WILEY | Hindawi

Research Article

Deep Echo State Network with Variable Memory Pattern for Solar Irradiance Prediction

Qian Li,[1] **Tao Li,**[1] **Jiangang Ouyang,**[1] **Dayong Yang** ,[1] **and Zhijun Guo** [2]

[1]*College of Advanced Manufacturing, Nanchang University, Nanchang 330031, China*
[2]*College of Information Engineering, Nanchang University, Nanchang 330031, China*

Correspondence should be addressed to Zhijun Guo; guozhijun@ncu.edu.cn

Accurate solar irradiance prediction plays an important role in ensuring the security and stability of renewable energy systems. Solar irradiance modeling is usually a time-dependent dynamic model. As a new kind of recurrent neural network, echo state network (ESN) shows excellent performance in the field of time series prediction. However, the memory length of classical ESN is fixed and finite, which makes it hard to map sufficient features of solar irradiance with long-range dependency. Therefore, a novel deep echo state network with variable memory pattern (VMP-DESN) is proposed in this brief. VMP-DESN consists of multiple connected reservoirs in series, and there exist different types of memory modes in VMP-DESN. To remember more input history information in the states, the time delay links are added in each reservoir and between every two reservoirs. The VMP-DESN is more flexible to deal with different input signals due to its variable memory modes in the reservoir states. Additionally, the effect of different memory patterns on the VMP-DESN performance is discussed in detail, including the antidisturbance ability, memory capacity, and prediction accuracy. Finally, the effectiveness of VMP-DESN is evaluated by predicting the real solar irradiance task.

1. Introduction

In recent years, accurate solar irradiance prediction has played an increasingly important role in the management, dispatch, and security of renewable energy systems [1]. To predict the solar irradiance series, various techniques have been proposed, such as statistical models [2, 3], artificial neural network (ANN) [4, 5], hybrid methods [6, 7], and so forth. Among those approaches, ANN has become the most popular method for solar irradiance prediction due to its strong nonlinear approximation ability [8].

Among ANN, one powerful algorithm is the recurrent neural network (RNN) [9]. Compared with feed-forward neural network (FNN), the RNN has dynamic characteristics and memory performance, by introducing connection loops in the hidden layer. Thus, it is more appropriate for RNN to model the time-dependent solar irradiance series than FNN. Echo state network, a simple yet improved variant of RNN [10], adopts a dynamic reservoir as the hidden layer. The dynamic reservoir contains a large number of randomly sparsely connected neurons, which can encode the input signal from the low-dimensional input space to the high-dimensional state space. In addition, only the output connection weights in ESN need to be trained by linear regression algorithms [11], while the input-reservoir and reservoir-reservoir weights are usually initialized randomly and remain unchanged in the training process. Therefore, ESN avoids the drawbacks of high computational complexity and gradient disappearance in traditional RNN. Due to those advantages, ESN has been widely applied to time series prediction [12–14], pattern recognition [15, 16], and control fields [17–19].

For those time series modeling with long-term dependency, the traditional ESN still shows limitations. On the one hand, the reservoir state update equation of ESN is fixed, which can only express the relation between partial input history and the current state. However, the input features far away from the current time cannot be preserved in the current states of ESN. If the states cannot contain enough input history, then the expected value will not be accurately

fitted in output layer. On the other hand, the traditional ESN only contains one reservoir layer, which makes it difficult to fully extract input features. Thus, different variants of ESN have been developed for the performance improvement, for instance, expanding single-reservoir to multiple reservoirs [20–23], changing the state update rules of the reservoir [24–26], and so on. The authors in [20, 21, 23] propose deep ESN (DESN) with multiple reservoirs in series array, which can process the input signal layer by layer. The experimental results show that the DESN has better prediction accuracy than traditional ESN due to the strong feature extraction ability of its deep topology. In [24, 25], some improved ESNs with variable state update equation are proposed, by adding the leaky integrator units in the reservoir. For these ESNs, more input and state history are preserved in the current state. Therefore, they have higher memory capacity than ESN, which also has been illustrated in some experiments. In addition, some other variants of ESN have been established and have been successfully applied to solar irradiance prediction, such as the chain-structure ESN (CESN) with multiple independent ESN modules [27, 28], multi-timescale ESN (MTS-ESN) with multiple reservoirs in parallel array [29], and so forth. These ESNs mainly focus on the topology changes for performance improvement, whose memory patterns are single and fixed.

It can be noticed that the aforementioned ESNs modify the topology or neuron model to improve the performance. Although some effective results have been achieved, they still face some challenges in practical complex application. Firstly, the memory capacity of DESN is finite and continuous in spite of the strong deep learning ability. The single memory pattern will affect the flexibility of DESN to process different kinds of input signals. Furthermore, the leaky ESNs mainly focus on single-reservoir topology although they have larger and variable memory capacity. The single-reservoir ESN has limited feature extraction capability, which will influence the modeling accuracy.

Therefore, a novel deep echo state network with variable memory pattern (VMP-DESN) is proposed in this paper in order to handle aforementioned challenges. To encode the input feature into a richer state space, VMP-DESN adopts multiple subreservoirs in series as the hidden layer. To remember more input history information, the time delay links are added in each reservoir and every two reservoirs. The VMP-DESN can be regarded as an extension of the DESN. Unlike the DESN, the memory capacity of VMP-DESN is variable due to the selective memory mode. Compared with single-reservoir leaky ESNs, the hidden layer of VMP-DESN consists of multiple subreservoirs, which can hierarchically process the input information. Therefore, it is more flexible and advantageous for VMP-DESN to deal with different input signals.

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

(i) A novel VMP-DESN model is proposed for solar irradiance prediction, which has multiple reservoirs topology and selective memory patterns. The VMP-

DESN is studied in terms of mathematical model, pattern classifications, and stability analysis, respectively.

(ii) The time delay links are added in each reservoir and between every two reservoirs in order to preserve more input history in the states. It is more flexible for VMP-DESN to process different input signals.

(iii) The effect of different memory patterns on VMP-DESN is quantitatively and qualitatively analyzed, including the prediction accuracy, antidisturbance ability, and memory capacity.

The remainder of this paper is organized as follows. Section 2 introduces the proposed VMP-DESN methodology in detail. In Section 3, the performance of VMP-DESN is comprehensively evaluated in terms of prediction accuracy, antidisturbance ability, and memory capacity. Finally, some conclusions are summarized in Section 4.

2. Methodology of a Novel VMP-DESN

The memory length of traditional ESN is usually finite and fixed, which cannot extract sufficient features of input signals with long-term dependency. Therefore, a novel deep echo state network with variable memory pattern (VMP-DESN) is proposed in this study to preserve more input history features. The developed VMP-DESN is more flexible to deal with different kinds of input signals, as a result of variable memory capacity. This section will describe the principle of the proposed VMP-DESN methodology.

2.1. VMP-DESN Model. As shown in Figure 1, VMP-DESN is composed of an input layer, a hidden layer with multiple subreservoirs in series and time delay links, and an output layer. On the one hand, the input signals can be encoded into a richer state space as a result of the hierarchical topology in VMP-DESN. On the other hand, it is more flexible to deal with various input signals as different types of memory patterns are included in VMP-DESN. To remember more input history features from the states, the time delay links are added in both each reservoir and between every two reservoirs. Therefore, VMP-DESN has selectively variable memory modes, which will be helpful for the reservoirs to map more characteristics of input signals with long-term dependency.

Assume VMP-DESN has M subreservoirs with the same neurons N for simplification, K and L are the number of input and output neurons, respectively. Denote the input and output signals at time step t as $\mathbf{u}(t) = [u_1(t), u_2(t), \ldots, u_K(t)]^T$ and $\mathbf{y}(t) = [y_1(t), y_2(t), \ldots, y_L(t)]^T$. The global reservoir states are given as $\mathbf{X}(t) = [\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_M(t)]^T$, where $\mathbf{x}_l(t) = (x_{l1}, x_{l2}, \ldots x_{lN})$ represents the states of reservoir layer l, $l = 1, 2, 3, \ldots M$. The input signal and reservoir states satisfy the compact sets through this paper. $\mathbf{W}^{in} \in \mathbb{R}_{N \times K}$ is the input connection weights matrix, $\mathbf{W}_l^{\text{res}} \in \mathbb{R}_{N \times N}$ denotes the internal connection weights matrix for reservoir l, and $\{\mathbf{W}\text{ext}/l\}_{l=2}^M \in \mathbb{R}_{N \times N}$
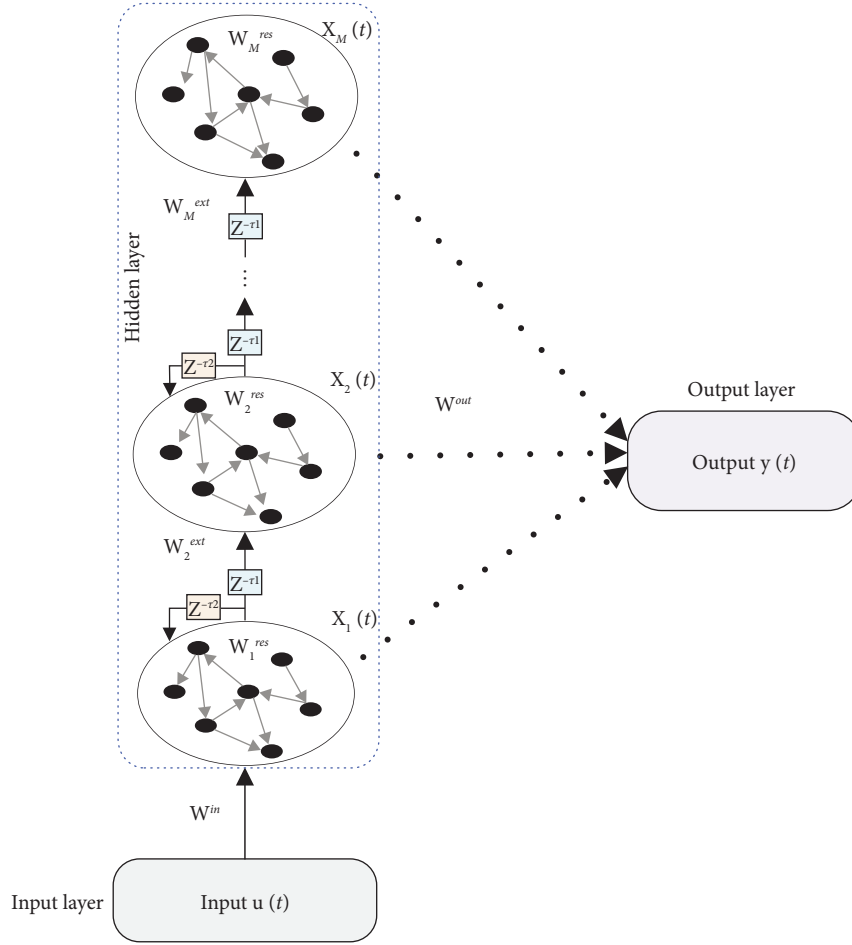
Figure 1: Structure of VMP-DESN.

is the external connection weights matrix between every two adjacent subreservoirs. τ_1 denotes the delayed time between every two adjacent reservoirs, while τ_2 is the delayed time in each reservoir.

The reservoir states of VMP-DESN are updated according to

$$\begin{aligned}
\mathbf{x}_1(t) &= \tanh\left(\mathbf{W}^{in}\mathbf{u}(t) + \mathbf{W}_1^{res}\mathbf{x}_1(t - \tau_2)\right), \\
\mathbf{x}_2(t) &= \tanh\left(\mathbf{W}_2^{ext}\mathbf{x}_1(t - \tau_1) + \mathbf{W}_2^{res}\mathbf{x}_2(t - \tau_2)\right), \\
\mathbf{x}_3(t) &= \tanh\left(\mathbf{W}_3^{ext}\mathbf{x}_2(t - \tau_1) + \mathbf{W}_3^{res}\mathbf{x}_3(t - \tau_2)\right), \\
&\ \ \vdots \\
\mathbf{x}_M(t) &= \tanh\left(\mathbf{W}_M^{ext}\mathbf{x}_{M-1}(t - \tau_1) + \mathbf{W}_M^{res}\mathbf{x}_M(t - \tau_2)\right).
\end{aligned} \tag{1}$$

It can be noticed that the input signal of layer l ($l > 1$) is the delayed state value of reservoir layer $l - 1$. The above equations can be generalized as follows:

$$\mathbf{x}_l(t) = \begin{cases} \tanh\left(\mathbf{W}^{in}\mathbf{u}(t) + \mathbf{W}_1^{res}\mathbf{x}_1(t - \tau_2)\right), & \text{if } l = 1, \\ \tanh\left(\mathbf{W}_l^{ext}\mathbf{x}_{l-1}(t - \tau_1) + \mathbf{W}_l^{res}\mathbf{x}_l(t - \tau_2)\right) & \text{if } 1 < l \leq M. \end{cases} \tag{2}$$

The network outputs of VMP-DESN are expressed as follows:

$$\mathbf{y}(t) = \mathbf{W}^{out}\mathbf{X}(t), \tag{3}$$

where $\mathbf{W}^{out} \in \mathbb{R}_{L \times MN}$ is the output connection weights matrix.

The method of training VMP-DESN is similar to classical ESN. That is to say, only the output connection matrix $\mathbf{W}^{out}$ needs to be learned in VMP-DESN, while other connection matrices remain fixed after proper initialization. The ridge regression training mechanism [30] is adopted to compute $\mathbf{W}^{out}$ by minimizing the cost function as follows:

$$\mathbf{W}^{out} = \underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{M}\mathbf{W} - \mathbf{D}\|_2^2 + \gamma\|\mathbf{W}\|_2^2. \tag{4}$$

The solution of equation (4) is given by

$$\left(\mathbf{W}^{out}\right)^T = \left(\mathbf{M}^T\mathbf{M} + \gamma\mathbf{E}\right)^{-1}\mathbf{M}^T\mathbf{D}, \tag{5}$$

where $\mathbf{M}$ is a L_{tr}-by-MN reservoir state matrix, $\mathbf{D}$ is a L_{tr}-by-L teacher signal vector matrix, and L_{tr} is the training length. $\mathbf{E}$ denotes a MN-by-MN identity matrix, and γ is the regulation parameter. During the training stage, $\mathbf{M}$ and $\mathbf{D}$ are collected as follows:

$$\mathbf{M} = \begin{bmatrix} \mathbf{x}_1(1) & \mathbf{x}_2(1) & \dots & \mathbf{x}_M(1) \\ \mathbf{x}_1(2) & \mathbf{x}_2(2) & \dots & \mathbf{x}_M(2) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}_1(L_{tr}) & \mathbf{x}_2(L_{tr}) & \dots & \mathbf{x}_M(L_{tr}) \end{bmatrix}_{L_{tr} \times MN}, \quad (6)$$

$$\mathbf{D} = \begin{bmatrix} d_1(1) & d_2(1) & \dots & d_L(1) \\ d_1(2) & d_2(2) & \dots & d_L(2) \\ \vdots & \vdots & \vdots & \vdots \\ d_1(L_{tr}) & d_2(L_{tr}) & \dots & d_L(L_{tr}) \end{bmatrix}_{L_{tr} \times L}, \quad (7)$$

where the t-th column of $\mathbf{M}$ includes the state signal of all reservoir layers at time t and the t-th column of $\mathbf{D}$ denotes the corresponding expected output signal.

The training procedure of VMP-DESN is given in Algorithm 1. According to Algorithm 1, the model parameters are firstly initialized by trial and error. Then, the states are updated according to equation (1), when the system is driven by input signal. At the same time, the reservoir states matrix and corresponding expected signal are collected according to equations (6)-(7), respectively. The output connection weights are finally computed as equation (5).

Remark 1. When the delayed time are set to $\tau_1 = 0$ and $\tau_2 = 1$, the DESN can be obtained. On this basis, if the number of reservoir is further set to 1 (i.e., $M = 1$), the DESN model will degrade to the standard ESN. In addition, the variable memory pattern ESN (VMP-ESN) can be obtained when $\tau_1 = 0$, $\tau_2 > 1$, and $M = 1$. DESN, ESN, and VMP-ESN are special cases of VMP-DESN, which will be taken as benchmarks to verify the effectiveness of VMP-DESN in later comparative simulation.

Remark 2. The proposed VMP-DESN includes multiple reservoirs in series array and different types of time delay links. On the one hand, the input signals can be encoded into a richer state space representation due to the hierarchical structure of VMP-DESN. On the other hand, the reservoir state update equation of VMP-DESN is variable and designable, which can deal with different input signals more flexibly.

2.2. Pattern Classification of VMP-DESN. From the dynamics of VMP-DESN in equation (2), it can be noticed that the VMP-DESN mainly includes two types of delayed links, i.e., the delay of internal state in each reservoir and the delay of state information transmission between every two subreservoirs. In order to explore the impact of different memory patterns on the network performance, three various VMPs-DESN are discussed in this current study, as shown in Table 1.

2.2.1. Pattern I: VMP1-DESN. To test the delay of internal state in each reservoir on network performance, the VMP1-DESN is developed with $\tau_1 = 0$ and $\tau_2 \geq 1$. The update rules of VMP1-DESN are rewritten as follows:

$$\mathbf{x}_l(t) = \begin{cases} \tanh\left(\mathbf{W}^{in}\mathbf{u}(t) + \mathbf{W}_1^{res}\mathbf{x}_1(t - \tau_2)\right), & \text{if } l = 1, \\ \tanh\left(\mathbf{W}_l^{ext}\mathbf{x}_{l-1}(t) + \mathbf{W}_l^{res}\mathbf{x}_l(t - \tau_2)\right), & \text{if } 1 < l \leq M. \end{cases} \quad (8)$$

It can be seen that only one kind of memory mode is retained in VMP1-DESN. VMP1-DESN is a special case of VMP-DESN (equation (2)) which can also be regarded as an extension of DESN. The ridge regression training mechanism in equation (4) can also be used to regulate $\mathbf{W}^{out}$ of VMP1-DESN.

2.2.2. Pattern II: VMP2-DESN. In order to verify the delay of state information transmission between every two sub-reservoirs on network performance, the VMP2-DESN is established with $\tau_1 \neq 0$ and $\tau_2 = 1$. The update equations of VMP2-DESN are expressed as follows:

$$\mathbf{x}_l(t) = \begin{cases} \tanh\left(\mathbf{W}^{in}\mathbf{u}(t) + \mathbf{W}_1^{res}\mathbf{x}_1(t - 1)\right), & \text{if } l = 1, \\ \tanh\left(\mathbf{W}_l^{ext}\mathbf{x}_{l-1}(t - \tau_1) + \mathbf{W}_l^{res}\mathbf{x}_l(t - 1)\right), & \text{if } 1 < l \leq M. \end{cases} \quad (9)$$

Obviously, the VMP2-DESN is also an extension of DESN. Compared with traditional ESN and DESN, the VMP2-DESN can obtain richer asynchronous state signals, which can express more input history features.

2.2.3. Pattern III: VMP3-DESN. VMP3-DESN includes two kinds of delayed links, i.e., the delay of internal state in each reservoir and the delay of state information transmission between every two reservoirs. The dynamics of VMP3-DESN are formulated as follows:

$$\mathbf{x}_l(t) = \begin{cases} \tanh\left(\mathbf{W}^{in}\mathbf{u}(t) + \mathbf{W}_1^{res}\mathbf{x}_1(t - \tau_2)\right), & \text{if } l = 1, \\ \tanh\left(\mathbf{W}_l^{ext}\mathbf{x}_{l-1}(t) - \tau_1) + \mathbf{W}_l^{res}\mathbf{x}_l(t - \tau_2)\right), & \text{if } 1 < l \leq M. \end{cases} \quad (10)$$

In VMP3-DESN, it should be satisfied that $\tau_1 \neq 0$ and $\tau_2 > 1$. It can be observed that the VMP3-DESN model is an extension of DESN, VMP1-DESN, and VMP2-DESN.

Remark 3. In order to study the influence of different memory patterns on network performance, VMP1-DESN, VMP2-DESN, and VMP3-DESN are developed, respectively, which correspond to different reservoir state update equation. VMP1-DESN, VMP2-DESN, and VMP3-DESN are three special forms of VMP-DESN (equation (2)) under different constraints. Therefore, the method to train the above three VMPs-DESN can refer to Algorithm 1. Furthermore, the effect of three various VMPs-DESN on network performance will be discussed and compared in later simulation.

2.3. Echo State Property Analysis of VMP-DESN. A valid ESN should satisfy the echo state property (ESP), which plays an important role in ensuring the asymptotical stability [31]. ESP denotes that the reservoir states only depend on the input signal with time and independent of its initial values. Similarly, the ESP of the VMP-DESN system is mainly discussed in this section, to guarantee the asymptotically stable operation. To facilitate the theoretical analysis of ESP,

$$\max_{l=1,2,\ldots,M} \rho\left(\mathbf{W}_l^{\text{res}}\right) < 1. \tag{16}$$

```
(1)  Input: u;
(2)  Output: W^out;
(3)  Initialization: {W^in, W_l^ext, W_l^res, K, N, L, M, γ, τ_1, τ_2};
(4)  for t←1 to L_tr
(5)    for l←1 to M
(6)      if l = 1 then
(7)        x_l(t)←tanh(W^in u(t) + W_1^res x_1(t − τ_2));
(8)      else
(9)        x_l(t)←tanh(W_l^ext x_{l−1}(t − τ_1) + W_l^yes x_l(t − τ_2));
(10)     end
(11)   end
(12)   y(t)←y_desired(t); X(t)←(x_1(t), x_2(t), ..., x_M(t));
(13) end
(14) M←(X(1), X(2), ..., X(L_tr));
(15) D←(y(1), y(2), ..., y(L_tr));
(16) (W^out)^T ← (M^T M + γE)^{−1} M^T D;
```

ALGORITHM 1: Training procedures of VMP-DESN.

the dynamics of VMP-DESN in equation (2) are rewritten as follows:

$$\mathbf{x}_l(t) = \tanh\left(\mathbf{W}_l^{in}\mathbf{I}_l(t) + \mathbf{W}_l^{\text{res}}\mathbf{x}_l\left(t - \tau_2\right)\right), \tag{11}$$

where the input signal $\mathbf{I}_l(t)$ is defined as

$$\mathbf{I}_l(t) = \begin{cases} \mathbf{u}(t), & \text{if } l = 1, \\ \mathbf{x}_{l-1}\left(t - \tau_1\right), & \text{if } 1 < l \le M, \end{cases} \tag{12}$$

and the input connection matrices $\mathbf{W}_l^{in}$ is denoted as

$$\mathbf{W}_l^{in} = \begin{cases} \mathbf{W}^{in}, & \text{if } l = 1, \\ \mathbf{W}_l^{ext}, & \text{if } 1 < l \le M. \end{cases} \tag{13}$$

Then, the Euclidean distance between $\mathbf{x}_l(t)$ and $\mathbf{x}_l'(t)$ randomly initialized from $\mathbf{x}_l(0)$ and $\mathbf{x}_l'(0)$ is computed by

$$\begin{aligned} \left\|\mathbf{x}_l(t) - \mathbf{x}_l'(t)\right\| &= \left\|\tanh\left(\mathbf{W}_l^{in}\mathbf{I}_l(t) + \mathbf{W}_l^{\text{res}}\mathbf{x}_l\left(t - \tau_2\right)\right)\right. \\ &\quad \left. - \tanh\left(\mathbf{W}_l^{in}\mathbf{I}_l(t) + \mathbf{W}_l^{\text{res}}\mathbf{x}_l'\left(t - \tau_2\right)\right)\right\|. \end{aligned} \tag{14}$$

According to Lagrange's mean value theorem and the fact $\tanh' \in (0,1)$, one can obtain that

$$\begin{aligned} \left\|\mathbf{x}_l(t) - \mathbf{x}_l'(t)\right\| &= \left\|\tanh\left(\mathbf{W}_l^{\text{res}}\mathbf{x}_l\left(t - \tau_2\right) - \mathbf{W}_l^{\text{res}}\mathbf{x}_l'\left(t - \tau_2\right)\right)\right\| \\ &\le \left\|\mathbf{W}_l^{\text{res}}\right\|\left\|\mathbf{x}_l\left(t - \tau_2\right) - \mathbf{x}_l'\left(t - \tau_2\right)\right\| \\ &\le \ldots\ldots \le \left\|\mathbf{W}_l^{\text{res}}\right\|^{t - \tau_2}\left\|\mathbf{x}_l(0) - \mathbf{x}_l'(0)\right\|, \end{aligned} \tag{15}$$

$\mathbf{x}_l(0)$ and $\mathbf{x}_l'(0)$ are randomly initialized state values, i.e., $\left\|\mathbf{x}_l(0) - \mathbf{x}_l'(0)\right\|$ is bounded. According to equation (15), it can be obviously obtained that $\left\|\mathbf{x}_l(t) - \mathbf{x}_l'(t)\right\| \longrightarrow 0$, when $\left\|\mathbf{W}_l^{\text{res}}\right\| < 1$ and $t \longrightarrow \infty$. Therefore, $\rho\left(\mathbf{W}_l^{\text{res}}\right) < 1$ can ensure the ESP of reservoir l, as a result of $\left\|\mathbf{W}_l^{\text{res}}\right\| \ge \rho\left(\mathbf{W}_l^{\text{res}}\right)$ ($\rho(\cdot)$ denotes the spectral radius of $\mathbf{W}_l^{\text{res}}$). Furthermore, the necessary condition to guarantee the ESP of VMP-DESN model is deduced as follows:

3. Experimental Design and Results

In this section, the effectiveness of the proposed VMP-DESN is substantiated by solar irradiance forecasting. The solar irradiance datasets in the whole year 2017 with hourly time interval are used for simulation, which are obtained from California Irrigation Management Information System (CIMIS) [32]. CIMIS provides public access to data on the solar irradiance and other details about the stations. For each dataset, the training period and testing period are shown in Table 2.

3.1. Performance Evaluation Metrics. In this current study, the performance of the proposed methodology is comprehensively evaluated through quantitative and qualitative analysis. Four different external metrics in [1] are used to quantitatively evaluate the accuracy of VMPs-DESN, i.e., root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and normalized root mean square error (nRMSE), as shown in equations (17)–(20).

$$\text{RMSE} = \sqrt{\frac{1}{l_{te}}\sum_{t=1}^{l_{te}}\left(y_{\text{forecasted},t} - y_{\text{actual},t}\right)^2}, \tag{17}$$

$$\text{MAE} = \frac{1}{l_{te}}\sum_{t=1}^{l_{te}}\left|y_{\text{forecasted},t} - y_{\text{actual},t}\right|, \tag{18}$$

$$\text{MAPE} = \frac{1}{l_{te}}\sum_{t=1}^{l_{te}}\left|\frac{y_{\text{forecasted},t} - y_{\text{actual},t}}{y_{\text{actual},t}}\right|, \tag{19}$$

$$n\text{RMSE} = \frac{1}{\overline{y}}\sqrt{\frac{1}{l_{te}}\sum_{t=1}^{l_{te}}\left(y_{\text{forecasted},t} - y_{\text{actual},t}\right)^2}, \tag{20}$$

where $y_{\text{forecasted},t}$ is the predicted value, $y_{\text{actual},t}$ is the target value, $\overline{y}$ is the mean of actual value, and l_{te} denotes the length of testing step. In addition, the antidisturbance ability and memory capacity of VMP-DESN will be further qualitatively analyzed in later simulation. The influence of disturbance on one network with strong antidisturbance ability will disappear quickly, which plays an important role in practical applications. Memory capacity refers to the information length that one network can remember in a short time.

3.2. Modeling Accuracy of VMP-DESN. The prediction performance of VMP-DESN is verified by one-hour-ahead solar irradiance prediction, and only the historical solar irradiance data are adopted to model in this study. In order to provide appropriate input and output signals for VMP-DESN, the time series analysis [23] is used to characterize the historical solar irradiance data. The network parameters are

TABLE 1: Different pattern cases of VMP-DESN.

Pattern	Model	Parameter setting
I	VMP1-DESN	$\tau_1 = 0, \tau_2 \geq 1$
II	VMP2-DESN	$\tau_1 \neq 0, \tau_2 = 1$
III	VMP3-DESN	$\tau_1 \neq 0, \tau_2 > 1$

determined by trial and error, and they are set as follows: the total reservoir size is set as 200, the number of reservoir layer is 4, the input number is 2, and the output layer is 1; connection weights in matrices $\mathbf{W}^{in}$ and $\{\mathbf{W}_l^{ext}\}_{l=2}^M$ are sampled from a uniform distribution over $[-0.1, 0.1]$, while the spectral radius of matrices $\{\mathbf{W}_l^{res}\}_{l=1}^M$ is set as 0.85 to meet the ESP condition. The regularization parameter γ is 10^{-3}.

The influence of different delayed time on the four-layer VMP-DESN is firstly studied. Evaluation period II in Seeley station is adopted to test the prediction accuracy versus delayed time. The iteration process is stopped when the testing error is greater than the initial error. The statistical results are shown in Table 3, in terms of RMSE, nRMSE, MAE, and MAPE, respectively. In addition, the effect of different memory patterns on VMP-DESN performance is also compared in Table 3. Note that VMP-ESN with $\tau_1 = 0, \tau_2 = 1$, and $M = 1$ corresponds to standard ESN and VMP1-ESN with $\tau_1 = 0, \tau_1 = 1$, and $M > 1$ corresponds to traditional DESN. From the results in Table 3, it can be obtained that (1) the prediction accuracy of DESN is higher than that of ESN with the same reservoir units; (2) the proposed VMP-DESN and VMP-ESN perform better than DESN and ESN; (3) VMP1-DESN, VMP2-DESN, and VMP3-DESN have equivalent prediction accuracy, indicating that the performance of DESN can be effectively improved by adding these three different memory patterns. Additionally, the prediction error of VMP-ESN, VMP1-DESN, and VMP3-DESN varies rapidly with the delayed time, and it starts to increase when the delayed time is greater than 3. However, the prediction error of VMP2-ESN varies slowly with the delayed time, and it starts to increase when $\tau_1 > 7$. The reason behind this phenomenon may be that the delayed link in reservoir can speed up the network response, while the delayed link between every two subreservoirs can stabilize the network performance. The performance of classical ESN, DESN, VMP-ESN, and VMPs-DESN with the best prediction accuracy is compared in Figure 2(a), and the corresponding relative errors are shown in Figure 2(b). Obviously, experimental results clearly show that the proposed VMP-DESN could match the actual value better than ESN, DESN, and VMP-ESN, especially for some inflection points.

Taking Seeley and Blythe NE stations as two examples, the performance of VMPs-DESN is further verified. The statistical results of different prediction models are recalculated and reported in Table 4, in terms of RMSE, nRMSE, MAE, and MAPE. According to the results obtained, it can be seen that VMPs-DESN can achieve higher accuracy than other models for the most testing evaluation periods. Comparative results between ESN and DESN indicate the advantage of hierarchical topology, while the comparative results between ESN and VMP-ESN verify the effectiveness of variable memory. In addition, the equivalent prediction accuracy of VMP1-DESN, VMP2-DESN, and VMP3-DESN further demonstrates the validity of three different types of memory patterns in improving the accuracy of DESN.

3.3. Antidisturbance Ability of VMP-DESN. The antidisturbance ability of VMP-DESN will be qualitatively evaluated in this subsection. The effect of two kinds of disturbances on the networks is mainly analyzed, i.e., the disturbances caused by differences in state initialization and noise in the input signal. For a certain model, the first kind of disturbance can be constructed by initializing two different reservoir states, while other model parameters remain the same. Similarly, the second type of disturbance can be built by adding small disturbance to the input signal, while the model parameters remain consistent. The effect caused by disturbance is expressed by the difference between reservoir states. In this current study, Euclidean distance is adopted to characterize the state differences, as shown in the following equation:

$$D|\mathbf{X}_1 - \mathbf{X}_2| = \sqrt{\sum_{i=1}^{MN} (x_{1i} - x_{2i})^2}, \quad (21)$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are the reservoir states generated under different conditions. If the distance is close to zero as the iteration progresses, it can indicate that the difference between the two states has disappeared. That is to say, the corresponding model has the antidisturbance capability. The faster the difference disappears, the stronger the antidisturbance ability of a model, and vice versa.

The impact of state initialization differences on above prediction models is firstly studied. The only variable is the initialized reservoir states, while other model parameters remain the same. The antidisturbance ability of ESN, DESN, VMP-ESN, and VMP-DESN is compared in Figure 3, with the initialized state values as $\mathbf{X}_1 = 0$ and $\mathbf{X}_2 = 0.01$. From Figure 3, it can be observed that (1) the state Euclidean distance of six prediction models gradually decreases to zero as the training progresses; (2) the disturbance curves of VMP-ESN decay are significantly slower than the other five models, and the decay speed of the disturbance curves corresponding to ESN, DESN, and VMP2-DESN is faster than the other three models; and (3) the state difference of ESN, DESN, and VMP2-DESN has decayed to zero, when the training reaches the time step 30. Therefore, it can be concluded that each of the six models has a certain antidisturbance ability, and the antidisturbance ability of ESN, DESN, and VMP2-DESN is stronger than the other three models. The reason is that ESN, DESN, and VMP2-DESN contain their own delayed states, which affect the ability to resist such kind of disturbance caused by initial states.

Then, the ability of six models to resist the second type of disturbance is further explored. The single variable is input signal, while the other model parameters remain unchanged.

TABLE 2: Different evaluation periods for prediction models.

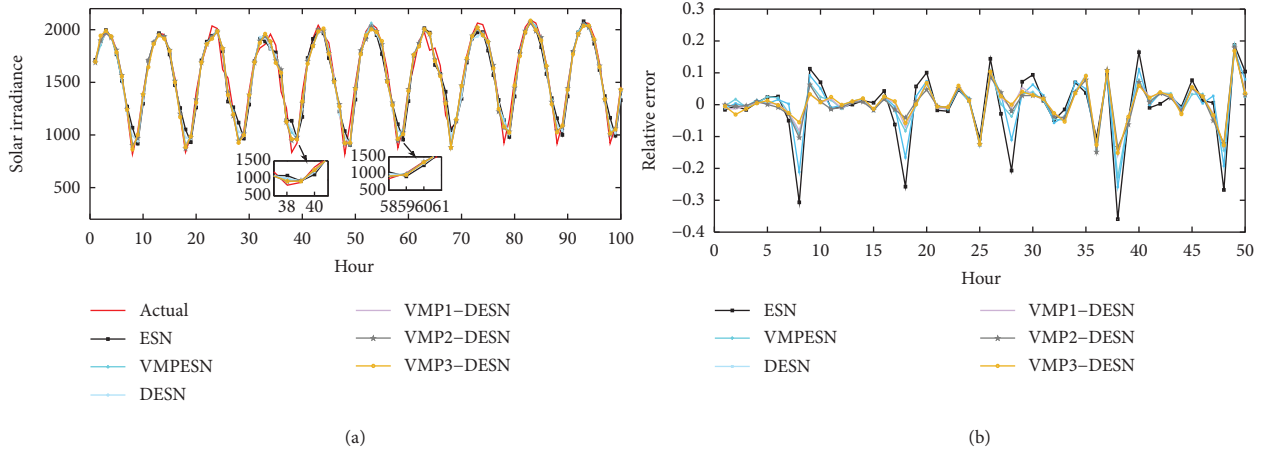| Evaluation period	Training period	Testing period
I	Jan. 1 ~ Feb. 28	Mar. 1 ~ Mar. 31
II	Apr. 1 ~ May. 31	Jun. 1 ~ Jun. 30
III	Jul. 1 ~ Aug. 31	Sep. 1 ~ Sep. 30
IV	Oct. 1 ~ Nov. 30	Dec. 1 ~ Dec. 31

(a)

(b)

FIGURE 2: (a) The actual values versus predicted solar irradiance by different models. (b) The corresponding relative error comparison.

TABLE 3: RMSE, nRMSE, MAE, and MAPE versus delayed time for evaluation period II at Seeley station.

		Seeley			
Model	Parameter	RMSE	nRMSE	MAE	MAPE
VMP-ESN $(\tau_1 = 0, M = 1)$	$\tau_2 = 1$	109.756	0.0696	80.639	0.0662
	$\tau_2 = 2$	77.474	0.0491	54.972	0.0442
	$\tau_2 = 3$	72.403	0.0459	48.903	0.0389
	$\tau_2 = 4$	77.188	0.0489	55.261	0.0444
	$\tau_2 = 5$	129.548	0.0821	96.560	0.0807
VMP1-DESN $(\tau_1 = 0, M = 4)$	$\tau_2 = 1$	84.499	0.0536	61.642	0.0495
	$\tau_2 = 2$	73.313	0.0465	53.647	0.0421
	$\tau_2 = 3$	67.390	0.0427	46.639	0.0360
	$\tau_2 = 4$	75.011	0.0476	54.490	0.0425
	$\tau_2 = 5$	125.346	0.0795	103.069	0.0813
VMP2-DESN $(\tau_2 = 1, M = 4)$	$\tau_1 = 1$	78.260	0.0496	57.778	0.0458
	$\tau_1 = 2$	74.674	0.0473	53.797	0.0425
	$\tau_1 = 3$	73.375	0.0465	51.799	0.0409
	$\tau_1 = 4$	70.243	0.0445	49.499	0.0385
	$\tau_1 = 5$	67.909	0.0431	46.905	0.0364
	$\tau_1 = 6$	67.425	0.0427	46.508	0.0357
	$\tau_1 = 7$	65.608	0.0416	44.645	0.0342
	$\tau_1 = 8$	65.686	0.0416	43.829	0.0333
	$\tau_1 = 9$	69.991	0.0444	46.500	0.0349
	$\tau_1 = 10$	78.880	0.0500	56.207	0.0437
VMP3-DESN $(\tau_1 = \tau_2, M = 4)$	$\tau_1 = 1$	78.260	0.0496	57.778	0.0458
	$\tau_1 = 2$	69.917	0.0443	49.495	0.0386
	$\tau_1 = 3$	65.952	0.0418	44.736	0.0342
	$\tau_1 = 4$	74.159	0.0470	53.202	0.0416
	$\tau_1 = 5$	125.869	0.0798	101.536	0.0812

Assume that the input signal at 200th time step receives a small disturbance and changes from the original value $\mathbf{u}(200)$ to $\mathbf{u}(200) + 0.005$. For better presentation, the semilogarithmic Euclidean distance curves of six models

under input disturbance are plotted and compared in Figure 4. As can be seen from Figure 4, there is an immediate difference between states at time step 200, when the disturbance is added to the input signal. Simultaneously, the
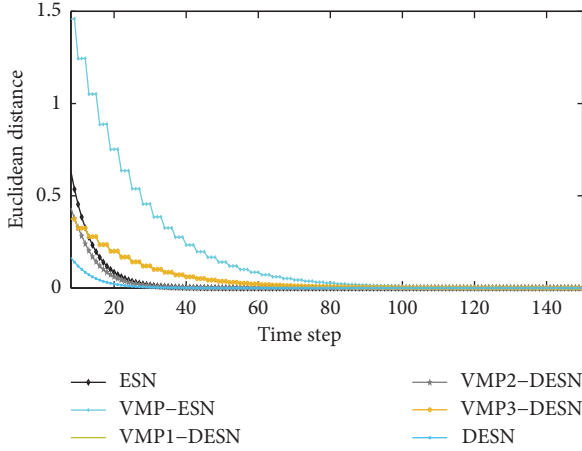
Figure 3: Euclidean distance curves comparison of six prediction models under different initial states.

Figure 4: Semilogarithmic Euclidean distance comparison of six prediction models under different input signals.

difference between states reaches the maximum at time step 200. As the iteration goes on, the impact of input disturbance on six models gradually decreases. In addition, the disturbance curves corresponding to DESN, VMP-DESN2, and ESN decay faster than the other three models, further indicating the stronger ability to resist input disturbance.

The quantitative results in Subsection 3.2 have illustrated that VMP1-DESN, VMP2-DESN, and VMP3-DESN corresponding to the three different types of memory patterns can achieve equivalent performance in improving the prediction accuracy. However, the qualitative analysis results in this subsection show that VMP2-DESN performs better than VMP1-DESN and VMP3-DESN in resisting the input and initial states disturbances.

3.4. Memory Capacity of VMP-DESN.
The memory capacity (MC) of several models is investigated in this section. MC refers to the information length that one network can remember in a short time, i.e., the ability to recall the input signal. According to [33], MC is denoted and computed as follows:

$$MC = \sum_{d=0}^{\infty} MC_d,$$

$$MC_d = R^2 \left(u(t-d), y_d(t) \right),$$

where $y_d(t)$ is the actual network output and $u(t-d)$ is the input signal with a delay d. $R(u(t-d), y_d(t))$ denotes the correlation coefficient between $u(t-d)$ and $y_d(t)$. The goal is to train the readout of a network to recall the input signal with a delay d. That is to say, $u(t-d)$ is the target output with the input signal $u(t)$. The closer MC_d is to 1, the higher the accuracy of a network is to recall the delayed input signal.

In this experiment, MC of traditional ESN, DESN, VMP-ESN, and the proposed VMPs-DESN is calculated and compared. The input signal $u(t)$ is sampled from a uniform distribution $[-1, 1]$, and the upper limit of d is set as 200 for practical consideration. The number of reservoir layer is 4,
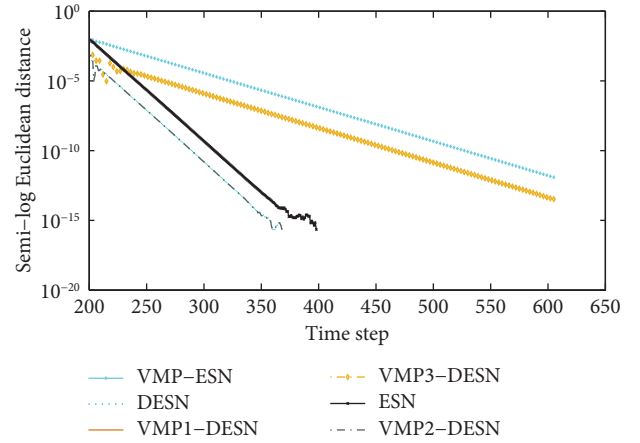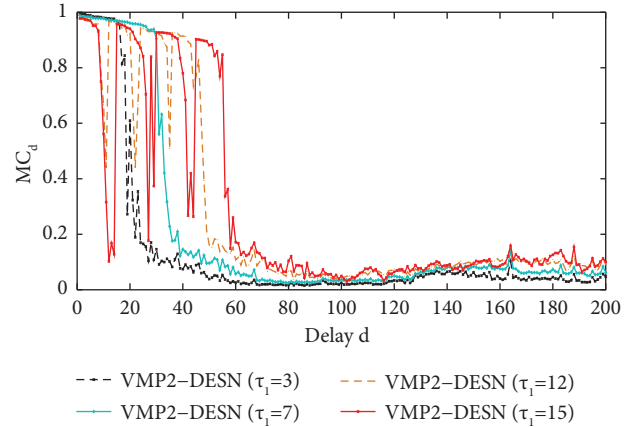

Figure 5: MC_d versus delay d under different delayed time τ_1 in VMP2-DESN.

and the total reservoir size is set as 200, while other model parameters remain unchanged.

Taking a four-layer VMP2-DESN as an example, the influence of delayed time τ_1 on memory capacity is discussed. Figure 5 shows the changes of MC_d with d under different τ_1 in VMP2-DESN. As illustrated in Figure 5, MC_d is almost monotonously decreasing over delay d, when τ_1 is relatively small, such as $\tau_1 = 3$ and $\tau_1 = 7$. However, MC_d decreases to zero through a series of discrete time spans, when τ_1 is large, such as $\tau_1 = 12$ and $\tau_1 = 15$. The reason behind this phenomenon may be the special memory mode in VMP2-DESN. When τ_1 is relatively small, there exist duplicate memories in different reservoir layers, and only a small amount of past input information can be recalled accurately by VMP2-DESN. Duplicate memories denote that some input features preserved in former reservoir layer are also contained in the latter reservoir layer. When τ_1 is large, the duplicate

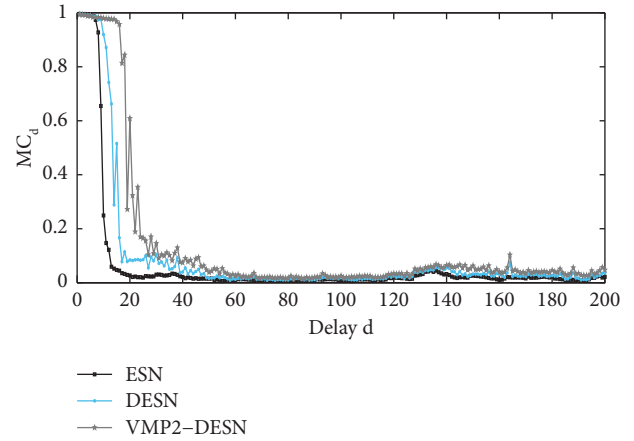TABLE 4: Statistical results comparison of different prediction models for Seeley and Blythe NE stations.

Evaluation period	Model	Seeley				Blythe NE			
		RMSE	nRMSE	MAE	MAPE	RMSE	nRMSE	MAE	MAPE
I	ESN	167.595	0.1384	123.126	0.1495	154.064	0.1460	110.396	0.1904
	DESN	158.339	0.1307	116.498	0.1373	136.354	0.1292	97.528	0.1631
	VMP-ESN	149.746	0.1236	97.804	0.1229	127.295	0.1206	87.527	0.1506
	VMP1-DESN	145.011	0.1197	99.792	0.1209	123.427	0.1170	87.351	0.1479
	VMP2-DESN	148.265	0.1224	98.790	0.1195	117.998	0.1118	82.129	0.1409
	VMP3-DESN	142.745	0.1179	96.283	0.1189	120.941	0.1146	84.456	0.1454
II	ESN	109.756	0.0696	80.639	0.0662	128.531	0.0856	90.449	0.0869
	DESN	84.499	0.0536	61.642	0.0495	90.723	0.0604	59.659	0.0582
	VMP-ESN	72.403	0.0459	48.903	0.0389	65.009	0.0433	43.961	0.0401
	VMP1-DESN	67.390	0.0427	46.639	0.0360	60.566	0.0403	41.645	0.0363
	VMP2-DESN	65.608	0.0416	44.645	0.0342	55.107	0.0367	33.162	0.0296
	VMP3-DESN	65.952	0.0418	44.736	0.0342	58.438	0.0389	40.099	0.0346
III	ESN	185.986	0.1517	126.347	0.1913	194.415	0.1713	127.720	0.4790
	DESN	168.241	0.1372	104.368	0.1707	172.362	0.1519	99.462	0.4144
	VMP-ESN	160.927	0.1313	107.736	0.1597	163.245	0.1438	96.228	0.3936
	VMP1-DESN	161.722	0.1319	104.934	0.1549	157.274	0.1386	90.041	0.3935
	VMP2-DESN	158.641	0.1294	99.464	0.1531	154.356	0.1360	87.808	0.3664
	VMP3-DESN	160.226	0.1307	99.302	0.1582	155.123	0.1367	88.044	0.3878
IV	ESN	233.259	0.3791	177.588	0.6571	119.322	0.1871	89.370	0.7635
	DESN	178.297	0.2897	132.655	0.4350	99.633	0.1562	69.862	0.4839
	VMP-ESN	165.194	0.2685	114.809	0.3371	108.399	0.1699	71.808	0.3168
	VMP1-DESN	154.903	0.2517	109.331	0.3229	100.391	0.1574	68.039	0.3304
	VMP2-DESN	152.537	0.2479	108.142	0.3285	95.838	0.1502	64.519	0.2994
	VMP3-DESN	154.311	0.2508	109.010	0.3193	99.679	0.1563	67.264	0.3258

TABLE 5: MC comparison of six different models.

Model	MC
ESN	12.377
DESN	18.754
VMP-ESN $(\tau_2 = 3)$	11.791
VMP-ESN $(\tau_2 = 7)$	11.612
VMP1-DESN $(\tau_2 = 3)$	17.629
VMP1-DESN $(\tau_2 = 7)$	15.933
VMP3-DESN $(\tau_1 = 3)$	19.682
VMP3-DESN $(\tau_1 = 7)$	18.476
VMP2-DESN $(\tau_1 = 3)$	27.156
VMP2-DESN $(\tau_1 = 7)$	41.617
VMP2-DESN $(\tau_1 = 12)$	54.289
VMP2-DESN $(\tau_1 = 15)$	56.645
VMP2-DESN $(\tau_1 = 17)$	56.225

FIGURE 6: MC_d versus delay d in ESN, DESN, and VMP2-DESN.

memories among reservoirs in VMP2-DESN can be effectively decreased to a certain extent. The input features far away from the current time can also be recalled accurately under this circumstance.

Table 5 and Figure 6 further compare the MC of several models. From the results in Table 5, it can be seen that (1) the MC of DESN is larger than that of ESN due to multiple reservoir layers; (2) VMP-DESN has higher MC than ESN because of the adjustable memory intervals. In addition, VMP2-DESN has the highest MC value among

three different types of VMP-DESN, which further illustrates the advantages of adding delay links between every two subreservoirs. Figure 6 shows MC_d versus delay d in ESN, DESN, and VMP2-DESN. Obviously, VMP2-DESN has a higher MC than ESN and DESN. For ESN and DESN, the memory capacity is limited under constant reservoir size. However, the memory of this proposed VMP-DESN can be designable in advance by adjusting the delayed time. Therefore, VMP-DESN can obtain higher

memory capacity under proper memory pattern and delayed time.

4. Conclusion

A novel deep echo state network is proposed in this paper, i.e., VMP-DESN. This model introduces different types of memory patterns, which is helpful to extract more input features of solar irradiance with long-range dependency. Compared with traditional ESN and DESN, the memory length of VMP-DESN is variable and designable. Therefore, it is more flexible for VMP-DESN to deal with various input signals. The effect of three memory patterns on VMP-DESN is explored by quantitative and qualitative analysis in detail. The quantitative simulation results illustrate that the VMPs-DESN under different memory modes can achieve equivalent performance in improving the prediction accuracy. From the qualitative analysis results, VMP-DESN under the second memory pattern has a stronger antidisturbance ability and a higher memory capacity than the other two modes.

In this current study, the delayed time between every two subreservoirs are set to be the same with the delayed time in each reservoir for simplification. For more complex modeling in practice, the parameters can be different. However, it should be noted that different delayed time will increase the difficulty of parameter optimization.

Data Availability

Data can be found at http://cimis.water.ca.gov.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Qian Li conceptualized the study; Qian Li and Zhijun Guo proposed the methodology; Qian Li, Tao Li, and Zhijun Guo provided software; Qian Li and Tao Li validated the study; Dayong Yang was responsible for formal analysis; Jiangang Ouyang investigated the study; Qian Li was responsible for resources; Qian Li performed data curation; Qian Li wrote original draft; Zhijun Guo reviewed and edited the manuscript; Qian Li visualized the study; Zhijun Guo and Dayong Yang supervised the study; Zhijun Guo was responsible for project administration; Qian Li was responsible for funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

References

[1] C. Voyant, G. Notton, S. Kalogirou et al., "Machine learning methods for solar radiation forecasting: a review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.

[2] H. Verbois, R. Huva, A. Rusydi, and W. Walsh, "Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning," *Solar Energy*, vol. 162, pp. 265–277, 2018.

[3] D. van der Meer, G. R. Chandra Mouli, G. Morales-Espana Mouli, L. R. Elizondo, and P. Bauer, "Energy management system with PV power forecast to optimally charge EVs at the workplace," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 1, pp. 311–320, 2018.

[4] P. Kumari and D. Toshniwal, "Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting," *Applied Energy*, vol. 295, Article ID 117061, 2021.

[5] V. Sharma, D. Yang, W. Walsh, and T. Reindl, "Short term solar irradiance forecasting using a mixed wavelet neural network," *Renewable Energy*, vol. 90, pp. 481–492, 2016.

[6] D. Saez, F. Avila, D. Olivares, C. Canizares, and L. Marin, "Fuzzy prediction interval models for forecasting renewable resources and loads in microgrids," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 548–556, 2015.

[7] A. Ngoc-Lan Huynh, R. C. Deo, M. Ali, S. Abdulla, and N. Raj, "Novel short-term solar radiation hybrid model: long short-term memory network integrated with robust local mean decomposition," *Applied Energy*, vol. 298, no. 10, Article ID 117193, 2021.

[8] A. Qazi, H. Fayaz, A. Wadi, R. G. Raj, N. Rahim, and W. A. Khan, "The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review," *Journal of Cleaner Production*, vol. 104, pp. 1–12, 2015.

[9] X. Cai, N. Zhang, G. K. Venayagamoorthy, and D. C. Wunsch, "Time series prediction with recurrent neural networks trained by a hybrid PSO-EA algorithm," *Neurocomputing*, vol. 70, no. 13-15, pp. 2342–2353, 2007.

[10] H. Jaeger and H. Haas, "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, no. 5667, pp. 78–80, 2004.

[11] A. Rodan and P. Tino, "Minimum complexity echo state network," *IEEE Transactions on Neural Networks*, vol. 22, no. 1, pp. 131–144, 2011.

[12] M. Xu, M. Han, T. Qiu, and H. Lin, "Hybrid regularized echo state network for multivariate chaotic time series prediction," *IEEE Transactions on Cybernetics*, vol. 49, no. 6, pp. 2305–2315, 2019.

[13] H. Zhang, B. Hu, X. Wang et al., "Self-organizing deep belief modular echo state network for time series prediction," *Knowledge-Based Systems*, vol. 222, Article ID 107007, 2021.

[14] Z. Hua, Z. Zheng, M. C. Péra, and F. Gao, "Remaining useful life prediction of PEMFC systems based on the multi-input echo state network," *Applied Energy*, vol. 265, Article ID 114791, 2020.

[15] A. J. Wootton, S. L. Taylor, C. R. Day, and P. W. Haycock, "Optimizing echo state networks for static pattern recognition," *Cognitive Computation*, vol. 9, no. 3, pp. 391–399, 2017.

[16] L. Sun, B. Jin, H. Yang, J. Tong, C. Liu, and H. Xiong, "Unsupervised EEG feature extraction based on echo state network," *Information Sciences*, vol. 475, pp. 1–17, 2019.

[17] C. Liu, H. Zhang, Y. Luo, and K. Zhang, "Echo state network-based online optimal control for discrete-time nonlinear systems," *Applied Mathematics and Computation*, vol. 409, Article ID 126324, 2021.

[18] Q. Chen, L. Shi, J. Na, X. Ren, and Y. Nan, "Adaptive echo state network control for a class of pure-feedback systems with input and output constraints," *Neurocomputing*, vol. 275, pp. 1370–1382, 2018.

[19] S. I. Han and J. M. Lee, "Fuzzy echo state neural networks and funnel dynamic surface control for prescribed performance of a nonlinear dynamic system," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 2, pp. 1099–1112, 2014.

[20] C. Gallicchio, A. Micheli, and L. Pedrelli, "Design of deep echo state networks," *Neural Networks*, vol. 108, pp. 33–47, 2018.

[21] Z. K. Malik, A. Hussain, and Q. J. Wu, "Multilayered echo state machine: a novel architecture and algorithm," *IEEE Transactions on Cybernetics*, vol. 47, no. 4, pp. 946–959, 2017.

[22] Q. Ma, L. Shen, G. W. Cottrell, and DeePr-Esn, "DeePr-ESN: a deep projection-encoding echo-state network," *Information Sciences*, vol. 511, pp. 152–171, 2020.

[23] Q. Li, Z. Wu, R. Ling, L. Feng, and K. Liu, "Multi-reservoir echo state computing for solar irradiance prediction: a fast yet efficient deep learning approach," *Applied Soft Computing*, vol. 95, Article ID 106481, 2020.

[24] H. Jaeger, M. Lukosevicius, D. Popovici, and U. Siewert, "Optimization and applications of echo state networks with leaky-integrator neurons," *Neural Networks*, vol. 20, no. 3, pp. 335–352, 2007.

[25] S. x. Lun, X. s. Yao, and H. f. Hu, "A new echo state network with variable memory length," *Information Sciences*, vol. 370-371, pp. 103–119, 2016.

[26] Y. C. Bo, P. Wang, X. Zhang, and B. Liu, "Modeling data-driven sensor with a novel deep echo state network," *Chemometrics and Intelligent Laboratory Systems*, vol. 206, Article ID 104062, 2020.

[27] Z. Wu, Q. Li, and H. Zhang, "Chain-structure echo state network with stochastic optimization: methodology and application," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 5, pp. 1974–1985, 2022.

[28] Q. Li, Z. Wu, and H. Zhang, "Spatio-temporal modeling with enhanced flexibility and robustness of solar irradiance prediction: a chain-structure echo state network approach," *Journal of Cleaner Production*, vol. 261, Article ID 121151, 2020.

[29] Z. Wu, Q. Li, and X. Xia, "Multi-timescale forecast of solar irradiance based on multi-task learning and echo state network approaches," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 300–310, 2021.

[30] F. Wyffels, B. Schrauwen, and D. Stroobandt, "Stable output feedback in reservoir computing using ridge regression," in *Proceedings of the International Conference on Artificial Neural Networks*, Prague, Czech Republic.

[31] H. Jaeger, "The "echo State" Approach to Analysing and Training Recurrent Neural Networks- with an Erratum Note," GMD Report 148, German National Research Center for Information Technology, Hanover, Germany, 2010.

[32] http://cimis.water.ca.gov.

[33] H. Jaeger, "Short-term memory in echo state networks," GMD Report 152, German National Research Center for Information Technology, Hanover, Germany, 2002.