarios for each of the six datasets. Finally, statistical validation of the generated scenarios is performed before moving onto OPF-based validation concerning the historical data and against the existing methodology of uncorrelated scenario generation.

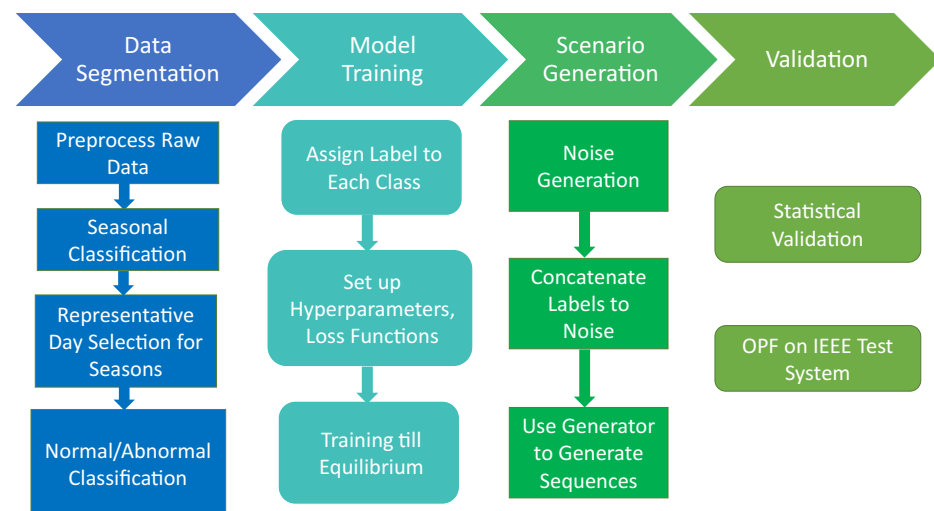


Figure 4. Flowchart of the proposed methodology. The first column captures the steps described in Section 2. The next two columns depict the training and use of the implicit generative model described in Section 3. The final column captures the thorough validation of the proposed methodology, which is described in Section 4.

4. Results and Analyses

The proposed approach was tested on a dataset provided by a power utility located in the US Southwest. The dataset comprised two years of hourly solar generation and load demand profiles at the transmission level. The nature of the data allowed for capture of

temporal and cross-correlations within the variables. However, as no spatial information was provided with the dataset, spatial correlations could not be captured. After preprocessing and normalizing the dataset, it was segregated into *summer*, *shoulder*, and *winter* seasons, followed by classification into *normal* and *abnormal* days. The C²RGAN was trained with these datasets. The C²RGAN-generated scenarios were then evaluated for their similarity to the historical data in the same category. Comparison of individual generated profiles for each variable to the historical profiles showed a good match, as shown in Figure 5 for summer normal real and generated load, and Figure 6 for summer normal real and generated solar, respectively. As is evident from the figures, the seasonal segmentation results in scenarios that closely track the temporal variations of the real dataset.

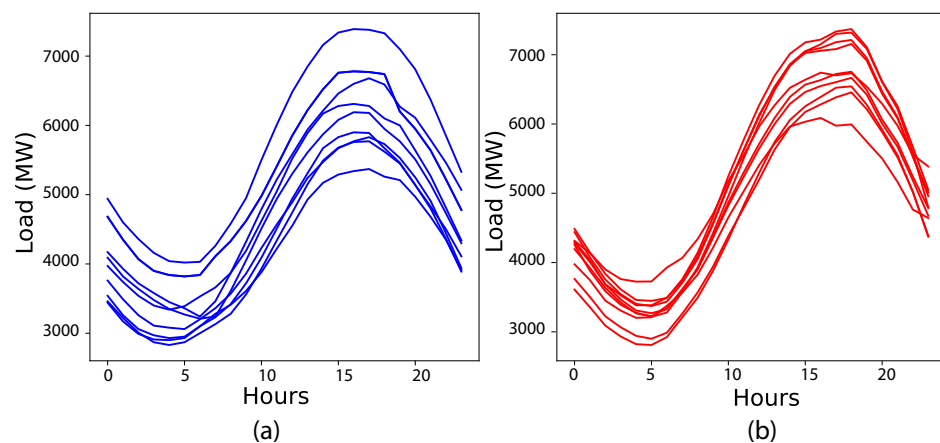


Figure 5. Selected summer season normal load profiles: (a) real; (b) generated.

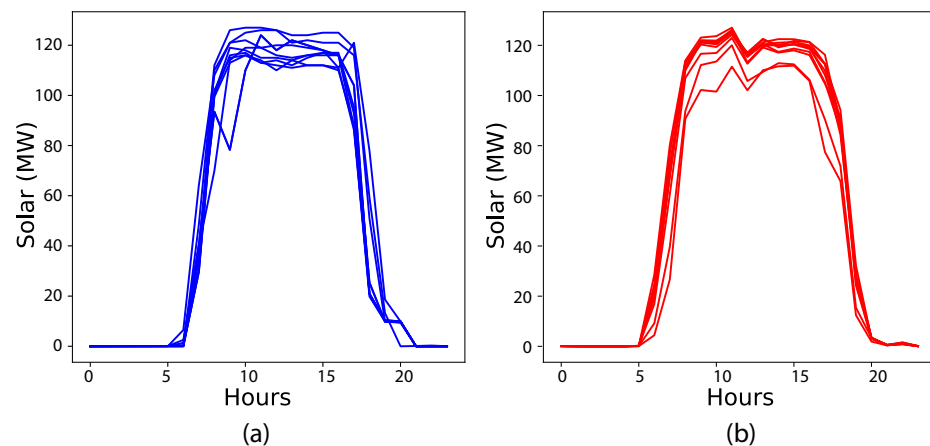


Figure 6. Selected summer season normal solar profiles: (a) real; (b) generated.

4.1. Statistical Validation of Proposed Implicit Generative Model

Going beyond visual confirmation, we performed rigorous statistical analysis to investigate the performance of the proposed scenario generation methodology. The statistical measure employed was the auto-correlation function (ACF), which defines how data points in a time series are related, on average, to the preceding data points.

Under normal conditions, the ACF shapes of the real and generated datasets for both load and solar were found to be very similar (see Figure 7a,b). The highest positive correlation at one hour for both variables confirms that the nearest temporal value has the highest correlation to any data point. However, since the normal solar peak and zero production times in summer are roughly 10 h each, the highest negative correlation occurs at a 10-h lag for the solar profile. The normal summer load pattern shown in Figure 5 is quasi-sinusoidal with peak and valley 12 h apart, which is consistent with the negative

ACF peak for the load at a 12-h lag. A similar pattern is observed for normal days in other seasons with slight variations in negative ACF peak location.

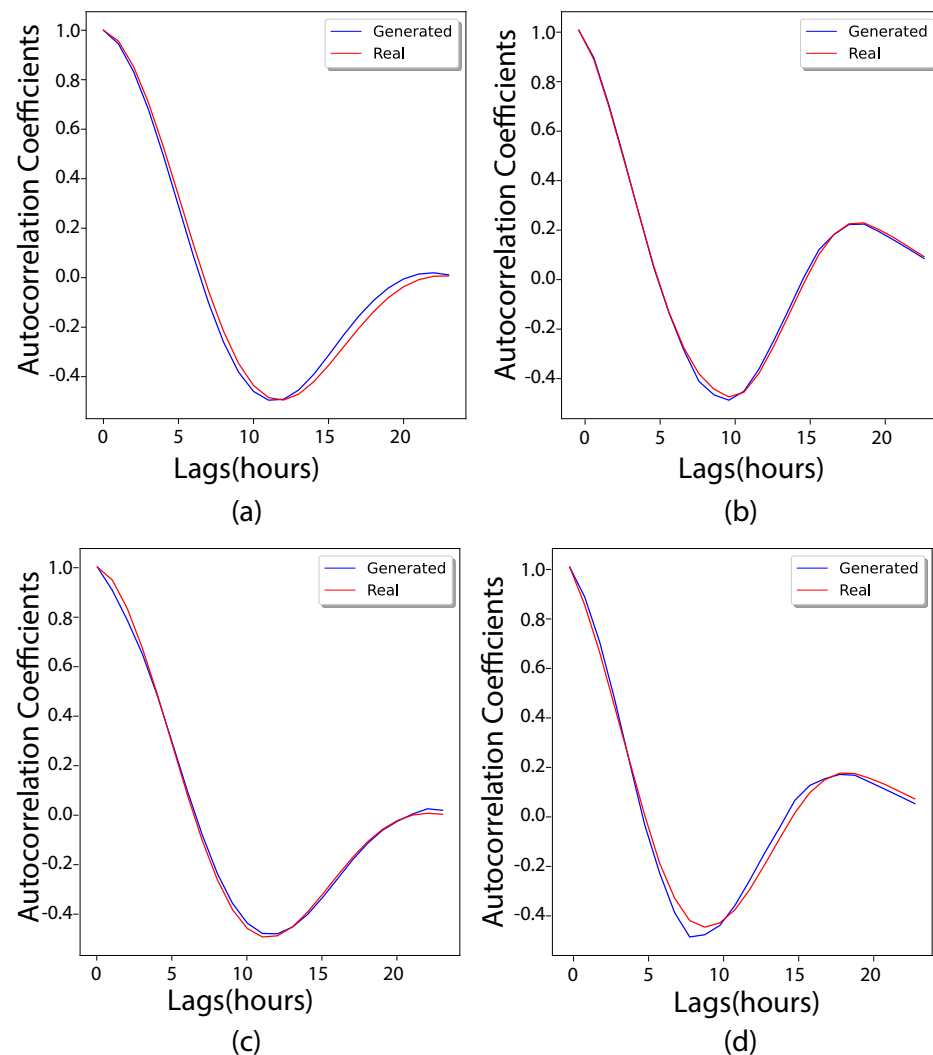


Figure 7. Summer season ACF for: (a) normal load; (b) normal solar; (c) abnormal load; (d) abnormal solar for generated and real data.

Under abnormal conditions, the load correlation shapes show a similar pattern as their counterparts under normal conditions, but a slight difference is observed between real and generated shapes for solar (see Figure 7c,d). This happens because the cross-correlated nature of the C²RGAN can bias one or both of its outputs (solar ACF in Figure 7d), as it is trained on both the variables. Therefore, its accuracy in producing matching scenarios for any one variable might be lower. However, we demonstrate in Section 4.2.2 that for actual power system applications, creating scenarios where the cross-correlations are considered results in more realistic outcomes.

4.2. Comparison with Uncorrelated Scenario Generation for Power System Application

To highlight the value of the correlated scenario generation process, two additional GANs were trained using the same historical data—one for the independent generation of load sequences and one for the independent generation of solar sequences. These univariate uncorrelated GANs (termed load GAN and solar GAN) generate seasonal (normal/abnormal) scenarios for load and solar generation, respectively. Note that many GAN-based scenario generation techniques proposed recently are univariate and hence,

uncorrelated (e.g., [22,23]). Therefore, the subsequent analysis is a comparison of the proposed methodology with the state of the art.

The selection of baseline days for uncorrelated scenarios is an important but challenging consideration. As load and solar are processed independently, there is no guaranteed or consistent overlap between the labeled training data for each set. This disjunction is more clearly pronounced for abnormal days. For example, an abnormal summer day for the load (very high load) can be vastly different from an abnormal summer day for solar generation (cloudy or rainy day). Thus, it is impossible to determine baseline days satisfying the same load and solar generation conditions.

One strategy could be to assume that the baseline days were identical for correlated and uncorrelated data. However, doing so will yield consistently favorable results for the correlated scenario generation approach since the baselines are drawn from its training dataset. Consequently, to avoid this possible (implicit) bias in favor of the proposed approach, the following strategy was devised in this paper: the baseline days for the uncorrelated scenarios were synthesized independently from the two training datasets (load and solar generation). Separate comparisons were then made between each approach's generated and baseline values.

4.2.1. Validation Using Optimal Power Flow (OPF) Analysis

To evaluate the performance of the generated scenarios for power systems applications, the generated solar and load profiles were applied to a modified IEEE 30-bus system [32]. A futuristic generation scenario was evaluated, where all the load buses also have solar generation. OPF was run under different ratios of solar generation peak to load peak, ranging from 0.3 to 1.2. Since the scenarios are derived from the historical dataset, the OPF converged for all the scenarios. To lend statistical validity to the exercise, 900 ($=30 \times 30$) scenarios were generated for both correlated and uncorrelated methodologies for each of the 6 classes (3 seasons $\times$ normal/abnormal). This enabled application of 30 distinct and randomly assigned profiles to all the buses of the system for one OPF computation. The OPF itself was run 30 times—each time with a completely different set of profiles—to ensure consistency of the results.

The distributions for each iteration were compared with the baseline to compute the distance between the two; the Wasserstein distance was used as a measure for this comparison. Additionally, the OPF results provided costs by the hour for each iteration. Finally, the voltage-angle data based on the hour/iteration/bus were collected for further analysis. The results were evaluated from multiple perspectives. Each methodology (correlated and uncorrelated scenario generation) was compared against its baseline to identify which would generate more realistic scenarios. Furthermore, comparisons were made over iterations to evaluate the consistency and on an hourly basis to identify if the gap between generated and baseline scenarios has any time-of-day dependence. The voltage-angle distribution plots were plotted for three different hours: 07, 12, and 17.

4.2.2. Results and Discussion for Abnormal Conditions

Figure 8 shows the shoulder season hourly OPF costs, averaged over 30 OPF iterations, for the solar-to-load ratio of 0.6. The correlated generated scenarios track their baseline much more closely than the uncorrelated scenarios. In addition, the baselines for the two cases show significant differences. The abnormal conditions typically signify lower solar generation (due to cloudy or rainy conditions), which is often accompanied by a lighter load (due to lower cooling requirements). However, the baseline of the uncorrelated case shows significantly higher OPF costs that result from the unrealistic combination of independently derived abnormal conditions (high load and no to low solar generation). The generated scenarios overestimate the costs (i.e., a combination of higher load and lower solar generation), resulting in grossly unrealistic scenarios.

A similar case is presented in Figure 9 for the summer abnormal situation, where the correlated scenario's costs are tracking the baseline costs well (similar to the shoulder

abnormal). The baselines for correlated and uncorrelated cases are more closely aligned compared to the shoulder abnormal (except for a few morning hours), but the uncorrelated generated scenarios are underestimating the cost by a large margin. Although not shown in the figures to ensure clarity, the behavior for the other solar-to-load ratio was consistent with these results.

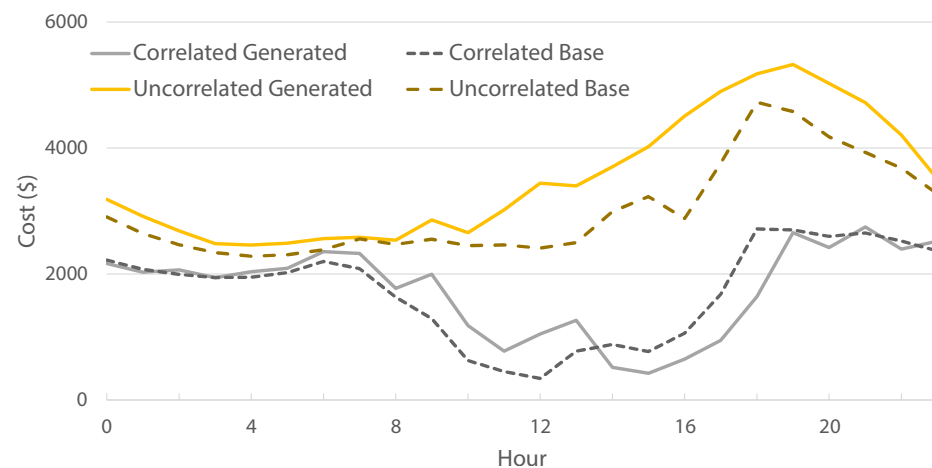


Figure 8. Shoulder abnormal hourly costs (ratio = 0.6).

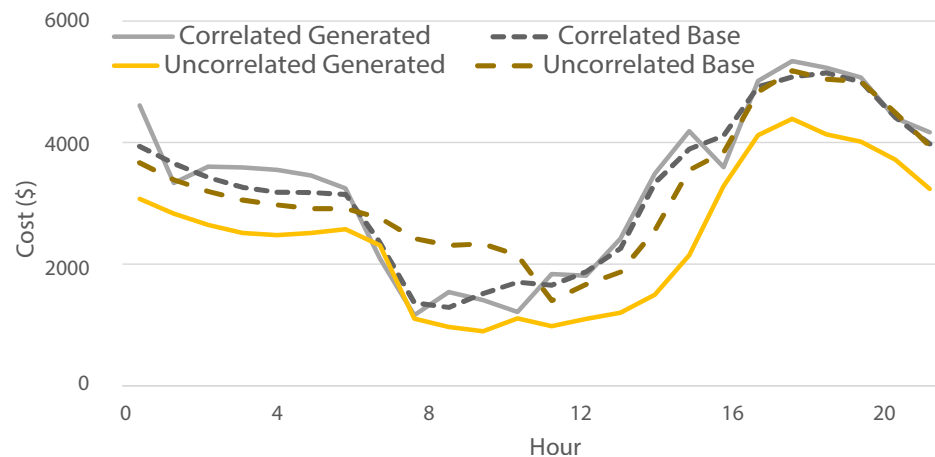


Figure 9. Summer abnormal hourly costs (ratio = 0.6).

In the case of the winter abnormal shown in Figure 10, the baseline costs between correlated and uncorrelated scenarios differ significantly—similar to the shoulder abnormal case. The correlated scenarios are much closer to their baseline than the uncorrelated ones. However, it is observed that the correlated scenarios are overestimating the costs between the hours of 8 AM and 6 PM, indicating that the generated solar scenarios are lower than the baseline. Under the winter abnormal conditions, the solar profiles are predominantly low with a few exceptions, so the GAN is getting trained to generate lower solar profiles. However, since the baseline does contain some higher solar generation profiles, there is some gap between the baseline and the correlated scenarios.

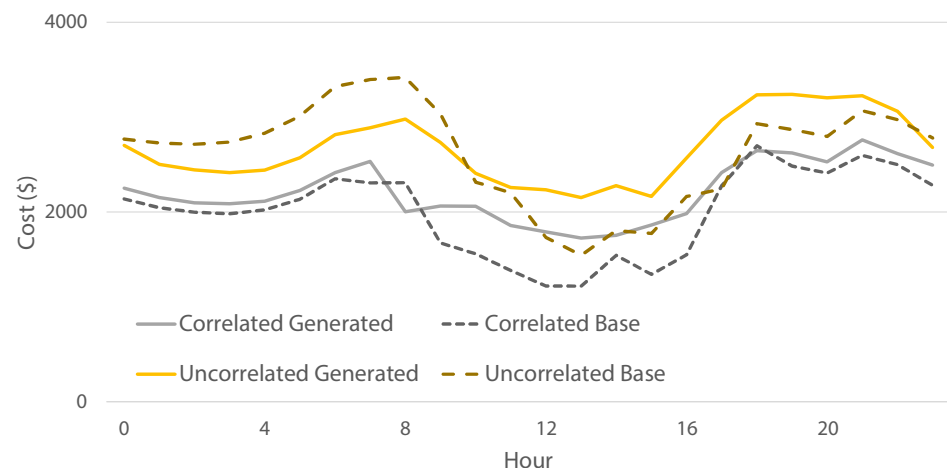


Figure 10. Winter abnormal hourly costs (ratio = 0.6).

In contrast to Figure 8, which provides averaged hourly OPF cost profiles over 30 iterations, Figure 11 depicts the average cost per hour for different OPF iterations for the solar-to-load ratio of 0.6. The correlated variations are narrower in range and closer to the baseline. This chart also underscores the baseline difference discussed above.

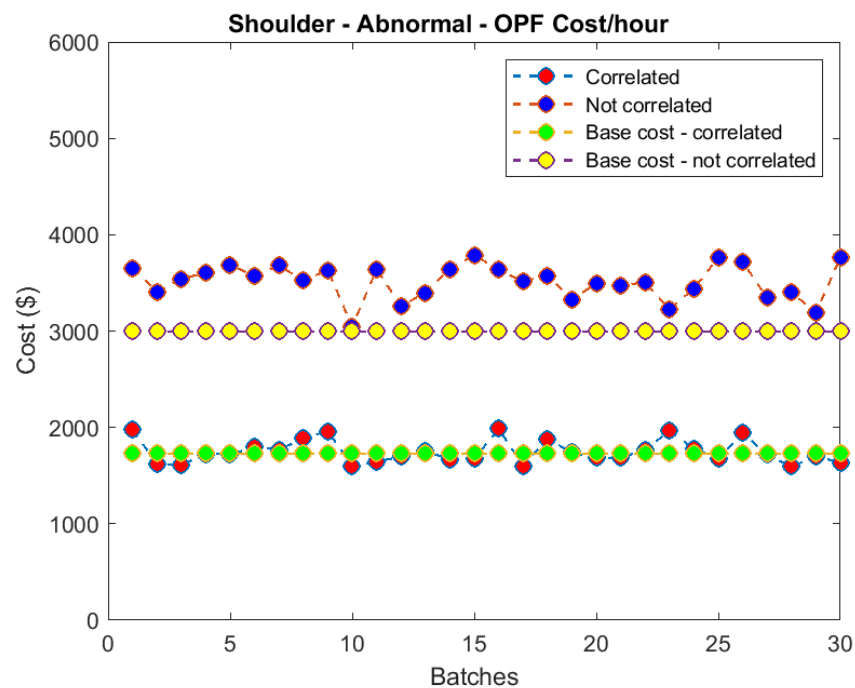


Figure 11. Shoulder abnormal OPF cost (ratio = 0.6).

Another perspective to view the differences between the correlated and uncorrelated approaches is to look at the voltage angle distributions for the 30-bus system. Figure 12 shows the probability density functions (PDFs) of the voltage angles for 5 PM for correlated and uncorrelated scenarios for all three seasons. The better overlap with the baseline distribution is clearly visible for the correlated scenarios. These plots are for one of the 30 iterations, but a similar pattern was observed for other hours and for all iterations, albeit with some variability. The larger difference between the correlated and uncorrelated scenarios in the shoulder season may be partially attributable to the data segmentation technique used in Section 2.2. However, the distinction between the two scenario generation methods is still evident in the other seasons.

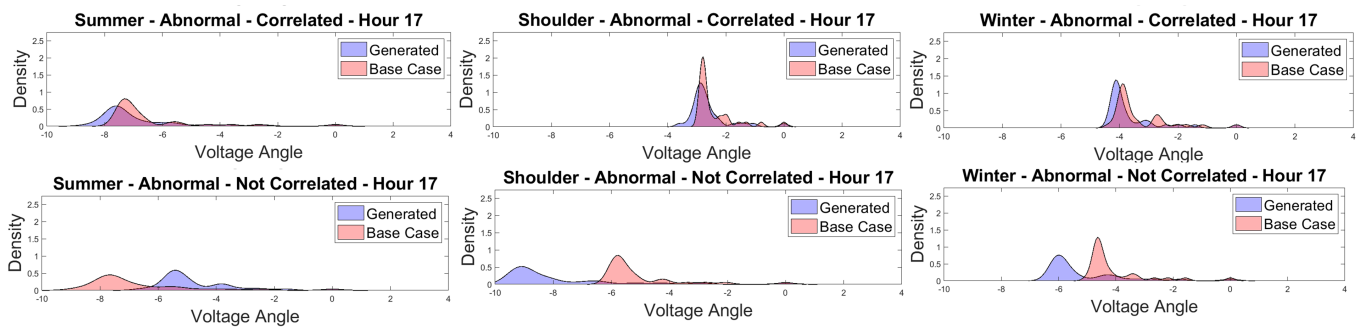


Figure 12. Voltage angle PDFs—abnormal, all seasons.

Table 3 shows the numerical results for abnormal seasonal daily OPFs. It covers solar-to-load ratios from 0.3 to 1.2 for all three seasons and correlated and uncorrelated conditions and reinforces the results and conclusions from earlier plots. The Wasserstein distances for correlated cases are lower than uncorrelated cases under most conditions, often by large margins. The Wasserstein distance should be a low number, but not 0, as we are aiming to obtain similar, but distinct scenarios. Correlated scenarios achieve this objective much better than uncorrelated scenarios, with a minor exception of high solar-to-load ratios in winter, for which the results are comparable. Moreover, the costs for the uncorrelated scenarios for the summer and shoulder seasons point to totally misleading results. For instance, even under abnormal conditions, summer costs should be highest due to high load, and shoulder costs should be lowest due to a combination of low load and good solar generation. However, the uncorrelated scenarios are showing the exact opposite behavior.

Table 3. Result summary for abnormal conditions.

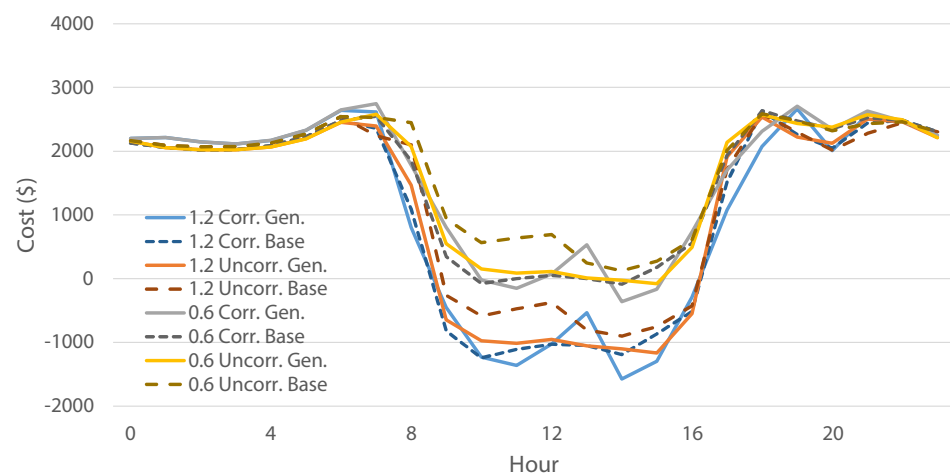
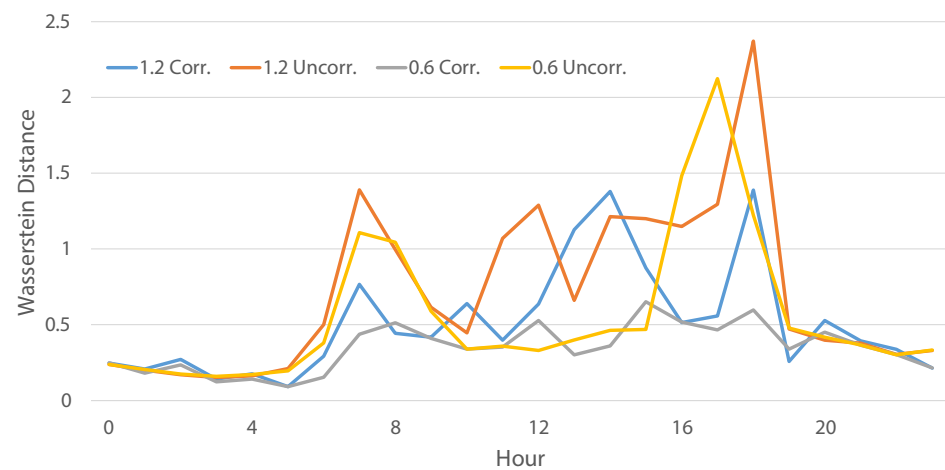
Solar-to-Load Ratio		Correlated				Uncorrelated			
		0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
Average	Summer	0.43	0.58	0.77	0.85	1.52	1.59	1.69	1.66
Wasserstein	Shoulder	0.40	0.57	0.71	0.77	1.00	1.06	1.12	1.17
Distance	Winter	0.31	0.47	0.67	0.84	0.58	0.59	0.62	0.65
Solar-to-Load Ratio		0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
Average	Summer	3951	3308	2762	2363	2992	2452	2007	1698
Daily	Shoulder	2087	1749	1466	1246	3776	3515	3268	3035
Cost (\$)	Winter	2342	2211	2084	1961	2796	2674	2556	2443

4.2.3. Results and Discussion for Normal Conditions

The difference between correlated and uncorrelated scenarios is not as significant under normal conditions. In fact, the uncorrelated scenarios showed a closer correlation to the baseline data in the summer season than the correlated scenarios, as shown in Table 4. This is understandable as the solar and load profiles for each season do not have many deviations under normal conditions, and the single variable nature of uncorrelated scenarios allows the corresponding GAN to be trained better for normal, independent signals. However, it was observed that for winter (see Figure 13), the uncorrelated scenarios are farther from their baselines (depicting lower costs) due to the overestimation of the solar generation. The Wasserstein distances for shoulder normal shown in Figure 14 indicate that the uncorrelated distances are higher than correlated ones for most hours of the day. The voltage-angle plots for three different hours for winter normal correlated scenarios (see Figure 15) demonstrate that the distributions are matching the baseline very well.

Table 4. Results summary for normal conditions.

Solar-to-Load Ratio		Correlated				Uncorrelated			
		0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
Average	Summer	0.61	0.69	0.77	0.75	0.30	0.33	0.39	0.45
Wasserstein	Shoulder	0.28	0.35	0.41	0.51	0.45	0.56	0.65	0.71
Distance	Winter	0.23	0.35	0.36	0.52	0.21	0.36	0.32	0.33
Solar-to-Load Ratio		0.3	0.6	0.9	1.2	0.3	0.6	0.9	1.2
Average	Summer	3499	2616	1960	1523	3417	2574	1950	1548
Daily	Shoulder	2038	1398	993	670	1832	1208	816	498
Cost (\$)	Winter	1978	1592	1323	1107	1940	1572	1333	1139

**Figure 13.** Winter normal hourly OPF costs (ratio = 0.6, 1.2).**Figure 14.** Shoulder normal hourly Wasserstein distances.

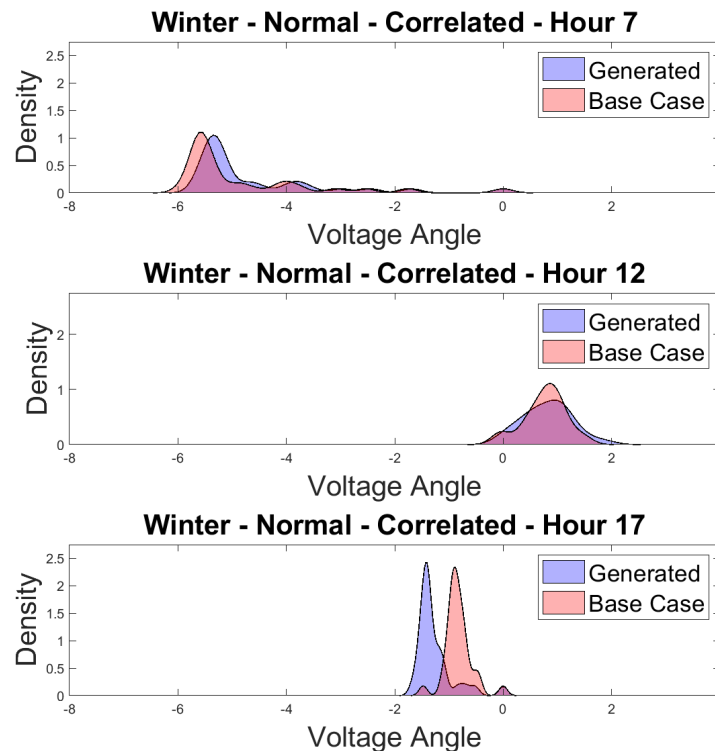


Figure 15. Voltage angle PDFs for winter normal–correlated.

Table 4 shows the normal seasonal summary results for solar-to-load ratios from 0.3 to 1.2 for all three seasons and correlated and uncorrelated conditions. It can be observed from the tables that the Wasserstein distances for normal conditions are similar (both are low) for correlated and uncorrelated scenarios. Similarly, the cost distinctions are minor under most conditions. However, the uncorrelated scenarios are consistently underestimating the costs for the shoulder season, which is in direct contrast to their behavior under abnormal conditions. As a result, the uncorrelated scenario-based OPF may demonstrate unreasonably high variations in OPF costs between normal and abnormal scenarios, leading to non-optimal outcomes from a long-term reliability planning perspective.

4.3. Practical Significance

Since many resource planning activities aim to distinguish abnormal conditions from normal conditions, it is helpful to compare how the generated abnormal scenarios differ from the generated normal scenarios. For correlated cases, the costs for the abnormal scenarios are consistently and reasonably higher than the costs of the normal scenarios due to the lower solar production on abnormal days. Winter days show the largest and most consistent gap through the day (see Figure 16), indicating the need for a longer traditional generation or battery backup requirements. For the shoulder (see Figure 17) and summer seasons, the gap between normal and abnormal is smaller and restricted to fewer hours of the day, indicating that the backup requirements may be less. For uncorrelated scenarios, consistency is absent: for the shoulder season, the abnormal scenarios grossly overestimate the net load (as shown in Figure 17); for summer, they show lower costs than the normal scenario, and no reasonable conclusions can be drawn from them. In summary, through the OPF application, we have demonstrated the ability of correlated scenario generation to create valid representative power system scenarios that are a prerequisite for long-term resource planning. In the future, we will apply the scenarios generated using the proposed approach to solve the optimal BESS sizing and siting problem [33].

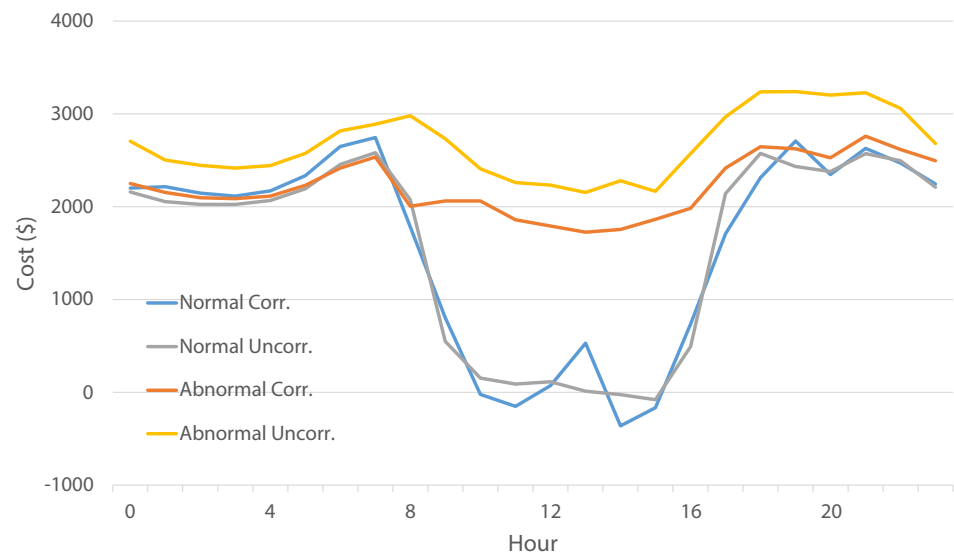


Figure 16. Winter normal–abnormal comparison (ratio = 0.6).

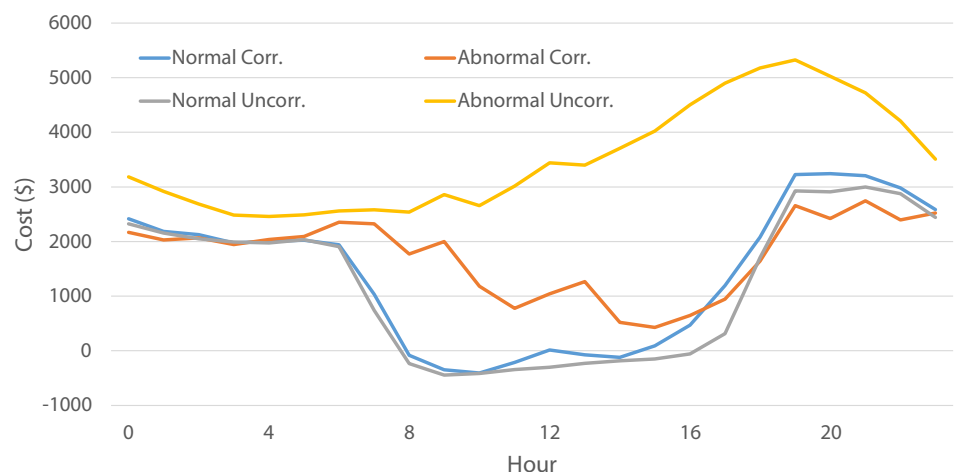


Figure 17. Shoulder normal–abnormal comparison (ratio = 0.6).

5. Conclusions

As the exploration of ways to understand and analyze the impacts of RG on grid reliability continues, synthetically generated representative scenarios will play an increasingly vital role. Due to legacy practices and/or ease of application, uncorrelated/univariate scenario generation is often used for such exploration. However, this may lead to outcomes that are not realistic. This paper demonstrates the utility of correlated multivariate scenario generation in understanding and analyzing normal and abnormal system conditions.

The proposed systematic end-to-end methodology for correlated scenario generation has the following components:

- Structured and model-free data segmentation.
- An informed selection/design of a cross-correlated conditional recurrent generative adversarial network (C^2 RGAN).
- Generation of correlated representative scenarios that augment the original dataset.
- Extensive and application-oriented validation that proves the value of the proposed methodology.

Overall, correlated scenario generation was seen to create more realistic profiles due to the integration of both solar generation and load demand in the training of the C^2 RGAN. From the OPF application evaluation, the following key conclusions are drawn:

- The correlated scenario generation resulted in lower and more accurate average hourly costs across the seasons (as shown in Figure 11 and Table 3).
- From the voltage angle distributions, it was observed that the correlated scenarios are more similar to the real case compared to uncorrelated scenarios (as shown in Figure 12).
- Seasonal performance analyses highlighted why inferences drawn from uncorrelated scenarios might be misleading (results from Tables 3 and 4).
- It was also shown that the results from uncorrelated scenarios are adequate for normal days, but it can lead to misplaced conviction about their applicability to abnormal scenarios (results from Tables 3 and 4).

The proposed methodology is voltage-level agnostic, scalable, and portable to different datasets, geographies, and end-application requirements. It can be used to analyze the reliability and resilience issues with various renewable energy penetration levels and come to definitive conclusions about deploying these resources. The proposed approach currently captures cross-correlations and temporal correlations that exist between RG and loads. With the right dataset and minor modifications, it can also be extended to capture spatial correlations between the different variables.

Author Contributions: D.D. and M.B. are co-first authors; conceptualization, A.P. and P.A.; methodology, A.P., D.D. and M.B.; software, M.B., D.D., H.S. and A.I.S.; validation, D.D., M.B. and A.P.; formal analysis, M.B. and D.D.; investigation, A.P., D.D., M.B. and A.I.S.; resources, A.P. and P.A.; data curation, A.I.S., H.S., D.D. and M.B.; writing—original draft preparation, D.D. and M.B.; writing—review and editing, A.P., A.I.S. and D.D.; visualization, M.B., D.D., H.S. and A.I.S.; supervision, A.P.; project administration, A.P. and P.A.; funding acquisition, A.P. and P.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Salt River Project (SRP) under grant 96-180C 2021-2022 EE-04, and the National Science Foundation (NSF) under grants OAC 1934766 and ECCS 2145063.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available as they were received under a non-disclosure agreement.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

The following abbreviations are used in this paper:

ACF	Auto-correlation Function
C ² RGAN	Cross-Correlated Conditional Recurrent Generative Adversarial Network
DTW	Dynamic Time Warping
GAN	Generative Adversarial Network
LSTM	Long Short-Term Memory
OPF	Optimal Power Flow
PDF	Probability Density Function
PV	Photovoltaic
RG	Renewable Generation
RNN	Recurrent Neural Network

References

1. Yan, R.; Wang, J.; Huo, S.; Qin, Y.; Zhang, J.; Tang, S.; Wang, Y.; Liu, Y.; Zhou, L. Flexibility improvement and stochastic multi-scenario hybrid optimization for an integrated energy system with high-proportion renewable energy. *Energy* **2023**, *263*, 125779. [\[CrossRef\]](#)
2. Hainsch, K.; Löffler, K.; Burandt, T.; Auer, H.; Crespo del Granado, P.; Pesciella, P.; Zwickl-Bernhard, S. Energy transition scenarios: What policies, societal attitudes, and technology developments will realize the EU Green Deal? *Energy* **2022**, *239*, 122067. [\[CrossRef\]](#)

3. Xie, L.; Huang, T.; Zheng, X.; Liu, Y.; Wang, M.; Vittal, V.; Kumar, P.; Shakkottai, S.; Cui, Y. Energy system digitization in the era of AI: A three-layered approach toward carbon neutrality. *Patterns* **2022**, *3*, 100640. [\[CrossRef\]](#)
4. Dumlao, S.M.G.; Ishihara, K.N. Weather-Driven Scenario Analysis for Decommissioning Coal Power Plants in High PV Penetration Grids. *Energies* **2021**, *14*, 2389. [\[CrossRef\]](#)
5. Padhee, M.; Pal, A.; Vance, K.A. Analyzing effects of seasonal variations in wind generation and load on voltage profiles. In Proceedings of the 2017 North American Power Symposium (NAPS), Morgantown, WV, USA, 17–19 September 2017; pp. 1–6.
6. Padhee, M.; Pal, A. Effect of solar PV penetration on residential energy consumption pattern. In Proceedings of the 2018 North American Power Symposium (NAPS), Fargo, ND, USA, 9–11 September 2018; pp. 1–6.
7. Mishra, C.; Pal, A.; Thorp, J.S.; Centeno, V.A. Transient stability assessment of prone-to-trip renewable generation rich power systems using Lyapunov's direct method. *IEEE Trans. Sustain. Energy* **2019**, *10*, 1523–1533. [\[CrossRef\]](#)
8. Buonanno, A.; Caliano, M.; Di Somma, M.; Graditi, G.; Valenti, M. A Comprehensive Tool for Scenario Generation of Solar Irradiance Profiles. *Energies* **2022**, *15*, 8830. [\[CrossRef\]](#)
9. Sund, L.; Talari, S.; Ketter, W. Stochastic Wind Power Generation Planning in Liberalised Electricity Markets within a Heterogeneous Landscape. *Energies* **2022**, *15*, 8109. [\[CrossRef\]](#)
10. Marulanda, G.; Bello, A.; Cifuentes, J.; Reneses, J. Wind Power Long-Term Scenario Generation Considering Spatial-Temporal Dependencies in Coupled Electricity Markets. *Energies* **2020**, *13*, 3427. [\[CrossRef\]](#)
11. Xie, Y.; Xu, Y. Transmission Expansion Planning Considering Wind Power and Load Uncertainties. *Energies* **2022**, *15*, 7140. [\[CrossRef\]](#)
12. Lee, D.; Baldick, R. Load and Wind Power Scenario Generation Through the Generalized Dynamic Factor Model. *IEEE Trans. Power Syst.* **2017**, *32*, 400–410. [\[CrossRef\]](#)
13. Hoeltgebaum, H.; Fernandes, C.; Street, A. Generating Joint Scenarios for Renewable Generation: The Case for Non-Gaussian Models With Time-Varying Parameters. *IEEE Trans. Power Syst.* **2018**, *33*, 7011–7019. [\[CrossRef\]](#)
14. Becker, R. Generation of time-coupled wind power infeed scenarios using pair-copula construction. *IEEE Trans. Sustain. Energy* **2017**, *9*, 1298–1306. [\[CrossRef\]](#)
15. Goh, H.H.; Peng, G.; Zhang, D.; Dai, W.; Kurniawan, T.A.; Goh, K.C.; Cham, C.L. A New Wind Speed Scenario Generation Method Based on Principal Component and R-Vine Copula Theories. *Energies* **2022**, *15*, 2698. [\[CrossRef\]](#)
16. Cui, M.; Zhang, J.; Wang, Q.; Krishnan, V.; Hodge, B.M. A data-driven methodology for probabilistic wind power ramp forecasting. *IEEE Trans. Smart Grid* **2017**, *10*, 1326–1338. [\[CrossRef\]](#)
17. Li, J.; Zhou, J.; Chen, B. Review of wind power scenario generation methods for optimal operation of renewable energy systems. *Applied Energy* **2020**, *280*, 115992. [\[CrossRef\]](#)
18. Cui, M.; Ke, D.; Sun, Y.; Gan, D.; Zhang, J.; Hodge, B.M. Wind power ramp event forecasting using a stochastic scenario generation method. *IEEE Trans. Sustain. Energy* **2015**, *6*, 422–433. [\[CrossRef\]](#)
19. Sideratos, G.; Hatziaargyriou, N.D. Probabilistic wind power forecasting using radial basis function neural networks. *IEEE Trans. Power Syst.* **2012**, *27*, 1788–1796. [\[CrossRef\]](#)
20. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the 28th Annual Conference on Neural Information Processing Systems 2014 (NIPS), Montreal, QC, Canada, 8–13 December 2014.
21. Liang, J.; Tang, W. Sequence generative adversarial networks for wind power scenario generation. *IEEE J. Sel. Areas Commun.* **2019**, *38*, 110–118. [\[CrossRef\]](#)
22. Jiang, C.; Mao, Y.; Chai, Y.; Yu, M.; Tao, S. Scenario generation for wind power using improved generative adversarial networks. *IEEE Access* **2018**, *6*, 62193–62203. [\[CrossRef\]](#)
23. Chen, Y.; Wang, Y.; Kirschen, D.; Zhang, B. Model-free renewable scenario generation using generative adversarial networks. *IEEE Trans. Power Syst.* **2018**, *33*, 3265–3275. [\[CrossRef\]](#)
24. Zhang, Y.; Ai, Q.; Xiao, F.; Hao, R.; Lu, T. Typical wind power scenario generation for multiple wind farms using conditional improved Wasserstein generative adversarial network. *Int. J. Electr. Power Energy Syst.* **2020**, *114*, 105388. [\[CrossRef\]](#)
25. Yuan, R.; Wang, B.; Sun, Y.; Song, X.; Watada, J. Conditional Style-based Generative Adversarial Networks for Renewable Scenario Generation. *IEEE Trans. Power Syst.* **2022**. [\[CrossRef\]](#)
26. Wei, H.; Hongxuan, Z.; Yu, D.; Yiting, W.; Ling, D.; Ming, X. Short-term optimal operation of hydro-wind-solar hybrid system with improved generative adversarial networks. *Appl. Energy* **2019**, *250*, 389–403. [\[CrossRef\]](#)
27. Choi, J.; Lee, J.I.; Lee, I.W.; Cha, S.W. Robust PV-BESS Scheduling for a Grid With Incentive for Forecast Accuracy. *IEEE Trans. Sustain. Energy* **2021**, *13*, 567–578. [\[CrossRef\]](#)
28. Dalal, D.; Pal, A.; Augustin, P. Representative Scenarios to Capture Renewable Generation Stochasticity and Cross-Correlations. In Proceedings of the 2022 IEEE Power & Energy Society General Meeting (PESGM), Denver, CO, USA, 17–21 July 2022; pp. 1–5. [\[CrossRef\]](#)
29. Keogh, E.; Ratanamahatana, C.A. Exact indexing of dynamic time warping. *Knowl. Inf. Syst.* **2005**, *7*, 358–386. [\[CrossRef\]](#)
30. Fabozzi, D.; Van Cutsem, T. Assessing the proximity of time evolutions through dynamic time warping. *IET Gener. Transm. Distrib.* **2011**, *5*, 1268–1276. [\[CrossRef\]](#)
31. Goodfellow, I.J. NIPS 2016 Tutorial: Generative Adversarial Networks. *arXiv* **2017**, arXiv:1701.00160.

32. IEEE 30 Bus System. Available online: <https://icseg.itl.illinois.edu/ieee-30-bus-system/> (accessed on 12 April 2022).
33. Padhee, M.; Pal, A.; Mishra, C.; Vance, K.A. A Fixed-Flexible BESS Allocation Scheme for Transmission Networks Considering Uncertainties. *IEEE Trans. Sustain. Energy* **2020**, *11*, 1883–1897. [[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.