

Article

Observer-Based State Estimation for Recurrent Neural Networks: An Output-Predicting and LPV-Based Approach

Wanlin Wang, Jinxiong Chen and Zhenkun Huang *

School of Science, Jimei University, Xiamen 361021, China; wanlin19981025@outlook.com (W.W.); chenjinxiong@wuyiu.edu.cn (J.C.)

* Correspondence: hzk974226@jmu.edu.cn

Abstract: An innovative cascade predictor is presented in this study to forecast the state of recurrent neural networks (RNNs) with delayed output. This cascade predictor is a chain-structured observer, as opposed to the conventional single observer, and is made up of several sub-observers that individually estimate the state of the neurons at various periods. This new cascade predictor is more useful than the conventional single observer in predicting neural network states when the output delay is arbitrarily large but known. In contrast to examining the stability of error systems solely employing the Lyapunov–Krasovskii functional (LKF), several new global asymptotic stability standards are obtained by combining the application of the Linear Parameter Varying (LPV) approach, LKF and convex principle. Finally, a series of numerical simulations verify the efficacy of the obtained results.

Keywords: cascade predictor; recurrent neural networks; delayed output; linear parameter varying approach



Citation: Wang, W.; Chen, J.; Huang, Z. Observer-Based State Estimation for Recurrent Neural Networks: An Output-Predicting and LPV-Based Approach. *Math. Comput. Appl.* **2023**, *28*, 104. <https://doi.org/10.3390/mca28060104>

Academic Editor: Guillermo Valencia-Palomo

Received: 5 August 2023

Revised: 12 October 2023

Accepted: 13 October 2023

Published: 25 October 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

Over the past decades, delayed recurrent neural networks were successfully applied in many fields, including pattern recognition, image processing, and combinatorial optimization [1–5], and the dynamic behaviors of RNNs have quickly become a research hotspot. At present, many stability results about the dynamic behavior of RNNs have been obtained [6–12]. Meanwhile, the state information of neurons is very important, because it may participate in the design process of control law, such as feedback control. Therefore, neural networks' state estimation research is of significant importance in practical applications.

The issue of state estimation for RNNs is currently of great interest to many scholars, and many significant results have been made [13–20]. In [13], the authors discussed the state estimation problem for delayed RNNs and obtained delay-independent results using LMI technique. The state estimation problem for Markov jump RNNs with distributed delays was discussed in [14]; the authors proposed an effective LMI technique to solve the problem of neuron states' estimation. The state estimation problem of uncertain RNNs was addressed via a robust state estimator in [15], and it was shown that the suggested robust estimator can be ensured by the feasibility of solving a set of LMIs. An interesting delay partition method was proposed in [17]; the authors used this method to investigate the state estimation problem for delayed static neural networks. In [20], the authors solved the memristive neural networks' (MNNs) state estimation problem by using a novel full-order state observer. It is well known that the effectiveness of the designed observer is usually related to system parameters and the size of various time delays. For example, in most of the works mentioned above, the output states do not have a time delay or the size of the delay is limited to a small range, and a full-order observer is designed based on the measured output. In [21], during the identification of RNN models, a subspace encoder is co-estimated to reconstruct the state of the model from past input and output data.

However, such an explicit form of an observer might run into difficulties if the state delay is not known, and needs an excessively large number of past input–output samples.

For arbitrary large known output delays, it is still an open problem to construct an effective observer to predict the current accurate states of the neuron. In fact, the proposal of the cascade predictor in the field of nonlinear systems has attracted much attention from researchers in recent years. In [22], the authors first proposed a cascade predictor for a class of triangular nonlinear systems that have only output delay; the cascade predictor is made up of a series of subsystems, and each subsystem has a similar structure. However, due to the complexity of the structure of the cascade predictor in [22], the observation estimation is not easily implemented by computer simulation. Since then, the cascade predictor has been studied and a series of results have been achieved on such triangular nonlinear systems with only output delay [23–25]. It can be observed that the cascade predictor has a promising application value in the state estimation of delay systems. It will be a challenge to incorporate cascade predictors into the state estimation of RNNs, and new state estimation approaches may arise.

Inspired by the arguments mentioned above, we focus on the state estimation of delayed RNNs based on cascade predictors. The designed cascade predictor is composed of a limited amount of subsystems; each subsystem estimates the neuron states at different delays, and the last subsystem estimates the current actual states of the neuron. The following are the paper's primary innovative ideas.

(1) This paper theoretically describes the reason why a single observer cannot observe the neuron state information when the output delay is large enough. Then, inspired by [22,23], we design a new cascade predictor for estimating the state of RNNs with state and output delays. To our best knowledge, this is the first time that a cascade predictor has been applied to state estimation in neural networks.

(2) For the activation function, most papers usually use the traditional Lipschitz condition hypothesis; however, for the activation with the large Lipschitz constant, it may indirectly lead to the conservatism of the design process and theoretical results. To overcome these difficulties, a new reformulated Lipschitz property of the activation function, which is the outcome of applying the LPV approach to the Lipschitz condition, is provided. This property is motivated by [26,27] and can lessen conservatism in the observer design process.

(3) In contrast to [14,16,20], the case when the output states have an arbitrarily large delay is explored, and the state prediction problem of delayed RNNs is resolved based on the measured output. A set of LMIs may be used to calculate the observer gain, and new adequate requirements for the global asymptotic stability of each error system are obtained based on the LKF, the LMI technique, and the convex principle.

The structure of this paper is as follows. The RNNs model and its associated assumptions are introduced in Section 2 of this article. The main results of this paper are presented in Section 3. The efficiency of the results obtained is demonstrated in Section 4 by numerical simulations. Section 5 closes with a general conclusion.

Notations: \mathcal{R} , \mathcal{Z} denote, respectively, the set of real numbers and the positive integer set. \mathcal{R}^n represents the n dimensional Euclidean space with the Euclidean norm $\|\cdot\|$. $\mathcal{R}^{n \times m}$ denotes the set of all $n \times m$ real matrices. The superscript “ T ” and “ -1 ” represent the transpose and inverse of a matrix. $X > Y$ ($X < Y$) means that $X - Y$ is a positive (negative) matrix. $\|A\|$ denotes the operator norm of matrix A , i.e., $A = (a_{ij})$, $\|A\| = \sqrt{\lambda_{\max}(A^T A)}$, where $\lambda_{\max}(A)$ is the largest eigenvalue of A . $\text{diag}\{\cdot\cdot\cdot\}$ represents a block diagonal matrix. The symbol “ $*$ ” denotes the symmetric term of the matrix. Let $\tau > 0$, $\mathcal{C}([-\tau, 0]; \mathcal{R}^n)$ denote the family of continuous functions ψ from $[-\tau, 0]$ to \mathcal{R}^n . I stands for an identity matrix with the proper dimensions.

2. Problem Formulation

Consider the following RNNs with delayed output [28,29]:

$$\begin{cases} \dot{x}(t) = -Ax(t) + W_0f(x(t)) + W_1f(x(t - h_x)), \\ y(t) = Cx(t - h_y), \\ x(s) = \phi(s), s \in [-\tau, 0], \end{cases} \tag{1}$$

where $x(t) = [x_1(t), \dots, x_n(t)]^T \in \mathbb{R}^n$ denotes the state vector, $A = \text{diag}\{a_1, \dots, a_n\}$ is a diagonal matrix with $a_i > 0$, W_0 and W_1 represent the connection weight matrices, $f(x(t)) = [f_1(x_1(t)), \dots, f_n(x_n(t))]^T \in \mathbb{R}^n$ denotes the activation function, C is a output matrix, $y(t)$ represents the measured output. h_x, h_y denote the known discrete delays, and $\phi(s) \in \mathcal{C}([-\tau, 0]; \mathbb{R}^n)$ is an initial condition, $\tau = \max\{h_x, h_y\}$.

The primary objective of this research is to construct an effective observer that can accurately predict the neuron states when the output delay h_y is arbitrarily large yet known. Next, for a subsequent analysis, the following corresponding lemmas are given.

Assumption 1. *The activation function $f_i(\cdot)$ is bounded and satisfies*

$$|f_i(u) - f_i(v)| \leq l_i|u - v|, \forall u, v \in \mathbb{R}, \tag{2}$$

where $f_i(0) = 0$ and $l_i > 0$ is a Lipschitz constant.

Some conservative conditions in the observer design may result from the Lipschitz condition of the activation function in (2). However, it is generally known that LPV approach can reduce the Lipschitz condition’s conservatism, making it useful for designing observers for nonlinear systems with a large Lipschitz constant [26,27]. Here, we will extend this method to RNNs (1) and derive the subsequent lemma.

Lemma 1. *The activation function $f(\cdot)$ has the following two properties that are equal:*

(1) *Lipschitz property: $f_i(\cdot)$ is l_i -Lipschitz, i.e.,*

$$|f_i(x_i) - f_i(y_i)| \leq l_i|x_i - y_i|, \forall x_i, y_i \in \mathbb{R}. \tag{3}$$

(2) *Lipschitz property reformulated: for all $i = 1, \dots, n$, there exist functions $\psi_{ii}(t) : \mathbb{R} \rightarrow \mathbb{R}$ and constants $\overline{\gamma}_{ii}, \underline{\gamma}_{ii}$, such that*

$$f(x) - f(y) = \sum_{i=1}^n \psi_{ii}(t)H_{ii}(x - y), \forall x, y \in \mathbb{R}^n \tag{4}$$

with $\underline{\gamma}_{ii} \leq \psi_{ii}(t) \leq \overline{\gamma}_{ii}$, where $H_{ii} = e_n(i)e_n^T(i)$ and $e_n(i) = [0, \dots, \overbrace{1}^{i\text{-th}}, \dots, 0]^T \in \mathbb{R}^n$.

The proofs of Lemma 1 are similar to Lemma 6 and Lemma 7 in [26]; we omit it here. Note that $\psi_{ii}(t)$ in (4) is expressed as follows

$$\psi_{ii}(t) = \begin{cases} \frac{f_i(x_i) - f_i(y_i)}{x_i - y_i}, & x_i \neq y_i, \\ 0, & x_i = y_i. \end{cases} \tag{5}$$

Remark 1. *Compared with the traditional global Lipschitz condition hypothesis of activation functions in [13–16,30], the reformulation (4) in Lemma 1 offers a best less conservative Lipschitz condition and deals with the activation functions $f(x)$ with the best accuracy. For instance, for $f(x) = [\tanh(x_1) + \frac{1}{2}\sin(x_1), \frac{1}{2}\cos(x_2) + \frac{1}{2}(|x_2 + 1| - |x_2 - 1|)]^T$, we have $\underline{\gamma}_{11} = -\frac{1}{2}$, $\overline{\gamma}_{11} = \frac{3}{2}$, $\underline{\gamma}_{22} = -\frac{1}{2}$, $\overline{\gamma}_{22} = \frac{3}{2}$. For $g(x) = [\frac{3}{4}(|x_1 + 1| - |x_1 - 1|), \frac{3}{2}\tanh(x_2)]^T$, we have $\underline{\gamma}_{11} = 0$, $\overline{\gamma}_{11} = \frac{3}{2}$, $\underline{\gamma}_{22} = 0$, $\overline{\gamma}_{22} = \frac{3}{2}$. If we use the global Lipschitz condition in Assumption 1,*

we can only obtain $|f_i(x_i) - f_i(y_i)| \leq \frac{3}{2}|x_i - y_i|$ and $|g_i(x_i) - g_i(y_i)| \leq \frac{3}{2}|x_i - y_i|, i = 1, 2$, we cannot accurately distinguish between $f(x)$ and $g(x)$, and the related properties of $f(x)$ and $g(x)$ cannot be effectively utilized.

Lemma 2 (Finsler’s Lemma [31]). For $x \in \mathcal{R}^n, M \in \mathcal{R}^{n \times n}$ is a symmetric matrix, and the matrix $\mathcal{G} \in \mathcal{R}^{m \times n}$, such that $\text{rank}(\mathcal{G}) < n$. The subsequent properties are equivalent:

- (1) $x^T M x < 0, \forall x \in \mathcal{R}^n / \mathcal{G} x = 0, x \neq 0,$
- (2) $\mathcal{G}^{\perp T} M \mathcal{G}^{\perp} < 0,$

where \mathcal{G}^{\perp} is a right orthogonal complement of \mathcal{G} .

Lemma 3 (Moon’s Inequality [32]). Assume that $x(s) \in \mathcal{R}^{n_a}, y(s) \in \mathcal{R}^{n_b}$ are defined on the interval Ω and $Y \in \mathcal{R}^{n_a \times n_b}$. Then, for matrices $D \in \mathcal{R}^{n_a \times n_a}, T \in \mathcal{R}^{n_a \times n_b}$ and $Z \in \mathcal{R}^{n_b \times n_b}$, the following holds:

$$-2 \int_{\Omega} x^T(s) Y y(s) ds \leq \int_{\Omega} \begin{bmatrix} x(s) \\ y(s) \end{bmatrix}^T \cdot \begin{bmatrix} D & T - Y \\ * & Z \end{bmatrix} \cdot \begin{bmatrix} x(s) \\ y(s) \end{bmatrix} ds,$$

where

$$\begin{bmatrix} D & T \\ * & Z \end{bmatrix} > 0.$$

3. Results

3.1. Single Observer

In this section, we will employ a full-order observer to handle the matter of RNNs’ state estimation. First, in order to accurately estimate the RNNs’ state information, we will design the full-order observer based on the measured delayed output as

$$\hat{\dot{x}}(t) = -A\hat{x}(t) + W_0 f(\hat{x}(t)) + W_1 f(\hat{x}(t - h_x)) + L(C\hat{x}(t - h_y) - y(t)), \tag{6}$$

where $\hat{x}(t)$ is an estimation of the state $x(t)$ of (1). Then, by defining the estimation error $e(t) = \hat{x}(t) - x(t)$, we can obtain the error system given, as follows

$$\dot{e}(t) = -Ae(t) + W_0 \Delta f(t) + W_1 \Delta f(t - h_x) + LCe(t - h_y), \tag{7}$$

where $\Delta f(t) = f(\hat{x}(t)) - f(x(t))$. Due to Lemma 1, there are functions $\psi_{ii}(t)$ and $\psi_{ii}^{h_x}(t)$, such that

$$\begin{cases} \Delta f(t) = \sum_{i=1}^n \psi_{ii}(t) H_{ii} e(t), \\ \Delta f(t - h_x) = \sum_{i=1}^n \psi_{ii}^{h_x}(t) H_{ii} e(t - h_x), \end{cases} \tag{8}$$

where $\psi_{ii}^{h_x}(t) = \psi_{ii}(t - h_x)$.

Define the time-varying matrices $\Psi(t) = \text{diag}\{\psi_{11}(t), \dots, \psi_{nn}(t)\}, \Psi_{h_x}(t) = \text{diag}\{\psi_{11}^{h_x}(t), \dots, \psi_{nn}^{h_x}(t)\}$, and bounded convex set \mathcal{H}_n , where the vertex set of \mathcal{H}_n is defined as

$$\mathcal{V}_{\mathcal{H}_n} = [\phi = \text{diag}[\phi_{11}, \phi_{22}, \dots, \phi_{nn}] \in \mathcal{R}^{n \times n} | \phi_{ii} \in \{\underline{\gamma}_{ii}, \overline{\gamma}_{ii}\}]. \tag{9}$$

It is obvious that time-varying matrix parameters $\Psi(t)$ and $\Psi_{h_x}(t)$ belong to the bounded convex set \mathcal{H}_n . Now, we define the following matrices:

$$\begin{cases} \mathcal{A}(\Psi(t)) = -A + W_0 \sum_{i=1}^n \psi_{ii}(t)H_{ii} = -A + W_0\Psi(t), \\ \mathcal{B}(\Psi_{h_x}(t)) = W_1 \sum_{i=1}^n \psi_{ii}^{h_x}(t)H_{ii} = W_1\Psi_{h_x}(t), \end{cases} \tag{10}$$

then, by using (10), the LPV error system (7) can be reconstructed as

$$\dot{e}(t) = \mathcal{A}(\Psi(t))e(t) + \mathcal{B}(\Psi_{h_x}(t))e(t - h_x) + LCe(t - h_y). \tag{11}$$

The sufficient condition for the global asymptotic stability of error system (11) is presented in the following theorem.

Theorem 1. *The error system (11) is globally asymptotically stable for all $h_y \in [0, h^*]$, if there exist matrices $P > 0, Q > 0, M > 0, Z > 0, S > 0$, a matrix R , and positive scalars $\rho_i > 0, i = 1, 2$, such that for $\forall \Psi, \Psi_{h_x} \in \mathcal{V}_{\mathcal{H}_n}$, the following LMIs hold with observer gain $L = P^{-1}R$:*

$$\begin{bmatrix} \Omega_1 & P\mathcal{B}(\Psi_{h_x}) + \frac{Z}{h_x} & RC + \frac{S}{h^*} & h_x\mathcal{A}^T(\Psi)P & h^*\mathcal{A}^T(\Psi)P \\ * & \Omega_2 & 0 & h_x\mathcal{B}^T(\Psi_{h_x})P & h^*\mathcal{B}^T(\Psi_{h_x})P \\ * & * & \Omega_3 & h_xC^TR^T & h^*C^TR^T \\ * & * & * & \Omega_4 & 0 \\ * & * & * & * & \Omega_5 \end{bmatrix} < 0, \tag{12}$$

where $\Omega_1 = P\mathcal{A}(\Psi) + \mathcal{A}^T(\Psi)P + Q + M - \frac{Z}{h_x} - \frac{S}{h^*}$, $\Omega_2 = -Q - \frac{Z}{h_x}$, $\Omega_3 = -M - \frac{S}{h^*}$, $\Omega_4 = -h_x(2\rho_1P - \rho_1^2Z)$ and $\Omega_5 = -h^*(2\rho_2P - \rho_2^2S)$.

Proof. Consider the following Lyapunov–Krasovskii functions as

$$\begin{aligned} V(t) = & e^T(t)Pe(t) + \int_{t-h_x}^t e^T(\tau)Qe(\tau)d\tau + \int_{t-h_y}^t e^T(\tau)Me(\tau)d\tau \\ & + \int_{-h_x}^0 \int_{t+\beta}^t \dot{e}^T(\tau)Z\dot{e}(\tau)d\tau d\beta + \int_{-h_y}^0 \int_{t+\beta}^t \dot{e}^T(\tau)S\dot{e}(\tau)d\tau d\beta, \end{aligned} \tag{13}$$

and the time derivative of $V(t)$ can be evaluated as

$$\begin{aligned} \dot{V}(t) = & 2e^T(t)P\dot{e}(t) + e^T(t)Qe(t) - e^T(t - h_x)Qe(t - h_x) + e^T(t)Me(t) \\ & - e^T(t - h_y)Me(t - h_y) + h_x\dot{e}^T(t)Z\dot{e}(t) - \int_{t-h_x}^t \dot{e}^T(\tau)Z\dot{e}(\tau)d\tau \\ & + h_y\dot{e}^T(t)S\dot{e}(t) - \int_{t-h_y}^t \dot{e}^T(\tau)S\dot{e}(\tau)d\tau. \end{aligned} \tag{14}$$

Applying the Jensen’s Inequality [33], we can obtain

$$- \int_{t-h_x}^t \dot{e}^T(\tau)Z\dot{e}(\tau)d\tau \leq -\frac{1}{h_x}\Delta e_{h_x}^T(t)Z\Delta e_{h_x}(t), \tag{15}$$

$$- \int_{t-h_y}^t \dot{e}^T(\tau)S\dot{e}(\tau)d\tau \leq -\frac{1}{h_y}\Delta e_{h_y}^T(t)S\Delta e_{h_y}(t), \tag{16}$$

where $\Delta e_{h_x}(t) = e(t) - e(t - h_x)$, $\Delta e_{h_y}(t) = e(t) - e(t - h_y)$.

By using (14)–(16), $\dot{V}(t)$ satisfies

$$\dot{V}(t) \leq \chi^T(t)Y(h_x, h_y)\chi(t), \tag{17}$$

where

$$\chi(t) = [\hat{e}^T(t), e^T(t), e^T(t - h_x), e^T(t - h_y), \Delta e_{h_x}^T(t), \Delta e_{h_y}^T(t)]^T,$$

$$Y(h_x, h_y) = \begin{bmatrix} h_x Z + h_y S & P & 0 & 0 & 0 & 0 \\ * & Q + M & 0 & 0 & 0 & 0 \\ * & * & -Q & 0 & 0 & 0 \\ * & * & * & -M & 0 & 0 \\ * & * & * & * & -\frac{Z}{h_x} & 0 \\ * & * & * & * & * & -\frac{S}{h_y} \end{bmatrix}.$$

Moreover, it follows from the error system (11) and the definition of $y_{h_x}(t)$ and $y_{h_y}(t)$ that $\Gamma(\Psi(t), \Psi_{h_x}(t))\chi(t) = 0$ with

$$\Gamma(\Psi(t), \Psi_{h_x}(t)) = \begin{bmatrix} I & -\mathcal{A}(\Psi(t)) & -\mathcal{B}(\Psi_{h_x}(t)) & -P^{-1}RC & 0 & 0 \\ 0 & -I & I & 0 & I & 0 \\ 0 & -I & 0 & I & 0 & 0 \end{bmatrix}.$$

Therefore, the error system (11) is globally asymptotically stable if, for all $\Gamma(\Psi(t), \Psi_{h_x}(t))\chi(t) = 0$ with $\chi(t) \neq 0$, there holds $\chi^T(t)Y(h_x, h_y)\chi(t) < 0$. Then, according to Lemma 2 and the convexity principle [34], $\chi^T(t)Y(h_x, h_y)\chi(t) < 0$ is equivalent to

$$(\Gamma^\perp(\Psi, \Psi_{h_x}))^T Y(h_x, h_y) \Gamma^\perp(\Psi, \Psi_{h_x}) < 0, \forall \Psi, \Psi_{h_x} \in \mathcal{V}_{\mathcal{H}_n}, \tag{18}$$

where $\Gamma^\perp(\Psi, \Psi_{h_x})$ is a right orthogonal complement of $\Gamma(\Psi, \Psi_{h_x})$ and

$$\Gamma^\perp(\Psi, \Psi_{h_x}) = \begin{bmatrix} \mathcal{A}(\Psi) & \mathcal{B}(\Psi_{h_x}) & P^{-1}RC \\ I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ I & -I & 0 \\ I & 0 & -I \end{bmatrix}. \tag{19}$$

Further, (18) can be rewritten as

$$\begin{bmatrix} P\mathcal{A}(\Psi) + \mathcal{A}^T(\Psi)P + Q + M & P\mathcal{B}(\Psi_{h_x}) & RC \\ * & -Q & 0 \\ * & * & -M \end{bmatrix} + h_x \Pi_1 Z \Pi_1^T - \frac{1}{h_x} \Pi_2 Z \Pi_2^T + h_y \Pi_1 S \Pi_1^T - \frac{1}{h_y} \Pi_3 S \Pi_3^T < 0, \forall \Psi, \Psi_{h_x} \in \mathcal{V}_{\mathcal{H}_n}, \tag{20}$$

where $\Pi_1 = [\mathcal{A}(\Psi), \mathcal{B}(\Psi_{h_x}), P^{-1}RC]^T$, $\Pi_2 = [I, -I, 0]^T$ and $\Pi_3 = [I, 0, -I]^T$. Since $Z > 0$ and $S > 0$, (20) cannot hold when h_y is large enough. Therefore, there exists an upper bound h^* for h_y such that (20) holds when $h_y \in [0, h^*]$, and $(\Gamma^\perp(\Psi, \Psi_{h_x}))^T Y(h_x, h_y) \Gamma^\perp(\Psi, \Psi_{h_x}) < 0, \forall \Psi, \Psi_{h_x} \in \mathcal{V}_{\mathcal{H}_n}$ is a sufficient condition for (18).

By employing a Schur complement [34], $(\Gamma^\perp(\Psi, \Psi_{h_x}))^T Y(h_x, h_y) \Gamma^\perp(\Psi, \Psi_{h_x}) < 0$ with $\forall \Psi, \Psi_{h_x} \in \mathcal{V}_{\mathcal{H}_n}$ is equivalent to

$$\begin{bmatrix} \Omega_1 & P\mathcal{B}(\Psi_{h_x}) + \frac{Z}{h_x} & RC + \frac{S}{h^*} & h_x \mathcal{A}^T(\Psi)Z & h^* \mathcal{A}^T(\Psi)S \\ * & \Omega_2 & 0 & h_x \mathcal{B}^T(\Psi_{h_x})Z & h^* \mathcal{B}^T(\Psi_{h_x})S \\ * & * & \Omega_3 & h_x C^T R^T P^{-1}Z & h^* C^T R^T P^{-1}S \\ * & * & * & -h_x Z & 0 \\ * & * & * & * & -h^* S \end{bmatrix} < 0, \tag{21}$$

with $\forall \Psi, \Psi_{h_x} \in \mathcal{V}_{\mathcal{H}_n}$, then, multiplying both sides of (21) on the left and on the right by $\text{diag}\{I, I, I, PZ^{-1}, PS^{-1}\}$ and its transpose, respectively, and using the inequalities $-PZ^{-1}P \leq -2\rho_1P + \rho_1^2Z$, $-PS^{-1}P \leq -2\rho_2P + \rho_2^2S$, we can deduce that (12) is a sufficient condition for (21). We finish the proof. \square

Remark 2. Different from the stability analysis of the nonlinear error system in [12,13,15,16,20], we provide an LPV formulation of the error system for RNNs with Lipschitz activation functions, which leads us to study the stability of the linear error system (11) by using the convexity principle. Obviously, our LPV-based approach is a useful tool for the state estimation of RNNs.

Remark 3. It follows from Theorem 1, which depends on the output delay, that the designed full-order observer (6) cannot predict the current state of RNNs (1) if $h_y \gg h^*$. This is due to our inability to choose the appropriate observer gain L to stabilize the error system (11). From the later numerical simulations, it is clear that this is a drawback of the full-order observer. In addition, the conditions shown in (12) can be checked against a set of fixed values by standard LMI routines and an estimate of h^* is obtained. The algorithm for finding a feasible solution to (12) is summarized as follows:

Step 1: Fix the value of h^* to constant \bar{h} and make an initial guess for \bar{h} .

Step 2: Fix the value of ρ_1, ρ_2 to some constants $\bar{\rho}_1, \bar{\rho}_2$ and make an initial guess for the values of $\bar{\rho}_1, \bar{\rho}_2$.

Step 3: Solve the LMI (12) for L with the fixed values $\bar{\rho}_1, \bar{\rho}_2$ and \bar{h} ; if a feasible value of L cannot be computed, return to step 2 to reset the initial values of $\bar{\rho}_1$ and $\bar{\rho}_2$; if a feasible value of L can be computed, return to step 1 and increase the value of \bar{h} until L cannot be solved.

3.2. Cascade Predictor

If $h \gg h^*$, the full-order observer (6) will fail. In this case, cascade predictor design can be used to solve this problem. Let $h = \frac{h_y}{m}, m \in \mathcal{Z}$ and define

$$x_i(t) : \begin{cases} \dot{x}_i(t) = -Ax_i(t) + W_0f(x_i(t)) + W_1f(x_i(t - h_x)), t \in [(m - i)h, +\infty), \\ x_i(s) = \phi(s - (m - i)h), s \in [-\tau + (m - i)h, (m - i)h], \end{cases} \tag{22}$$

where $i = 1, 2, \dots, m$. Through mathematical analysis, we obtain $x_i(t) = x(t - h_y + i \cdot h) = x_{i+1}(t - h), i = 1, 2, \dots, m - 1$ and $x_m(t) = x(t)$. Then, the cascade predictor can be constructed as

$$\begin{cases} \dot{\hat{x}}_1(t) = -A\hat{x}_1(t) + W_0f(\hat{x}_1(t)) + W_1f(\hat{x}_1(t - h_x)) + L_1(C\hat{x}_1(t - h) - y(t)), \\ \dot{\hat{x}}_2(t) = -A\hat{x}_2(t) + W_0f(\hat{x}_2(t)) + W_1f(\hat{x}_2(t - h_x)) + L_2(C\hat{x}_2(t - h) - \hat{y}_1(t)), \\ \vdots \\ \dot{\hat{x}}_m(t) = -A\hat{x}_m(t) + W_0f(\hat{x}_m(t)) + W_1f(\hat{x}_m(t - h_x)) + L_m(C\hat{x}_m(t - h) - \hat{y}_{m-1}(t)), \end{cases} \tag{23}$$

where $\phi_i(s) \in \mathcal{C}([-\tau_1, 0]; \mathcal{R}^n)$ is an initial condition for the subsystem $\hat{x}_i, i = 1, 2, \dots, m$, $\tau_1 = \max\{h_x, h\}$, and $\hat{y}_i(t) = Cx_i(t), i = 1, 2, \dots, m - 1$. In the cascade predictor (23), the subsystem $\hat{x}_i(t)$ estimates the state $x_i(t), i = 1, 2, \dots, m - 1$, and the subsystem $\hat{x}_m(t)$ estimates the state $x(t)$.

Remark 4. The idea behind the cascade predictor (23) is that regardless of how long the output delay h_y is, we can split it into m small time periods. Then, each sub-observer $\hat{x}_i(t)$ in the cascade predictor estimates the delayed state $x(t - h_y + i \frac{h_y}{m})$, and the last sub-observer $\hat{x}_m(t)$ estimates the current state $x(t)$. Compared with [14–20], which discuss the state estimation of neural networks with only a small output delay or even no output delay using a full-order observer, the output delay h_y in this paper is arbitrarily large yet known, and, in this sense, this is an advancement in the study of neural networks' state estimation.

Remark 5. Moreover, the idea of this novel predictor was first proposed in [22], and the authors used this predictor with a chain structure for a class of triangular nonlinear systems with only the output delay. In this paper, we will discuss the state estimation problem of RNNs with both the state delay and output delay using this novel cascade predictor.

Next, define the estimation error $e_i(t) = \hat{x}_i(t) - x_i(t), i = 1, 2, \dots, m$; then, similar to (10)–(11), which use the LPV approach, we can obtain the following error systems

$$\begin{cases} \dot{e}_1(t) = \mathcal{A}(\Psi(t))e_1(t) + \mathcal{B}(\Psi_{h_x}(t))e_1(t - h_x) + L_1Ce_1(t - h), \\ \dot{e}_j(t) = \mathcal{A}(\Psi(t))e_j(t) + \mathcal{B}(\Psi_{h_x}(t))e_j(t - h_x) + L_jCe_j(t - h) - L_j e_{j-1}(t), j = 2, 3, \dots, m. \end{cases} \tag{24}$$

Theorem 2. For given output delay h_y and scalar $m \in \mathcal{L}$, the error systems (24) are globally asymptotically stable if there exist matrices $P_j > 0, Q_j > 0, M_j > 0, Z_j > 0, S_j > 0, X_i > 0, D_i > 0, i = 1, 2, \dots, m$, matrices $R_i, Y_i, T_i, i = 1, 2, \dots, m$ and a positive scalar $\gamma > 0$, such that $L_i = P_i^{-1}R_i, i = 1, 2, \dots, m$ and the following LMIs are satisfied:

$$\begin{bmatrix} \Xi_i & -T_i + P_i\mathcal{B}(\Psi_{h_x}) & -Y_i + R_iC & \gamma\mathcal{A}(\Psi)P_i \\ * & -Q_i & 0 & \gamma\mathcal{B}^T(\Psi_{h_x})P_i \\ * & * & -M_i & \gamma C^T R_i^T \\ * & * & * & -2\gamma P_i + h_x Z_i + h S_i \end{bmatrix} < 0, \tag{25}$$

with $\forall \Psi \in \mathcal{V}_{\mathcal{H}_n}, \forall \Psi_{h_x} \in \mathcal{V}_{\mathcal{H}_n}$ and

$$\begin{bmatrix} X_i & Y_i \\ * & S_i \end{bmatrix} < 0, \tag{26}$$

$$\begin{bmatrix} D_i & T_i \\ * & Z_i \end{bmatrix} < 0, \tag{27}$$

where $\Xi_i = P_i\mathcal{A}(\Psi) + \mathcal{A}^T(\Psi)P_i + T_i + T_i^T + Y_i + Y_i^T + h_x D_i + h X_i + Q_i + M_i$.

Proof. The stability of the error systems (24) will be proved gradually:

Step 1: we consider the first error system $e_1(t)$ in (24):

$$\dot{e}_1(t) = \mathcal{A}(\Psi(t))e_1(t) + \mathcal{B}(\Psi_{h_x}(t))e_1(t - h_x) + L_1Ce_1(t - h). \tag{28}$$

Using the Newton–Leibniz formula, we have

$$\begin{cases} e_1(t - h_x) = e_1(t) - \int_{t-h_x}^t \dot{e}_1(\tau)d\tau, \\ e_1(t - h) = e_1(t) - \int_{t-h}^t \dot{e}_1(\tau)d\tau, \end{cases} \tag{29}$$

then, $\dot{e}_1(t)$ in (28) can be reconstructed as

$$\dot{e}_1(t) = [\mathcal{A}(\Psi(t)) + \mathcal{B}(\Psi_{h_x}(t)) + L_1C]e_1(t) - \mathcal{B}(\Psi_{h_x}(t)) \int_{t-h_x}^t \dot{e}_1(\tau)d\tau - L_1C \int_{t-h}^t \dot{e}_1(\tau)d\tau. \tag{30}$$

Constructing the Lyapunov–Krasovskii functions as follows

$$\begin{aligned}
 V_1(t) = & \overbrace{e_1^T(t)P_1e_1(t)}^{V_{11}(t)} + \overbrace{\int_{t-h_x}^t e_1^T(\tau)Q_1e_1(\tau)d\tau}^{V_{12}(t)} + \overbrace{\int_{t-h}^t e_1^T(\tau)M_1e_1(\tau)d\tau}^{V_{13}(t)} \\
 & + \overbrace{\int_{-h_x}^0 \int_{t+\eta}^t \dot{e}_1^T(\tau)Z_1\dot{e}_1(\tau)d\tau d\eta}^{V_{14}(t)} + \overbrace{\int_{-h}^0 \int_{t+\eta}^t \dot{e}_1^T(\tau)S_1\dot{e}_1(\tau)d\tau d\eta}^{V_{15}(t)}. \tag{31}
 \end{aligned}$$

then, the derivative of $V_{11}(t)$ along (30) is as follows

$$\begin{aligned}
 \dot{V}_{11}(t) = & e_1^T(t) [P_1\mathcal{A}(\Psi(t)) + \mathcal{A}^T(\Psi(t))P_1 + P_1\mathcal{B}(\Psi_{h_x}(t)) + \mathcal{B}^T(\Psi_{h_x}(t))P_1 + R_1C + C^TR_1] \\
 & \times e_1(t) - 2e_1^T(t)P_1\mathcal{B}(\Psi_{h_x}(t)) \int_{t-h_x}^t \dot{e}_1(\tau)d\tau - 2e_1^T(t)R_1C \int_{t-h}^t \dot{e}_1(\tau)d\tau. \tag{32}
 \end{aligned}$$

According to Lemma 2 and (26)–(27), we have

$$\begin{aligned}
 -2e_1^T(t)P_1\mathcal{B}(\Psi_{h_x}(t)) \int_{t-h_x}^t \dot{e}_1(\tau)d\tau \leq & h_x e_1^T(t)D_1e_1(t) + 2e_1^T(t)[T_1 - P_1\mathcal{B}(\Psi_{h_x}(t))] \\
 & \times [e_1(t) - e_1(t - h_x)] + \int_{t-h_x}^t \dot{e}_1^T(\tau)Z_1\dot{e}_1(\tau)d\tau, \tag{33}
 \end{aligned}$$

$$\begin{aligned}
 -2e_1^T(t)R_1C \int_{t-h}^t \dot{e}_1(\tau)d\tau \leq & h e_1^T(t)X_1e_1(t) + 2e_1^T(t)[Y_1 - R_1C][e_1(t) - e_1(t - h)] \\
 & + \int_{t-h}^t \dot{e}_1^T(\tau)S_1\dot{e}_1(\tau)d\tau. \tag{34}
 \end{aligned}$$

By (32)–(34), $\dot{V}_{11}(t)$ satisfies

$$\begin{aligned}
 \dot{V}_{11} \leq & e_1^T(t) [P_1\mathcal{A}(\Psi(t)) + \mathcal{A}^T(\Psi(t))P_1 + T_1^T + T_1 + Y_1^T + Y_1 + h_xD_1 + hX_1]e_1(t) \\
 & - 2e_1^T(t)[T_1 - P_1\mathcal{B}(\Psi_{h_x}(t))]e_1(t - h_x) - 2e_1^T(t)[Y_1 - R_1C]e_1(t - h) \\
 & + \int_{t-h_x}^t \dot{e}_1^T(\tau)Z_1\dot{e}_1(\tau)d\tau + \int_{t-h}^t \dot{e}_1^T(\tau)S_1\dot{e}_1(\tau)d\tau. \tag{35}
 \end{aligned}$$

The derivatives of $V_{12}(t)$, $V_{13}(t)$, $V_{14}(t)$ and $V_{15}(t)$ are as follows

$$\dot{V}_{12}(t) = e_1^T(t)Q_1e_1(t) - e_1^T(t - h_x)Q_1e_1(t - h_x), \tag{36}$$

$$\dot{V}_{13}(t) = e_1^T(t)M_1e_1(t) - e_1^T(t - h)M_1e_1(t - h), \tag{37}$$

$$\dot{V}_{14}(t) = h_x \dot{e}_1^T(t)Z_1\dot{e}_1(t) - \int_{t-h_x}^t \dot{e}_1^T(\tau)Z_1\dot{e}_1(\tau)d\tau, \tag{38}$$

$$\dot{V}_{15}(t) = h \dot{e}_1^T(t)S_1\dot{e}_1(t) - \int_{t-h}^t \dot{e}_1^T(\tau)S_1\dot{e}_1(\tau)d\tau, \tag{39}$$

and for arbitrary constant $\gamma > 0$, we have

$$-2\gamma \dot{e}_1^T(t)P_1[\dot{e}_1(t) - \mathcal{A}(\Psi(t))e_1(t) - \mathcal{B}(\Psi_{h_x}(t))e_1(t - h_x) - P_1^{-1}R_1Ce_1(t - h)] = 0. \tag{40}$$

Combining (35)–(40), we obtain

$$\dot{V}_1(t) \leq \zeta_1^T(t)\Omega_1\zeta_1(t), \tag{41}$$

where $\xi_1(t) = [e_1^T(t), e_1^T(t - h_x), e_1^T(t - h), \dot{e}_1^T(t)]^T$ and

$$\Omega_1 = \begin{bmatrix} \Xi_1 & -T_1 + P_1\mathcal{B}(\Psi_{h_x}(t)) & -Y_1 + R_1C & \gamma\mathcal{A}(\Psi(t))P_1 \\ * & -Q_1 & 0 & \gamma\mathcal{B}^T(\Psi_{h_x}(t))P_1 \\ * & * & -M_1 & \gamma C^T R_1^T \\ * & * & * & -2\gamma P_1 + h_x Z_1 + h S_1 \end{bmatrix}.$$

Due to $\forall \Psi \in \mathcal{V}_{\mathcal{H}_n}, \forall \Psi_{h_x} \in \mathcal{V}_{\mathcal{H}_n}$, according to the convex principle, one obtains

$$\Omega_1 < 0, \forall \Psi, \Psi_{h_x} \in \mathcal{V}_{\mathcal{H}_n}, \tag{42}$$

which leads to $\dot{V}_1(t) < 0$. Then, it follows from (41) that

$$\dot{V}_1(t) \leq -\lambda_{\min}(-\Omega_1)\xi_1^T(t)\xi_1(t) \leq -\lambda_{\min}(-\Omega_1)e_1^T(t)e_1(t), \tag{43}$$

this indicates that the error system $e_1(t)$ is globally asymptotically stable. Obviously, conditions (25) ensure that $\Omega_1 < 0$ holds.

Step j: To recursively prove the stability of the error system $e_j(t)$, we assume that $e_{j-1}(t)$ is globally asymptotically stable. Similar to (28) and using the Newton–Leibniz formula, $e_j(t)$ ($j = 2, 3, \dots, m$) can be rewritten, as follows

$$\begin{aligned} \dot{e}_j(t) = & [\mathcal{A}(\Psi(t)) + \mathcal{B}(\Psi_{h_x}(t)) + L_j C]e_j(t) - \mathcal{B}(\Psi_{h_x}(t)) \int_{t-h_x}^t \dot{e}_j(\tau) d\tau - L_j C \int_{t-h}^t \dot{e}_j(\tau) d\tau \\ & - L_j C e_{j-1}(t). \end{aligned} \tag{44}$$

Then, we construct the following Lyapunov–Krasovskii functions

$$\begin{aligned} V_j(t) = & \overbrace{e_j^T(t)P_j e_j(t)}^{V_{j1}(t)} + \overbrace{\int_{t-h_x}^t e_j^T(\tau)Q_j e_j(\tau) d\tau}^{V_{j2}(t)} + \overbrace{\int_{t-h}^t e_j^T(\tau)M_j e_j(\tau) d\tau}^{V_{j3}(t)} \\ & + \overbrace{\int_{-h_x}^0 \int_{t+\eta}^t \dot{e}_j^T(\tau)Z_j \dot{e}_j(\tau) d\tau d\eta}^{V_{j4}(t)} + \overbrace{\int_{-h}^0 \int_{t+\eta}^t \dot{e}_j^T(\tau)S_j \dot{e}_j(\tau) d\tau d\eta}^{V_{j5}(t)}. \end{aligned} \tag{45}$$

Taking the derivative of $V_{j1}(t)$ along with (44), we have

$$\begin{aligned} \dot{V}_{j1}(t) = & e_j^T(t) [P_j \mathcal{A}(\Psi(t)) + \mathcal{A}^T(\Psi(t))P_j + P_j \mathcal{B}(\Psi_{h_x}(t)) + \mathcal{B}^T(\Psi_{h_x}(t))P_j + R_j C + C^T R_j] \times \\ & e_j(t) - 2e_j^T(t)P_j \mathcal{B}(\Psi_{h_x}(t)) \int_{t-h_x}^t \dot{e}_j(\tau) d\tau - 2e_j^T(t)R_j C \int_{t-h}^t \dot{e}_j(\tau) d\tau \\ & - e_j^T(t)R_j C e_{j-1}(t). \end{aligned} \tag{46}$$

Now, by using Young’s inequality [27], we obtain

$$-e_j^T(t)R_j C e_{j-1}(t) \leq \varepsilon_1 e_j^T(t)e_j(t) + \frac{1}{\varepsilon_1} \|R_j C\|^2 \|e_{j-1}\|^2, \tag{47}$$

with $\varepsilon_1 > 0$. Then, similar to (33)–(34) in step 1 and using (46)–(47), we have

$$\begin{aligned} \dot{V}_{j1} \leq & e_j^T(t) [P_j \mathcal{A}(\Psi(t)) + \mathcal{A}^T(\Psi(t))P_j + T_j^T + T_j + Y_j^T + Y_j + h_x D_j + h X_j + \varepsilon_1 I] e_j(t) \\ & - 2e_j^T(t) [T_j - P_j \mathcal{B}(\Psi_{h_x}(t))] e_j(t - h_x) - 2e_j^T(t) [Y_j - R_j C] e_j(t - h) \\ & + \int_{t-h_x}^t \dot{e}_j^T(\tau)Z_j \dot{e}_j(\tau) d\tau + \int_{t-h}^t \dot{e}_j^T(\tau)S_j \dot{e}_j(\tau) d\tau + \frac{1}{\varepsilon_1} \|R_j C\|^2 \|e_{j-1}\|^2. \end{aligned} \tag{48}$$

For arbitrary constants $\gamma > 0$ and $\varepsilon_2 > 0$, we obtain

$$\begin{cases} -2\gamma \dot{e}_j^T(t) P_j [\dot{e}_j(t) - \mathcal{A}(\Psi(t)) e_j(t) - \mathcal{B}(\Psi_{h_x}(t)) e_j(t - h_x) - P_j^{-1} R_j C e_j(t - h) \\ + P_j^{-1} R_j C e_{j-1}(t)] = 0, \\ -2\gamma e_j^T(t) R_j C e_{j-1}(t) \leq \varepsilon_2 e_j^T(t) e_j(t) + \frac{\gamma^2}{\varepsilon_2} \|R_j C\|^2 \|e_{j-1}\|^2. \end{cases} \tag{49}$$

Then, combining the derivative of $\sum_{i=2}^5 V_{ji}(t)$ and using (48)–(49), we have

$$\dot{V}_j(t) \leq \xi_j^T(t) \hat{\Omega}_j \xi_j(t) + \left(\frac{1}{\varepsilon_1} + \frac{\gamma^2}{\varepsilon_2}\right) \|R_j C\|^2 \|e_{j-1}\|^2, \tag{50}$$

where $\xi_j(t) = [e_j^T(t), e_j^T(t - h_x), e_j^T(t - h), \dot{e}_j^T(t)]^T$,

$$\hat{\Omega}_j = \begin{bmatrix} \hat{\Xi}_j & -T_j + P_j \mathcal{B}(\Psi_{h_x}(t)) & -Y_j + R_j C & \gamma \mathcal{A}(\Psi(t)) P_j \\ * & -Q_j & 0 & \gamma \mathcal{B}^T(\Psi_{h_x}(t)) P_j \\ * & * & -M_j & \gamma C^T R_j^T \\ * & * & * & -2\gamma P_j + h_x Z_j + h S_j + \varepsilon_2 I \end{bmatrix}$$

and $\hat{\Xi}_j = P_j \mathcal{A}(\Psi(t)) + \mathcal{A}^T(\Psi(t)) P_j + T_j + T_j^T + Y_j + Y_j^T + h_x D_j + h X_j + Q_j + M_j + \varepsilon_1 I$.
 Similar to Ω_1 in step 1, if

$$\hat{\Omega}_j < 0, \forall \Psi, \Psi_{h_x} \in \mathcal{V}_{\mathcal{H}_n} \tag{51}$$

is true, we have

$$\dot{V}_j(t) \leq -\lambda_{\min}(-\hat{\Omega}_j) e_j^T(t) e_j(t) + \left(\frac{1}{\varepsilon_1} + \frac{\gamma^2}{\varepsilon_2}\right) \|R_j C\|^2 \|e_{j-1}\|^2. \tag{52}$$

Then, employing the comparison Lemma [35], we can conclude that if $e_{j-1}(t)$ is globally asymptotically stable, then $e_j(t)$ is also globally asymptotically stable.

Furthermore, it not difficult to observe that $\hat{\Omega}_j = \Omega_j + \varepsilon_1 \Pi_1^T \Pi_1 + \varepsilon_2 \Pi_2^T \Pi_2$, where $\Pi_1 = [I, 0, 0, 0]^T$, $\Pi_2 = [0, 0, 0, I]^T$, and

$$\Omega_j = \begin{bmatrix} \Xi_j & -T_j + P_j \mathcal{B}(\Psi_{h_x}) & -Y_j + R_j C & \gamma \mathcal{A}(\Psi) P_j \\ * & -Q_j & 0 & \gamma \mathcal{B}^T(\Psi_{h_x}) P_j \\ * & * & -M_j & \gamma C^T R_j^T \\ * & * & * & -2\gamma P_j + h_x Z_j + h S_j \end{bmatrix} < 0.$$

Since ε_1 and ε_2 are selected arbitrarily, $\Omega_j < 0$ ensures that $\hat{\Omega}_j < 0$ holds when ε_1 and ε_2 are sufficiently small. Finally, it observes that conditions (25) ensure that $\Omega_j < 0$ holds. We finish the proof. \square

Remark 6. It follows from Theorem 2 that for a given output delay h_y , whether the LMI set (25), (26) and (27) have feasible solutions depends on the size of the parameter $h = \frac{h_y}{m}$. Obviously, for a large output delay of h_y , a sufficiently large m can ensure that the LMI sets (25), (26) and (27) have feasible solutions. However, the complexity of the cascade predictor is proportional to m ; in other words, a larger m will reduce the observational performance of the cascade predictor. Therefore, we should choose a suitable value of m to balance the stability requirement and the complexity limitation of the cascade predictor.

4. Numerical Simulation

This section provides a series of numerical simulations to demonstrate the efficacy of our results.

Let $x(t) = [x_1(t), x_2(t)]^T$. Consider the RNNs (1) with the following parameters:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, W_0 = \begin{bmatrix} 2.0 & -0.1 \\ -5.0 & 3.0 \end{bmatrix}, W_1 = \begin{bmatrix} -1.5 & -0.1 \\ -0.2 & -2.5 \end{bmatrix}, C = \begin{bmatrix} 3 & 0.3 \\ 0.3 & 3 \end{bmatrix},$$

$$h_x = 1, \phi(s) = [2, 2]^T, f(x(t)) = [\tanh(x_1(t)), \tanh(x_2(t))]^T,$$

where h_y exists if we take the value later. It can be observed from Figure 1 that these RNNs have complex chaotic behaviors. Then, from Theorem 1, it follows that there is a nonfeasible solution to the LMI set (12) when $h^* > 0.24$. Therefore, when $h_y \leq 0.24$, we only use the full-order observer (6). When $h_y > 0.24$, we use the cascade predictor (23).

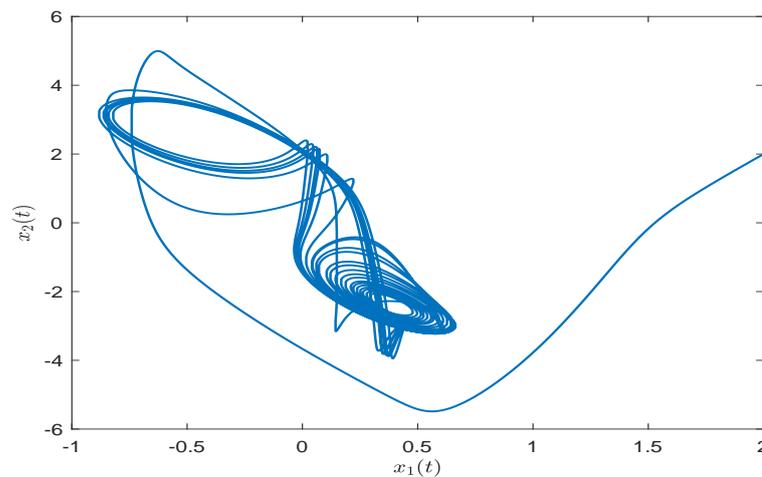


Figure 1. The phase space trajectory of RNNs (1).

Example 1. For $h_y = 0.23$, assume that the initial conditions of the full-order observer \hat{x} are $\hat{x}(s) = [-2, -2]^T, s \in [-1, 0)$. Then, from Theorem 1, we can obtain feasible solutions:

$$P = \begin{bmatrix} 15.9540 & 0.3889 \\ 0.3889 & 2.3080 \end{bmatrix}, L = \begin{bmatrix} -1.4728 & 0.1894 \\ 1.1641 & -2.3357 \end{bmatrix}.$$

The simulation results are shown in Figures 2 and 3. Figure 2 represents the trajectory profiles of $\hat{x}(t)$ and $x(t)$, which implies the validity of our designed full-order observer (6). Figure 3 represents the converged trajectory of $\|e(t)\| = \|\hat{x}(t) - x(t)\|$, which implies the convergence performance of the full-order observer (6). However, when $h = 0.25$, it can be observed from Figures 4 and 5 that the full-order observer cannot accurately estimate the state of the original system and the estimation error $\|e(t)\|$ becomes larger, since $h = 0.25$ is greater than $h^* = 0.24$.

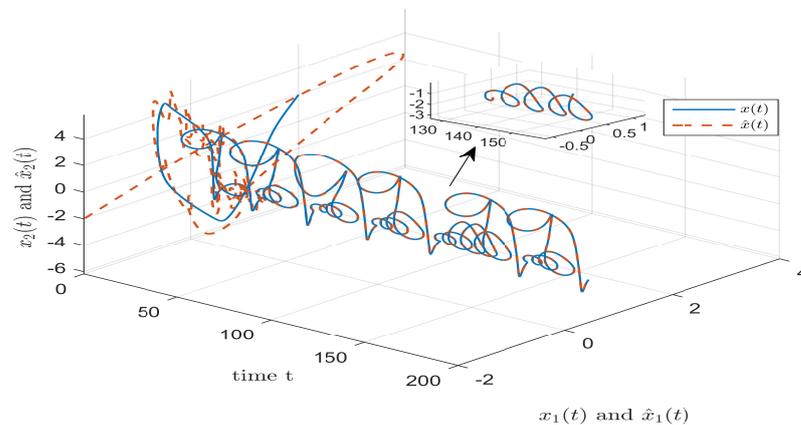


Figure 2. Profile on trajectories of $x(t)$ and $\hat{x}(t)$.

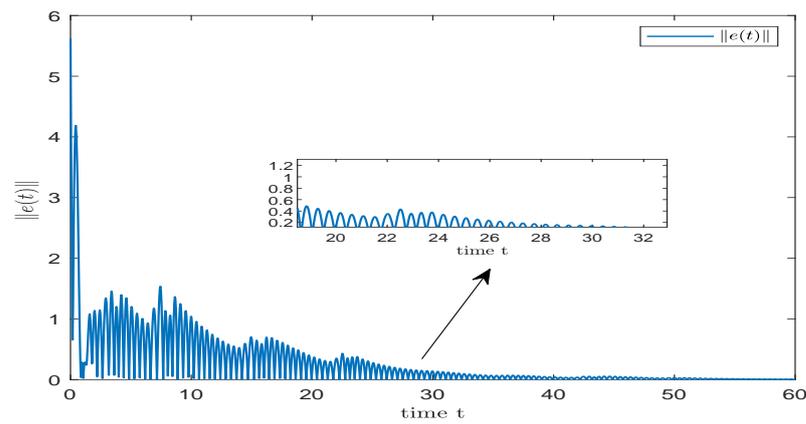


Figure 3. The estimation error $\|e(t)\|$.

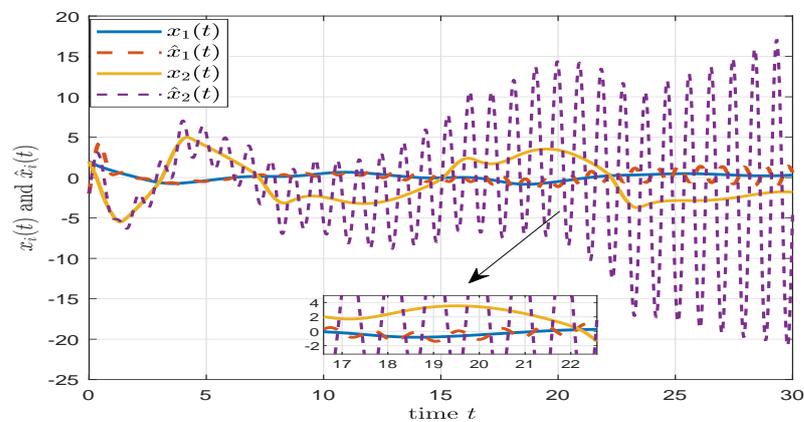


Figure 4. The state $x_i(t)$ and $\hat{x}_i(t)$, $i = 1, 2$.

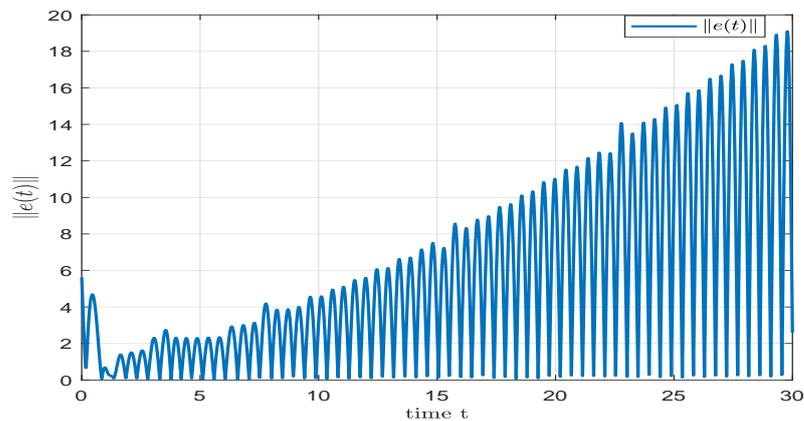


Figure 5. The estimation error $\|e(t)\|$.

Example 2. This example considers the case $h_y > 0.24$; thus, we only use the cascade predictor (23).

(i) For $h_y = 0.5$, we select $m = 5$ and $h = 0.1$ and assume that the initial conditions of \hat{x}_i , $i = 1, \dots, 5$ are $\hat{x}_i(s) = [-2, -2]^T$, $s \in [-1, 0)$. Then, from Theorem 2, we obtain feasible solutions:

$$P_i = \begin{bmatrix} 142.1455 & -3.8554 \\ -3.8554 & 16.5776 \end{bmatrix}, L_i = \begin{bmatrix} -1.5129 & 0.1483 \\ 1.5729 & -2.8071 \end{bmatrix} \quad (i = 1, \dots, 5).$$

As illustrated in Figures 6 and 7, the cascade predictor is valid and the estimation error $\|e_5(t)\| = \|\hat{x}_5(t) - x(t)\|$ finally converges to 0.

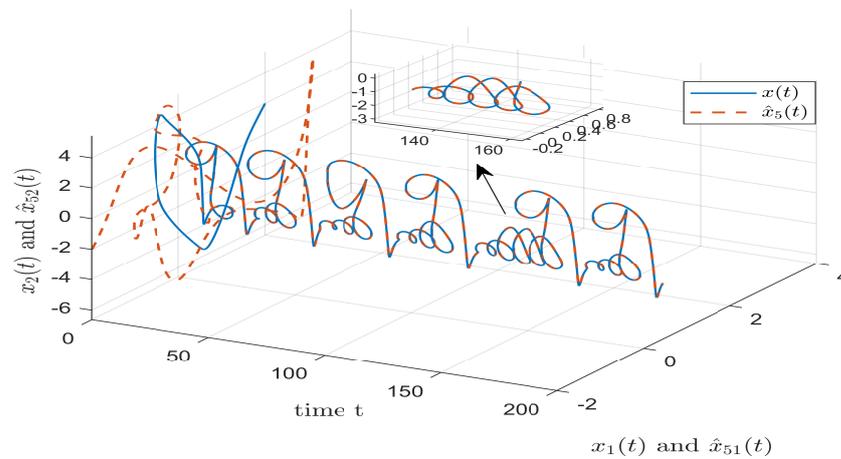


Figure 6. Profile on trajectories of $x(t)$ and $\hat{x}_5(t)$.

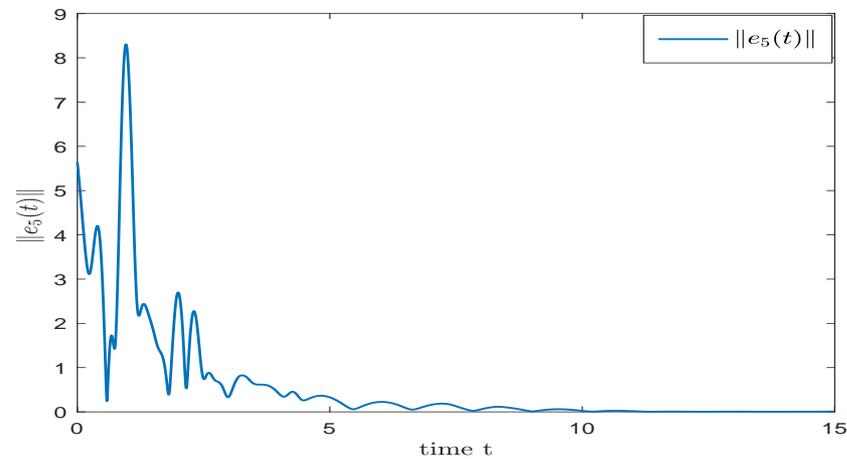


Figure 7. The estimation error $\|e_5(t)\|$.

(ii) For $h_y = 1$, we select $m = 10$ and $h = 0.1$ and assume that the initial conditions of \hat{x}_i , $i = 1, \dots, 10$ are $\hat{x}_i(s) = [-1, -1]^T$, $s \in [-1, 0)$. Since the value of h is equal to the value of h in (i), the observer gain L_i ($i = 1, \dots, 10$) can be equal to the observer gain in (i). Then, from Figures 8 and 9, it is clear that the cascade predictor is valid, and the observer error $\|e_{10}(t)\| = \|\hat{x}_{10}(t) - x(t)\|$ finally converges to 0.

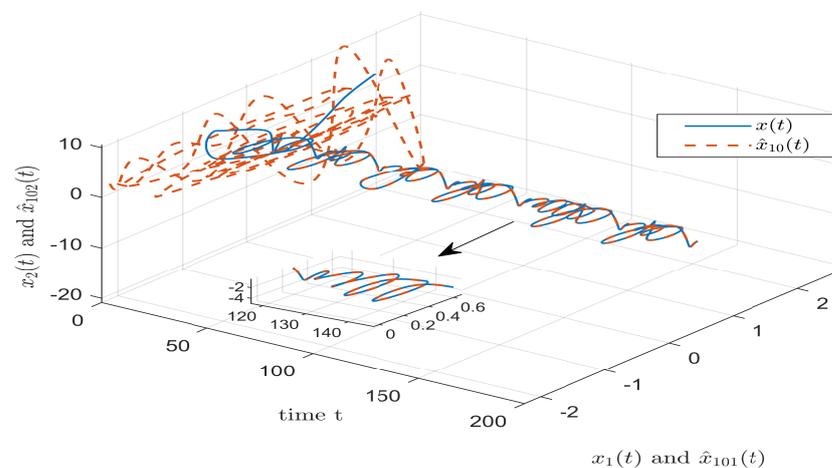


Figure 8. Profile on trajectories of $x(t)$ and $\hat{x}_{10}(t)$.

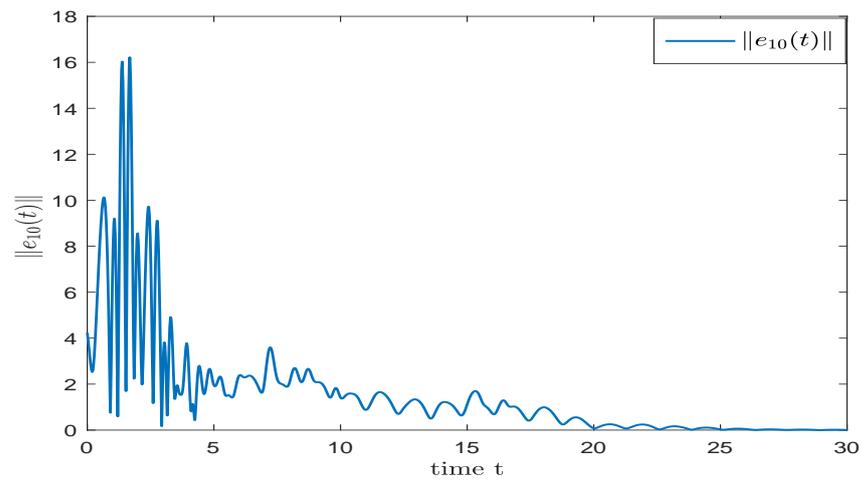


Figure 9. The estimation error $\|e_{10}(t)\|$.

Example 3. This example will further discuss the effect of the size of the output delay on the convergence of the two predictors. The simulation results are shown in Figures 10–12, and the influence of m and h on the convergence time is given in Table 1, where “*” denoting the single observer is not valid ($h_y > h^*$).

From Table 1, we can clearly observe that for both predictors, the output delay is directly proportional to the convergence time, and the larger the output delay, the longer the convergence time. In addition, from the experimental results, it can be concluded that, although the cascade predictor can solve the problem of an arbitrarily large output delay, as h_y increases, we will have to choose more subsystems to transmit the state information, which leads to the accumulation of error information and increases the cost of observation.

Table 1. The convergence time for two types of predictors by setting observer error $\|e\| = 0.1$.

Predictor \ h_y	$h_y = 0.1$	$h_y = 0.2$	$h_y = 0.5$	$h_y = 1$	$h_y = 1.5$	$h_y = 2$
Convergence time	Simple observer	1.6	8.5	*	*	*
	Cascade predictor	1.8 (m = 1)	3.6 (m = 2)	9.1 (m = 5)	16.3 (m = 10)	27.4 (m = 15)
				31.5 (m = 20)		

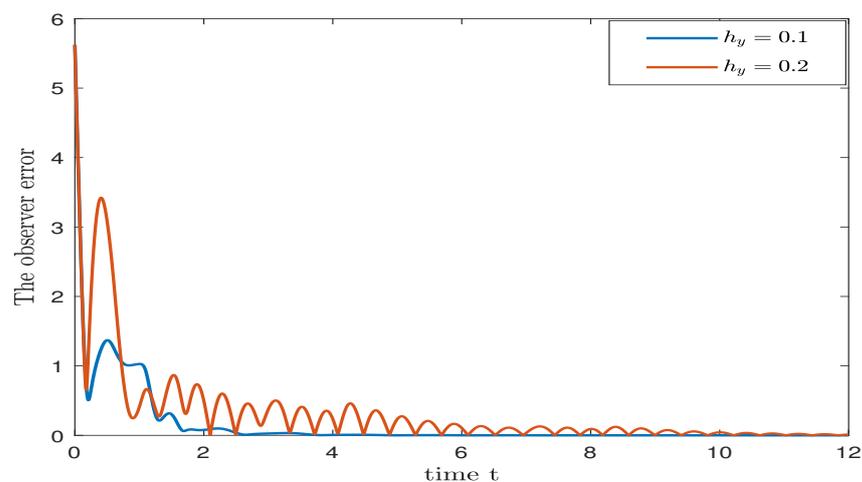


Figure 10. Convergence of single observer at different delays.

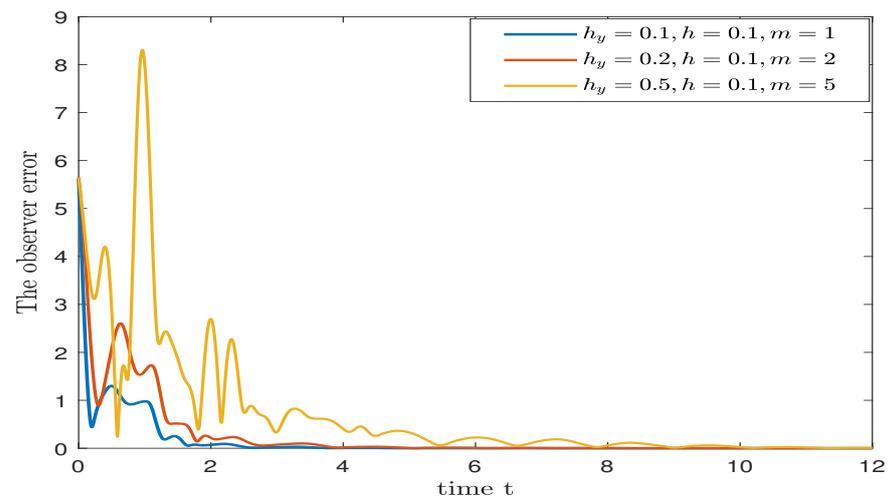


Figure 11. Convergence of cascade predictor at different delays.

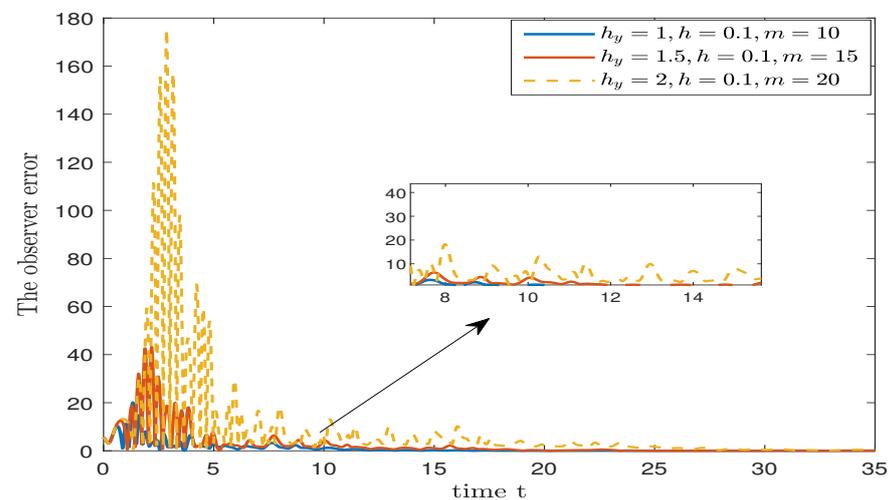


Figure 12. Convergence of cascade predictor at different delays.

5. Conclusions

In this research, we investigate the RNNs' state estimation by proposing an output-predicting and LPV approach. Due to the LPV approach, LKF and convex principle, several new conditions for the global asymptotic stability of the error system have been established. Compared with the traditional observer in [14–20], the chain-structured cascade predictor is more useful in the state estimation of neural networks. Different from [12,13,15,16,20], we use the LPV approach to convert nonlinear error dynamic systems into linear error systems, which greatly reduces the difficulty of the stability analysis. Finally, a series of numerical simulations show the effectiveness of the cascade predictor.

Author Contributions: Conceptualization, W.W. and Z.H.; methodology, W.W.; software, W.W. and J.C.; formal analysis, W.W.; writing—original draft preparation, W.W.; writing—review and editing, W.W. and Z.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61573005, and in part by the Natural Science Foundation of Fujian Province under Grant 2018J01417 and Grant 2019J01330.

Data Availability Statement: Not applicable.

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

References

1. Chua, L.; Yang, L. Cellular neural networks: Application. *IEEE Trans. Circuits Syst.* **1998**, *35*, 1273–1290. [[CrossRef](#)]
2. Cichocki, A.; Unbehauen, R. *Neural Networks for Optimization and Signal Processing*; Wiley: Chichester, UK, 1993.
3. Joya, G.; Atencia, M.; Sandoval, F. Hopfield neural networks for optimization: Study of the different dynamics. *Neurocomputing* **2002**, *43*, 219–237. [[CrossRef](#)]
4. Li, W.; Lee, T. Hopfield neural networks for affine invariant matching. *IEEE Trans. Neural Netw.* **2001**, *12*, 1400–1410. [[CrossRef](#)]
5. Yong, S.; Scott, P.; Nasrabadi, N. Object recognition using multilayer Hopfield neural network. *IEEE Trans. Image Process.* **1997**, *6*, 357–372. [[CrossRef](#)] [[PubMed](#)]
6. Wang, Z.; Liu, L.; Shan, Q.; Zhang, H. Stability criteria for recurrent neural networks with time-varying delay based on secondary delay partitioning method. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 2589–2595. [[CrossRef](#)] [[PubMed](#)]
7. Zhang, C.; He, Y.; Jiang, L.; Wu, M. Delay-dependent stability criteria for generalized neural networks with two delay components. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 1263–1276. [[CrossRef](#)]
8. Zhang, X.; Han, Q. Global asymptotic stability analysis for delayed neural networks using a matrix-based quadratic convex approach. *Neural Netw.* **2014**, *54*, 57–69. [[CrossRef](#)] [[PubMed](#)]
9. Wang, Z.; Zhang, H.; Jiang, B. LMI-based approach for global asymptotic stability analysis of recurrent neural networks with various delays and structures. *IEEE Trans. Neural Netw.* **2011**, *22*, 1032–1045. [[CrossRef](#)] [[PubMed](#)]
10. Liu, Y.J.; Lee, S.M.; Kwon, O.M.; Park, J.H. New approach to stability criteria for generalized neural networks with interval time-varying delays. *Neurocomputing* **2015**, *149*, 1544–1551. [[CrossRef](#)]
11. Wu, Z.; Shi, P.; Su, H.; Chu, J. Delay-dependent stability analysis for switched neural networks with time-varying delay. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2011**, *41*, 1522–1530. [[CrossRef](#)]
12. Zhu, Q.; Cao, J. Stability of Markovian jump neural networks with impulse control and time varying delays. *Nonlinear Anal. Real World Appl.* **2012**, *13*, 2259–2270. [[CrossRef](#)]
13. Wang, Z.; Ho, D.W.C.; Liu, X. State estimation for delayed neural networks. *IEEE Trans. Neural Netw.* **2005**, *16*, 279–284. [[CrossRef](#)] [[PubMed](#)]
14. Wang, Z.; Liu, R.; Liu, Y. State estimation for jumping recurrent neural networks with discrete and distributed delays. *Neural Netw.* **2009**, *22*, 41–48. [[CrossRef](#)] [[PubMed](#)]
15. Wang, Z.; Wang, J.; Wu, Y. State estimation for recurrent neural networks with unknown delays: A robust analysis approach. *Neurocomputing* **2017**, *227*, 29–36. [[CrossRef](#)]
16. Huang, H.; Feng, G.; Cao, J. Robust state estimation for uncertain neural networks with time-varying delay. *IEEE Trans. Neural Netw.* **2008**, *19*, 1329–1339. [[CrossRef](#)]
17. Huang, H.; Feng, G.; Cao, J. State estimation for static neural networks with time-varying delay. *Neural Netw.* **2010**, *23*, 1202–1207. [[CrossRef](#)]
18. Ren, J.; Zhu, H.; Zhong, S.; Ding, Y.; Shi, K. State estimation for neural networks with multiple time delays. *Neurocomputing* **2015**, *151*, 501–510. [[CrossRef](#)]
19. Liu, M.; Chen, H. H_∞ state estimation for discrete-time delayed systems of the neural network type with multiple missing measurements. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 2987–2998. [[CrossRef](#)]
20. Guo, M.; Zhu, S.; Liu, X. Observer-based state estimation for memristive neural networks with time-varying delay. *Knowl.-Based Syst.* **2022**, *246*, 108707. [[CrossRef](#)]
21. Beintema, G.I.; Schoukens, M.; Toth, R. Deep subspace encoders for nonlinear system identification. *Automatica* **2023**, *156*, 111210. [[CrossRef](#)]
22. Germani, A.; Manes, C.; Pepe, P. A new approach to state observation of nonlinear systems with delayed output. *IEEE Trans. Autom. Control* **2002**, *47*, 96–101. [[CrossRef](#)]
23. Ahmed-Ali, T.; Cherrier, E.; Lamnabhi-Lagarrigue, F. Cascade high gain predictors for a class of nonlinear systems. *IEEE Trans. Autom. Control* **2012**, *57*, 224–229. [[CrossRef](#)]
24. Farza, M.; M'Saad, M.; Menard, T.; Fall, M.L.; Gehan, O.; Pigeon, E. Simple cascade observer for a class of nonlinear systems with long output delays. *IEEE Trans. Autom. Control* **2015**, *60*, 3338–3343. [[CrossRef](#)]
25. Farza, M.; Hernandez-Gonzalez, O.; Menard, T.; Targui, B.; M'Saad, M.; Astorga-Zaragoza, C.M. Cascade observer design for a class of uncertain nonlinear systems with delayed outputs. *Automatica* **2018**, *89*, 125–134. [[CrossRef](#)]
26. Zemouche, A.; Boutayeb, M. On LMI conditions to design observers for Lipschitz nonlinear systems. *Automatica* **2013**, *49*, 585–591. [[CrossRef](#)]
27. Adil, A.; Hamaz, A.; N'Doye, I.; Zemouche, A.; Laleg-Kirati, T.M.; Bedouhene, F. On high-gain observer design for nonlinear systems with delayed output measurements. *Automatica* **2022**, *141*, 11281. [[CrossRef](#)]
28. Huang, H.; Huang, T.; Chen, X.; Qian, C. Exponential stabilization of delayed recurrent neural networks: A state estimation based approach. *Neural Netw.* **2013**, *48*, 153–157. [[CrossRef](#)] [[PubMed](#)]
29. Zhang, Z.; He, Y.; Zhang, C.; Wu, M. Exponential stabilization of neural networks with time-varying delay by periodically intermittent control. *Neurocomputing* **2016**, *207*, 469–475. [[CrossRef](#)]
30. Huang, H.; Feng, G. Delay-dependent h_∞ and generalized h_2 filtering for delayed neural networks. *IEEE Trans. Circuits Syst. Regul. Pap.* **2009**, *56*, 846–857. [[CrossRef](#)]

31. Gonzalez, A. Improved results on stability analysis of time-varying delay systems via delay partitioning method and Finsler's lemma. *J. Frankl. Inst.* **2022**, *359*, 7632–7649. [[CrossRef](#)]
32. Moon, Y.S.; Park, P.; Kwon, W. H.; Lee, Y.S. Delay-dependent robust stabilization of uncertain state-delayed systems. *Int. J. Control* **2001**, *74*, 1447–1455. [[CrossRef](#)]
33. Gu, K.L.; Kharitonov, V.; Chen, J. *Stability of Time-Delay Systems*; Springer: Berlin, Germany, 2003.
34. Body, S.; Ghaoui, L.E.; Feron, E.; Balakrishnan, V. *Linear Matrix Inequalities in System and Control Theory*; SIAM: Philadelphia, PA, USA, 1994.
35. Khalil, H.K. *Nonlinear Systems*; Prentice-Hall: Englewood Cliffs, NJ, USA, 2002.

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.