# Nonlinear Autoregression with Convergent Dynamics on Novel Computational Platforms

Jiayin Chen and Hendra I. Nurdin

*Abstract*—Nonlinear stochastic modeling is useful for describing complex engineering systems. Meanwhile, neuromorphic (brain-inspired) computing paradigms are developing to tackle tasks that are challenging and resource intensive on digital computers. An emerging scheme is reservoir computing which exploits nonlinear dynamical systems for temporal information processing. This paper introduces reservoir computers with output feedback as stationary and ergodic infinite-order nonlinear autoregressive models. We highlight the versatility of this approach by employing classical and quantum reservoir computers to model synthetic and real data sets, further exploring their potential for control applications.

*Index Terms*—Nonlinear stochastic modeling; Convergent dynamics; Reservoir computing; Quantum dynamical systems.

## I. INTRODUCTION

The on-going quest for modeling complex systems has motivated a fruitful development in nonlinear stochastic modeling [1], [2]. The threshold model has been applied for ecology and hydrology times series and the autoregressive conditional heteroscedastic model is useful for volatility clustering [1]. Well-known nonlinear system identification models include the Volterra series [3], neural networks [4], nonlinear autoregressive exogenous models [5] and block-oriented models [6].

To solve tasks that are costly on digital computers, neuromorphic computing imitates human learning with the energy efficiency of the human brain. An emerging neuromorphic scheme is reservoir computing (RC) [7], [8], which exploits nonlinear dynamics (the "reservoir") to process time-varying input signals. In this work, we establish a theoretical framework for using reservoir computers (also abbreviated as RCs) as *infinite-order nonlinear autoregressive models with exogenous inputs, or NARX($\infty$) models* for applications such as time series modelling and system identification. Such NARX($\infty$) can be also expressed as *infinite-order nonlinear moving average models with exogenous inputs, or NMAX($\infty$) models*. Exploiting this equivalence, we show that such NARX($\infty$) models are asymptotically stationary and ergodic in the sense of Birkhoff-Khinchin [9, Theorem 24.1].

The attractiveness of our approach is that any nonlinear reservoir dynamics with the convergence property can induce stationary and ergodic NARX($\infty$). The reservoir dynamics are often randomly chosen but fixed at the outset and do not require precise tuning, only a linear output function is optimized to approximate target outputs. Echo-state networks (ESNs), a pioneering software RC implementation, has demonstrated

remarkable performance in chaotic system modeling [10] and time series modeling [11]. Their training efficiency enables RCs for fast signal processing. An FPGA RC reached 160 MHz rate for chaotic dynamics prediction and a photonic RC classified speech at a million words per second. RCs are also used in edge computing to reduce computation and transmission overhead; see [7], [8] and references therein. Quantum reservoir computers (QRCs) have recently been proposed to harness nonlinear quantum dynamics [12], [13], [14]. The micro-second quantum dynamics and their low energy requirements make them of interest as RC hardware [15].

Central to our development is the convergence property [16], [17] of a dynamical system, also known as the echo-state property [18] in the RC literature. Roughly speaking, a convergent dynamic has a unique reference state solution defined and bounded both backwards and forwards in time. All other solutions asymptotically converge to the reference state solution, independent of their initial conditions.

References [19], [20] investigate RCs for forecasting, reconstruction and filtering under stationary inputs. Ref. [21] explores RC properties that make them effective in stochastic modeling tasks. Different from these previous works, we develop a general theory for realizing NARX($\infty$) models with RCs, taking into account the stability of the model, and conditions for the asymptotic stationarity and ergodicity. Ref. [19] also considers optimizing the reservoir parameters to maximize the RC's information processing capacity via a Taylor expansion. Although we do not consider reservoir design problem here, it will be an interesting future research theme continuing from this work.

To highlight the versatility of our proposal, we employ ESNs and QRCs to model data sets collected from diverse fields of interest. We cast parameter estimation for these RCs as convex optimization problems. Numerical experiments indicate that ESNs and QRCs with only a few tunable parameters are able to describe these data sets.

This paper is organized as follows. Sec. II introduces RCs and the convergence property. Sec. III presents NARX($\infty$) models realized by convergent RCs and establishes their stationarity and ergodicity. Sec. IV details parameter estimation for ESNs and QRCs as NARX($\infty$). Sec. V presents numerical experiments. A conclusion is presented in Sec. VI.

**Notations**: $\mathbb{R}$ ($\mathbb{Z}$) are reals (integers), $\mathbb{Z}_- = \{\ldots, -1, 0\}$ and $\mathbb{Z}_+ = \{0, 1, \ldots\}$. $x \in \mathbb{R}^n$ is an $n \times 1$ vector and $x^\top$ is its transpose. $(\mathbb{R}^n)^{\mathbb{Z}}$ is the set of infinite sequences, i.e., $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ with $u = \{u_k\}_{k \in \mathbb{Z}}$ and $u_k \in \mathbb{R}^n$. $(\mathbb{R}^n)^{\mathbb{Z}_-}$ is the set of left-infinite sequences. $P_n^{\mathbb{Z}_-} : (\mathbb{R}^n)^{\mathbb{Z}} \to (\mathbb{R}^n)^{\mathbb{Z}_-}$ is the

J. Chen and H. I. Nurdin are with the School of Electrical Engineering and Telecommunications, UNSW Australia, Sydney NSW 2052, Australia (email: jiayin.chen@unsw.edu.au, h.nurdin@unsw.edu.au)

projection. For any $\tau \in \mathbb{Z}$, $z_n^{-\tau}$ is the time shift operator, i.e., for any $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ and $k \in \mathbb{Z}$, $z_n^{-\tau}(u)|_k := u_{k-\tau}$ (we drop the subscript and superscript $n$ when $n = 1$). $\|\cdot\|$ is the Euclidean norm.

## II. RESERVOIR COMPUTING

We consider RCs described by reservoir dynamics (also called the activation function in machine learning) $f : \mathbb{R}^N \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^N$ and output function $h : \mathbb{R}^N \to \mathbb{R}$, for all $k \in \mathbb{Z}$

$$\begin{cases} x_{k+1} = f(x_k, u_k, e_k), \\ \hat{y}_k = h(x_k), \end{cases} \tag{1}$$

Here, $x_k \in \mathbb{R}^N$ is the state, $u_k \in \mathbb{R}^n$ is the input, $e_k \in \mathbb{R}$ models an external noise, and $\hat{y}_k \in \mathbb{R}$ is the RC output. The dynamics $f = f_\gamma$ is often parametrized by a parameter $\gamma \in \mathbb{R}^p$ chosen according to the task. Here, we consider an arbitrary but fixed $f$ by uniformly randomly choosing $\gamma$ and fixing it at the onset. Only $h$ is optimized to match the target outputs. More generally, $\gamma$ can be optimized according to some criterion, a problem known as the reservoir design problem [19], as alluded to previously.

### A. The convergence property

RCs described by (1) with the convergence property [16], [17] induce input-output maps, mapping from $u \in (\mathbb{R}^n)^{\mathbb{Z}}, e \in \mathbb{R}^{\mathbb{Z}}$ to $\hat{y} \in \mathbb{R}^{\mathbb{Z}}$. Let $\phi(k; k_0, \xi)$ be a solution to (1) parameterized by $u$ and $e$, starting at time $k_0$ with initial condition $x_{k_0} = \xi$. That is, for all $k \geq k_0$, $\phi(k + 1; k_0, \xi) = f(\phi(k; k_0, \xi), u_k, e_k)$ and $\phi(k_0; k_0, \xi) = \xi$. A function $\beta : [0, \infty) \times \mathbb{Z}_+ \to \mathbb{R}$ is $\mathcal{KL}$ if $\beta(0, \cdot) = 0$, continuous and strictly increasing in the first argument, non-increasing in the second argument with $\lim_{t \to \infty} \beta(s, t) = 0$ for all $s \in [0, \infty)$ [22].

*Definition 1 (Convergence property [16], [17]):* An RC described by (1) has the convergence property (or is convergent) if for any $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ and $e \in \mathbb{R}^{\mathbb{Z}}$,

(i) there exists a unique and bounded solution $x^* \in (\mathbb{R}^N)^{\mathbb{Z}}$ to (1) that satisfies $x_{k+1}^* = f(x_k^*, u_k, e_k)$ for all $k \in \mathbb{Z}$ and $\sup_{k \in \mathbb{Z}} \|x_k^*\| < \infty$;

(ii) there exists $\beta \in \mathcal{KL}$ (independent of $u, e$) such that, for any $k, k_0 \in \mathbb{Z}$ with $k \geq k_0$ and any $\xi \in \mathbb{R}^N$,

$$\|x_k^* - \phi(k; k_0, \xi)\| \leq \beta(\|x_{k_0}^* - \xi\|, k - k_0). \tag{2}$$

Note that the convergence property is a property of $f$. The unique and bounded solution $x^*$ in Definition 1 is the reference state solution (determined by $u, e$ and $f$). Equation (2) imposes that as $k_0 \to -\infty$, any solution $\phi(k; k_0, \xi)$ to (1) asymptotically converges to the reference state solution $x^*$, independent of its initial condition $\xi$.

The following theorem provides sufficient conditions to ensure that (1) is convergent, which will be employed in parameter estimation of NARX($\infty$) models realized by convergent feedback RC dynamics; see Sec. IV below.

*Theorem 2:* [17, Theorem 13] An RC with a compact state-space described by (1) is convergent if there exists some $P = P^\top$, $P > 0$ and some $\theta \in (0, 1)$ (independent of $u$ and $e$) such that for any $u_k \in \mathbb{R}^n$, $e_k \in \mathbb{R}$ and any $x_1, x_2 \in \mathbb{R}^N$,

$$\|f(x_1, u_k, e_k) - f(x_2, u_k, e_k)\|_P \leq \theta \|x_1 - x_2\|_P, \tag{3}$$

where $\|x\|_P := \sqrt{x^\top P x}$.

### B. Filters and functionals

If an RC governed by (1) is convergent, then it induces a unique time-invariant and causal filter $U_{f,h} : (\mathbb{R}^n)^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ such that when evaluated at any time $k \in \mathbb{Z}$, $\hat{y}_k = U_{f,h}(u, e)|_k := h(x_k^*)$, where $x^*$ is the reference state solution to (1) (see [18], [23]). There is a bijection between $U_{f,h}$ and its associated functional $F_{f,h} : (\mathbb{R}^n)^{\mathbb{Z}^-} \times \mathbb{R}^{\mathbb{Z}^-} \to \mathbb{R}$, defined as $F_{f,h}(u', e') := U_{f,h}(\tilde{u}', \tilde{e}')|_0$ [3]. Here $\tilde{u}', \tilde{e}'$ are arbitrary extensions of $u', e'$ to $(\mathbb{R}^n)^{\mathbb{Z}}, \mathbb{R}^{\mathbb{Z}}$. We can recover $U_{f,h}$ from $F_{f,h}$ via $U_{f,h}(u, e)|_k = F_{f,h}(P_n^{\mathbb{Z}^-} \circ z_n^{-k}(u), P^{\mathbb{Z}^-} \circ z^{-k}(e))$ for any $k \in \mathbb{Z}$. This bijection allows us to establish the measurability of $U_{f,h}$ by showing that $F_{f,h}$ is measurable in Lemma 5, and the Birkhoff-Khinchin ergodicity of $U_{f,h}$ by showing that $F_{f,h}$ is integrable in Lemma 6.

## III. NARX($\infty$) MODELS BY CONVERGENT DYNAMICS

We are interested in implementing NARX($\infty$) models using RCs with output-feedback (also see Fig. 1), for all $k \in \mathbb{Z}$

$$\begin{cases} x_{k+1} = g(x_k, u_k, y_k), \\ \hat{y}_k = h(x_k), \end{cases} \tag{4}$$

where $x_k \in \mathbb{R}^N$ is the state and $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ is the input. The target output $y_k \in \mathbb{R}$ is related to the one-step ahead prediction $\hat{y}_k$ via $y_k = \hat{y}_k + e_k$, where $e_k \in \mathbb{R}$ is a noise source for the model. Later on, we will consider $e$ and $u$ modeled by discrete-time stochastic processes. We also consider an equivalent representation of (4) given by (1) where $u$ and $e$ are viewed as external inputs and

$$f(x_k, u_k, e_k) := g(x_k, u_k, h(x_k) + e_k). \tag{5}$$

We say that (4) is convergent if and only if (1) is convergent. In system identification, $u$ is often designed to be persistently exciting [2] to excite all modes of the plant. In time series modeling, RCs make (one-step ahead) predictions $\hat{y}_k$ of $y_k$ based on a single sample path $y$ from the generating model, and are not driven by input $u$. That is, (4) becomes

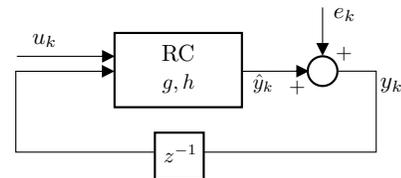$$\begin{cases} x_{k+1} = g(x_k, y_k), \\ \hat{y}_k = h(x_k). \end{cases}$$



Fig. 1. Schematic of RCs operating in an output-feedback configuration described by (4), where $z^{-1}$ is the one-step time delay operator.

This section introduces a probabilistic framework and shows that RCs described by (4) (under certain conditions on $f$ and $h$) implement NARX($\infty$) models defined later in (9). We then show that outputs of the NARX($\infty$) models are stationary and ergodic. The following result will be central.

*Theorem 3:* Consider a convergent RC described equivalently by (4) or (1). Let $U_{f,h}$ be the unique filter induced by (1). If $h$ is uniformly continuous, then for any $u \in (\mathbb{R}^n)^{\mathbb{Z}}$, $e \in \mathbb{R}^{\mathbb{Z}}$ and any $k \in \mathbb{Z}$,

$$
\begin{aligned}
y_k &= U_{f,h}(u,e)|_k + e_k \\
&= h \circ f(x_{k-1}, u_{k-1}, e_{k-1}) + e_k \\
&= h \circ f(f(x_{k-2}, u_{k-2}, e_{k-2}), u_{k-1}, e_{k-1}) + e_k \qquad (6) \\
&\;\;\vdots \\
&= \mathcal{F}(u_{k-1}, u_{k-2}, \ldots, e_{k-1}, e_{k-2}, \ldots) + e_k,
\end{aligned}
$$

where the following point-wise limit

$$
\begin{aligned}
U_{f,h}(u,e)|_k &= \mathcal{F}(u_{k-1}, u_{k-2}, \ldots, e_{k-1}, e_{k-2}, \ldots) \\
&:= \lim_{k_0 \to -\infty} h \circ f(\ldots f(f(\xi, u_{k_0}, e_{k_0}), u_{k_0+1}, e_{k_0+1}) \ldots) \qquad (7)
\end{aligned}
$$

exists and is independent of the initial condition $\xi \in \mathbb{R}^N$.

**Proof.** Equation (6) follows from (1). To show the point-wise limit (7) exists, fix $k \in \mathbb{Z}$ and $\xi \in \mathbb{R}^N$. For any $k_0 \in \mathbb{Z}$ with $k_0 \leq k$, consider a solution $\phi(k; k_0, \xi)$ to (1). Then (7) can be written as $U_{f,h}(u,e)|_k = \mathcal{F}(u_{k-1}, u_{k-2}, \ldots, e_{k-1}, e_{k-2}, \ldots) = \lim_{k_0 \to -\infty} h \circ \phi(k; k_0, \xi)$. We show that $\{h \circ \phi(k; k_0, \xi)\}_{k_0 \leq k}$ is a Cauchy sequence and thus (7) exists. Since $h$ is uniformly continuous, it suffices to show $\{\phi(k; k_0, \xi)\}_{k_0 \leq k}$ is Cauchy.

Let $k_0 \geq k_0'$ and $x^*$ be the reference state solution to (1). By (2) in Definition 1, we have

$$
\begin{aligned}
&\|\phi(k; k_0, \xi) - \phi(k; k_0', \xi)\| \\
&\leq \|x_k^* - \phi(k; k_0, \xi)\| + \|x_k^* - \phi(k; k_0', \xi)\| \\
&\leq \beta(\|x_{k_0}^* - \xi\|, k - k_0) + \beta(\|x_{k_0'}^* - \xi\|, k - k_0') \\
&\leq \beta(R_\xi, k - k_0) + \beta(R_\xi, k - k_0'),
\end{aligned}
$$

where $R_\xi = \max\{\|x_{k_0}^* - \xi\|, \|x_{k_0'}^* - \xi\|\} < \infty$. Since $\beta \in \mathcal{KL}$, for any $\epsilon > 0$ there exists $k_0^* \in \mathbb{Z}$ such that for all $k_0' \leq k_0 \leq k_0^*$, $\beta(R_\xi, k - k_0) + \beta(R_\xi, k - k_0') < \epsilon$. It follows that $\{\phi(k; k_0, \xi)\}_{k_0 \leq k}$ is Cauchy for any fixed $k$ and $\xi$.

To show that (7) is independent of $\xi$, let $\phi(k; k_0, \xi')$ be another solution to (1) starting at another initial condition $\xi' \neq \xi$. Mimicking the argument above gives

$$
\begin{aligned}
&\|\phi(k; k_0, \xi) - \phi(k; k_0, \xi')\| \\
&\leq \beta(\|x_{k_0}^* - \xi\|, k - k_0) + \beta(\|x_{k_0}^* - \xi'\|, k - k_0).
\end{aligned}
$$

The limit (7) is independent of $\xi$ now follows from taking $k_0 \to -\infty$ and uniform continuity of $h$. ∎

In Theorem 3, we have written $x_{k+1} = f(x_k, u_k, e_k)$ as in (1). Equivalently, we can write $x_{k+1} = g(x_k, u_k, y_k)$ as in (4), where $g$ and $f$ are related via (5). This leads to the following.

*Corollary 4:* Consider a convergent RC described equivalently by (4) or (1). If $h$ is uniformly continuous, then for any $u \in (\mathbb{R}^n)^{\mathbb{Z}}$, $e \in \mathbb{R}^{\mathbb{Z}}$ and any $k \in \mathbb{Z}$,

$$
y_k = \mathcal{G}(u_{k-1}, u_{k-2}, \ldots, y_{k-1}, y_{k-2}, \ldots) + e_k,
$$

where the following point-wise limit

$$
\begin{aligned}
&\mathcal{G}(u_{k-1}, u_{k-2}, \ldots, y_{k-1}, y_{k-2}, \ldots) \\
&:= \lim_{k_0 \to -\infty} h \circ g(\ldots g(g(\xi, u_{k_0}, y_{k_0}), u_{k_0+1}, y_{k_0+1}) \ldots) \qquad (8)
\end{aligned}
$$

exists and is independent of initial condition $\xi \in \mathbb{R}^N$.

## A. NARX($\infty$) models

We apply Corollary 4 to show that convergent RCs described by (4) or (1), such that $f$ is continuous and $h$ is uniformly continuous, implement NARX($\infty$) models defined in (9) below. By Theorem 3, such NARX($\infty$) models can also be written as NMAX($\infty$) models defined in (10) below.

Let $(\Omega, \Sigma, \mathbb{P})$ be a complete probability space on which all random variables are defined. We say that $\mathbf{z}$ is a stochastic process if $\mathbf{z} : (\Omega, \Sigma) \to ((\mathbb{R}^m)^{\mathbb{Z}}, (\mathcal{R}^m)^{\mathbb{Z}})$ is measurable, where $(\mathcal{R}^m)^{\mathbb{Z}}$ is the $\sigma$-algebra generated by cylindrical sets in $(\mathbb{R}^m)^{\mathbb{Z}}$; see e.g. [9, Sec. 36]. Further, for any $\omega \in \Omega$, $\mathbf{z}(\omega) = \{\mathbf{z}_k(\omega)\}_{k \in \mathbb{Z}}$ is a realization of $\mathbf{z}$.

Consider RCs described by (4) or (1) under $\mathbb{R}^n$-valued and $\mathbb{R}$-valued stochastic processes $\mathbf{u}$ and $\mathbf{e}$. The state $\mathbf{x}_k$ and the one-step ahead prediction $\hat{\mathbf{y}}_k$ are described by a stochastic RC:

$$
\begin{cases}
\mathbf{x}_{k+1} = g(\mathbf{x}_k, \mathbf{u}_k, \mathbf{y}_k), \\
\hat{\mathbf{y}}_k = h(\mathbf{x}_k),
\end{cases}
$$

where $\mathbf{y}_k = \hat{\mathbf{y}}_k + \mathbf{e}_k$. Stochasticity of $\mathbf{y}$ arises solely from the stochasticity of $\mathbf{u}$ and $\mathbf{e}$, and the maps $g, h$ are deterministic.

For a convergent RC described by (4) or (1) with uniformly continuous $h$, by Corollary 4, $\mathbf{y}_k$ is described by a NARX($\infty$) model such that for all $k \in \mathbb{Z}$,

$$
\mathbf{y}_k = \mathcal{G}(\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, \ldots, \mathbf{y}_{k-1}, \mathbf{y}_{k-2}, \ldots) + \mathbf{e}_k. \qquad (9)
$$

For each $\omega \in \Omega$, the point-wise limit (8) exists and is independent of initial condition $\xi \in \mathbb{R}^N$.

Equivalently, for a convergent RC described by (4) or (1) with uniformly continuous $h$, by Theorem 3, $\mathbf{y}_k$ is also described by a NMAX($\infty$) model, such that for all $k \in \mathbb{Z}$

$$
\begin{aligned}
\mathbf{y}_k &= U_{f,h}(\mathbf{u}, \mathbf{e})|_k + \mathbf{e}_k \\
&= \mathcal{F}(\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, \ldots, \mathbf{e}_{k-1}, \mathbf{e}_{k-2}, \ldots) + \mathbf{e}_k.
\end{aligned} \qquad (10)
$$

For each $\omega \in \Omega$, the point-wise limit (7) exists and is independent of initial condition $\xi \in \mathbb{R}^N$.

We now show that if in addition, $f$ defined by (5) is continuous, then the output $\mathbf{y}$ of a NARX($\infty$) model given by (9) is a well-defined stochastic process.

*Lemma 5:* Consider a convergent RC described equivalently by (4) or (1). Let $U_{f,h} : ((\mathbb{R}^n)^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}, (\mathcal{R}^n)^{\mathbb{Z}} \times \mathcal{R}^{\mathbb{Z}}) \to (\mathbb{R}^{\mathbb{Z}}, \mathcal{R}^{\mathbb{Z}})$ be the unique filter induced by (1). Suppose that $h : \mathbb{R}^N \to \mathbb{R}$ is uniformly continuous and $f : \mathbb{R}^N \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^N$ defined by (5) is continuous. Then $U_{f,h}$ is measurable and for any stochastic processes $\mathbf{u}$ and $\mathbf{e}$, the output $\mathbf{y}$ of the NARX($\infty$) model defined by (9) is a stochastic process.

**Proof.** Since $\mathbf{y}$ is also the output of the corresponding NMAX($\infty$) model (10), it follows that $\mathbf{y}$ is a stochastic process if $U_{f,h}$ is measurable. Recall the bijection between $U_{f,h}$ and its functional $F_{f,h}$. This bijection implies that $U_{f,h}$ is measurable if and only if $F_{f,h}$ is measurable [20, Sec. II]. Hence it suffices to show $F_{f,h} : ((\mathbb{R}^n)^{\mathbb{Z}_-} \times \mathbb{R}^{\mathbb{Z}_-}, (\mathcal{R}^n)^{\mathbb{Z}_-} \times \mathcal{R}^{\mathbb{Z}_-}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable. By Theorem 3, for any $u' \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ and $e' \in \mathbb{R}^{\mathbb{Z}_-}$, $F_{f,h}(u', e') = \lim_{k_0 \to -\infty} h \circ f(\ldots f(\xi, u_{k_0}', e_{k_0}') \ldots)$, where the limit exists and is independent of $\xi$. Fix a $\xi \in \mathbb{R}^N$, for any $k_0 \in \mathbb{Z}_-$, define $F_{f,h}^{k_0} : (\mathbb{R}^n)^{\mathbb{Z}_-} \times \mathbb{R}^{\mathbb{Z}_-} \to \mathbb{R}$ by

$$
F_{f,h}^{k_0}(u', e') := h \circ f(\ldots f(\xi, u_{k_0}', e_{k_0}') \ldots).
$$

Then $\lim_{k_o \to -\infty} F_{f,h}^{k_0}(u', e') = F_{f,h}(u', e')$ point-wise and $F_{f,h}$ is measurable if $F_{f,h}^{k_0}$ is measurable for all $k_0 = \{\ldots, -2, -1\}$ [9, Theorem 13.4]. To show this, we write

$$F_{f,h}^{k_0}(u', e') = h \circ \overline{f}_{k_0}(\mathcal{P}_n^{k_0}(u'), \mathcal{P}^{k_0}(e')),$$

where $\mathcal{P}_n^{k_0} := \prod_{j=k_0}^{-1} P_n^j : (\mathbb{R}^n)^{\mathbb{Z}_-} \to (\mathbb{R}^n)^{-k_0}$, $\mathcal{P}^{k_0} := \prod_{j=k_0}^{-1} P^j : \mathbb{R}^{\mathbb{Z}_-} \to \mathbb{R}^{-k_0}$. Here, $P_n^j(u') = u_j'$ and $P^j(e') = e_j'$. Furthermore, $\overline{f}_{k_0} : (\mathbb{R}^n)^{-k_0} \times \mathbb{R}^{-k_0} \to \mathbb{R}^N$ is given by

$$\overline{f}_{k_0}(\mathcal{P}_n^{k_0}(u'), \mathcal{P}^{k_0}(e')) = f(\ldots f(\xi, u_{k_0}', e_{k_0}') \ldots).$$

Since $h, \mathcal{P}_n^{k_0}$ and $\mathcal{P}^{k_0}$ are measurable, it remains to show that $\overline{f}_{k_0}$ is measurable. To this end, to simplify notation, for any $i, j \in \mathbb{Z}_+$, let $u_{k_0-j:-1-i}'$ and $e_{k_0-j:-1-i}'$ denote the concatenation of $\{u_{k_0-j}', \ldots, u_{-1-i}'\}$ and $\{e_{k_0-j}', \ldots, e_{-1-i}'\}$ into a column vector, respectively. We can define $\overline{f}_{k_0}$ recursively via $\overline{f}_{k_0-1}(u_{k_0-1:-1}', e_{k_0-1:-1}') = f(\overline{f}_{k_0}(u_{k_0-1:-2}', e_{k_0-1:-2}'), u_{-1}', e_{-1}')$. Using this recursion and continuity of $f$, an inductive argument on $k_0$ shows that $\overline{f}_{k_0}$ is continuous. Hence $F_{f,h}$ and $U_{f,h}$ are measurable, and $\mathbf{y}$ defined by (9) is a stochastic process. $\blacksquare$

### B. Stationarity and ergodicity

In this section, we derive conditions under which $\mathbf{y}$ defined by (9) is stationary and/or ergodic given that $\mathbf{u}, \mathbf{e}$ are stationary [1, Definition 2.2] and/or ergodic [9, Sec. 24&36]. We equip $\mathbb{R}^m$ with the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^m)$. An $\mathbb{R}$-valued process $\tilde{\mathbf{z}}$ is Birkhoff-Khinchin ergodic if $\lim_{L \to \infty} \frac{1}{L} \sum_{k=0}^{L-1} \tilde{\mathbf{z}}_k = \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}_0]$ almost surely (a.s.), where $\mathbb{E}_{\mathbb{P}}[\cdot]$ is the expectation over $\mathbb{P}$.

To establish statistical properties of the output of a NARX($\infty$) model defined by (9), we again exploit the bijection with its associated NMAX($\infty$) model given by (10).

*Lemma 6:* Consider a convergent RC described equivalently by (4) or (1). Under the assumptions of Lemma 5, the process $\mathbf{y}$ defined by (9) is stationary (resp. ergodic) if $\mathbf{u}$ and $\mathbf{e}$ are stationary (resp. ergodic). Furthermore, suppose that $h \circ \mathbf{x}_0^*$ and $\mathbf{e}_0$ are integrable, where $\mathbf{x}^*$ is the reference state solution to (1). Then under the assumptions of Lemma 5, $\mathbf{y}$ is Birkhoff-Khinchin ergodic if $\mathbf{u}, \mathbf{e}$ are stationary and ergodic.

**Proof.** By (10), we have $\mathbf{y}_k = U_{f,h}(\mathbf{u}, \mathbf{e})|_k + \mathbf{e}_k$, where $U_{f,h}$ is the unique filter induced by (1). By Lemma 5, $U_{f,h}$ is measurable. Now $\mathbf{y}$ is stationary (resp. ergodic), given that $\mathbf{u}, \mathbf{e}$ are stationary (resp. ergodic), follows from [9, Theorem 36.4]. To show the second part of the Lemma, by discussions on [9, p. 526] and the Birkhoff-Khinchin ergodic theorem [9, Theorem 24.1], it suffices to show that $F_{f,h}(P_n^{\mathbb{Z}_-}(\mathbf{u}), P^{\mathbb{Z}_-}(\mathbf{e})) + \mathbf{e}_0 : (\Omega, \Sigma, \mathbb{P}) \to \mathbb{R}$ is integrable, where $F_{f,h}$ is the functional induced by (1). Recall that

$$\int_\Omega \left| F_{f,h}(P_n^{\mathbb{Z}_-}(\mathbf{u}(\omega)), P^{\mathbb{Z}_-}(\mathbf{e}(\omega))) + \mathbf{e}_0(\omega) \right| \mathbb{P}(d\omega)$$
$$\leq \int_\Omega \left( |h(\mathbf{x}_0^*(\omega))| + |\mathbf{e}_0(\omega)| \right) \mathbb{P}(d\omega).$$

Integrability of $F_{f,h}(P_n^{\mathbb{Z}_-}(\mathbf{u}), P^{\mathbb{Z}_-}(\mathbf{e})) + \mathbf{e}_0$ now follows from integrability of $h \circ \mathbf{x}_0^*$ and $\mathbf{e}_0$. $\blacksquare$

*Remark 7:* When $h$ is uniformly continuous (as required in Theorem 3), a sufficient condition for integrability of $h \circ \mathbf{x}_0^*$ is that there exists $0 < M < \infty$ such that $\|\mathbf{x}_0^*\| \leq M$ a.s..

*Remark 8:* A similar argument as in the proof of Lemma 6 shows that under the assumptions of Lemma 6, the one-step ahead prediction $\hat{\mathbf{y}}$ and $\hat{\mathbf{y}}^2$ are Birkhoff-Khinchin ergodic if $\mathbf{u}, \mathbf{e}$ are stationary and ergodic.

We conclude this section by imposing the following standard assumptions on $\mathbf{u}$ and $\mathbf{e}$:

*Assumption 9:* $\mathbf{u}$ and $\mathbf{e}$ are independent, and $\mathbf{e}$ is identically and independently distributed (iid). This implies that $\mathbf{e}$ is stationary and ergodic [9, Sec. 36].

*Remark 10:* Assumption 9 lays the basis for analyzing the RC residual $\hat{\mathbf{e}}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k$; see [1] for further detail. If the RC prediction $\hat{\mathbf{y}}$ describes the target data $\mathbf{y}$ adequately, $\hat{\mathbf{e}}$ should be a proxy for $\mathbf{e}$. To test this, we test if $\hat{\mathbf{e}}$ are uncorrelated using sample autocorrelation (ACF) via the "ggAcf" R command [24], and test if $\mathbf{e}$ is Gaussian using the Lilliefors test [25] and Q-Q plot (via the "lillietest" and the "qqplot" Matlab commands). In system identification, we further test if $\hat{\mathbf{e}}$ is independent of $\mathbf{u}$ based on their sample cross-correlation via the "ggCcf" R command [24].

## IV. PARAMETER ESTIMATION

This section presents parameter estimation of ESNs and QRCs as NARX($\infty$) models. These RCs are described by (4) or (1), where $f$ and $h$ satisfy the assumptions in Lemma 6. Further, Remark 7 holds since ESNs and QRCs admit a compact state-space. By Lemma 6 and Remark 8, if these RCs are convergent, then they realize NARX($\infty$) models whose output $\mathbf{y}$ and one-step ahead prediction $\hat{\mathbf{y}}$ are stationary and ergodic. Here, $\hat{\mathbf{y}}_k = W^\top \overline{h}(\mathbf{x}_k) + W_c$, where $W \in \mathbb{R}^N, W_c \in \mathbb{R}$ are output parameters and $\overline{h}$ is not parametrized.

Given time series data $\mathbf{y}_k, \mathbf{u}_k$ for $k = 0, \ldots, L$, the first $L_1$ data are for washing out the effect of the RC's initial condition, the next $L_t = L_2 - L_1$ data are for training and the remaining data are for validation. We apply Theorem 2 to (1) to ensure the RC's convergence, the resulting optimization of $W, W_c$ becomes a convex constrained least squares problem, $\min_{W,W_c} \frac{1}{L_t} \sum_{k=L_1+1}^{L_2} |\mathbf{y}_k - \hat{\mathbf{y}}_k|^2$, subject to $G(W) \leq 0$. Note that if $\mathbf{y}^2$ and $\mathbf{y}\hat{\mathbf{y}}$ are Birkhoff-Khinchin ergodic, then parameter estimation is consistent, i.e., as $L_t \to \infty$, the above cost function becomes $\min_{W,W_c} \mathbb{E}_{\mathbb{P}}[|\mathbf{y}_0 - \hat{\mathbf{y}}_0|^2]$ a.s..

### A. Echo-state networks (ESNs)

An ESN with state $x_k \in \mathbb{R}^N$ is governed by

$$\begin{cases} x_{k+1} = \tanh(Ax_k + Bu_k + Cy_k), \\ \hat{y}_k = W^\top x_k + W_c \end{cases} \quad (11)$$

where $\tanh(\cdot)$ applies to a vector elementwise. Elements of $A, B, C$ are drawn independently and uniformly from $[-1, 1]$ (for time series modeling, we set $C = 0$). To apply Theorem 2, we re-express (11) in the form of (1) by substituting $y_k = \hat{y}_k + e_k = W^\top x_k + W_c + e_k$ into (11):

$$x_{k+1} = f_{\text{ESN}}(x_k, u_k, e_k)$$
$$:= \tanh((A + CW^\top)x_k + C(W_c + e_k) + Bu_k).$$

Since ESNs admit a compact state-space, by Theorem 2, an ESN is convergent if for any $x_1, x_2 \in \mathbb{R}^N$ and any $k \in \mathbb{Z}$,

$$\|f_{\mathrm{ESN}}(x_1, u_k, e_k) - f_{\mathrm{ESN}}(x_2, u_k, e_k)\|_P$$
$$\leq \sigma_{\mathrm{m}}(A + CW^\top)\|x_1 - x_2\|_P < \|x_1 - x_2\|_P,$$

where $\sigma_{\mathrm{m}}(\cdot)$ is the maximum singular value. That is, an ESN is convergent if $\sigma_{\mathrm{m}}(A + CW^\top) < 1$. We optimize $W, W_c$ for the ESN using YALMIP [26] by passing the non-strict inequality constraint:

$$\begin{bmatrix} I & A + CW^\top \\ (A + CW^\top)^\top & I \end{bmatrix} \geq 10^{-3} I. \tag{12}$$

### B. Quantum reservoir computers (QRCs)

Consider an $N$-qubit QRC proposed in [14] described by

$$\begin{cases} \rho_{k+1} = (1 - \epsilon)T(u_k, y_k)\rho_k + \epsilon\rho_*, \\ \hat{y}_k = \sum_{i=1}^N W_i \mathrm{Tr}(Z^{(i)}\rho_k) + W_c, \end{cases} \tag{13}$$

where $\rho_k$ is a $2^N \times 2^N$ density operator (a positive semidefinite Hermitian matrix with trace one), $\rho_*$ is a fixed density operator whose $(1,1)$-th element is one and all other elements are zero, $Z^{(i)}$ is the Pauli-$Z$ operator on qubit $i$ and $\epsilon = 0.9$. The dynamics $T(u_k, y_k)$ is a completely positive trace-preserving (CPTP) map determined by $u_k$ and $y_k = \hat{y}_k + e_k$. A CPTP map sends a density operator to another density operator [14]. We consider

$$T(u_k, y_k) = \frac{1}{n+1}\left( \sum_{j=1}^n g(u_k^{(j)})T_j + g(y_k)T_{n+1} \right.$$
$$\left. + \left( n + 1 - \sum_{j=1}^n g(u_k^{(j)}) - g(y_k) \right)T_{n+2} \right),$$

where $u_k^{(j)}$ is the $j$-th component of $u_k \in \mathbb{R}^n$. We choose $g(s) = \frac{1}{1+\exp(-s)}$ with a globally Lipschitz constant $L_g = 1/4$. Generally, different $g$'s can be applied to $u_k^{(j)}$ and $y_k$ (for time series modeling, we set $g(y_k) = 0$). The CPTP maps $T_l$ for $l = 1, \ldots n + 2$ are governed by arbitrary but fixed unitary matrices $U_l$ such that $T_l(\rho_k) = U_l \rho_k U_l^\dagger$, where $\dagger$ denotes the adjoint; see [27] for the details. Such QRCs can be implemented on current quantum machines [14].

Let $\|A\|_1 := \mathrm{Tr}(\sqrt{A^\dagger A})$ for any matrix $A$. For any CPTP map $T$, let $\|T\|_{1-1} := \sup_{\|A\|_1 = 1}\|T(A)\|_1$. Since QRCs admit a compact state-space, by (3) in Theorem 2 (Theorem 2 also applies to $\|\cdot\|_1$), a QRC is convergent if for any density operators $\rho_j (j = 1, 2)$, any $k \in \mathbb{Z}$, $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ and $e \in \mathbb{R}^{\mathbb{Z}}$,

$$\|(1 - \epsilon)T(u_k, y_{k,1})\rho_1 + \epsilon\rho_* - ((1 - \epsilon)T(u_k, y_{k,2})\rho_2 + \epsilon\rho_*)\|_1$$
$$\leq \left( (1 - \epsilon) + \frac{2L_g(1 - \epsilon)}{n+1}\sum_{i=1}^N |W_i| \right)\|\rho_1 - \rho_2\|_1$$
$$\leq \theta\|\rho_1 - \rho_2\|_1,$$

where the first inequality follows from $\|T_{n+1} - T_{n+2}\|_{1-1} \leq 2$ [14] and $|g(y_1) - g(y_2)| \leq L_g|y_1 - y_2| \leq L_g \sum_{i=1}^N |W_i|\|\rho_1 - \rho_2\|_1$. Hence, a QRC described by (13) is convergent if there exists some $\theta \in (0, 1)$ such that

$$\sum_{i=1}^N |W_i| - \frac{\theta + \epsilon - 1}{1 - \epsilon}\frac{n+1}{2L_g} \leq 0. \tag{14}$$

Throughout, we set $\theta = 0.99$. We optimize $W$ using the "fmincon" command in `Matlab`.

We remark that (12) and (14) are only sufficient. Nevertheless, numerical experiments suggest that ESNs and QRCs under these constraints are adequate in describing the data. It is a future direction to find relaxation of these constraints.

## V. NUMERICAL EXAMPLES

We employ convergent ESNs and QRCs to model time series. For each time series, we randomly generate 50 ESNs with $N = 2, \ldots, 10$, and 50 QRCs with qubit number $N = 2, \ldots, 5$. The fitted ESNs and QRCs are selected via final prediction error (FPE) criterion using validation data defined as $\mathrm{FPE} := \left( \frac{L_v + N + 1}{L_v - N + 1} \right)\mathrm{MSE}$ and $\mathrm{MSE} := \frac{1}{L_v}\sum_{k=L_2+1}^L \hat{\mathbf{e}}_k^2$, where $L_v = L - L_2$ and $\mathbf{e}_k$ is the residual. We remark that FPE assumes $W, W_c$ are unbiased [28]. Although this cannot be ensured in general, the decrease in MSE is small as $N$ increases and FPE prefers the lowest order $N = 2$. It is a future research theme to develop model selection methods for the proposed scheme. We employ the root mean-squared error (RMSE) to compare among different models, where $\mathrm{RMSE} := \sqrt{\mathrm{MSE}}$; see [29] for further discussions. We report the selected RC's RMSE and the average RMSE of 50 randomly generated RCs with the same dimension $N$ (or qubit number) as the selected RC.

We estimate $\mathrm{Tr}(Z^{(i)}\rho_j)$ by averaging $M_m$ measurements on quantum machines, whose variance decreases as $1/M_m$ [14]. Here, we assume that the sampling error is negligible by taking $M_m$ sufficiently large. To investigate the effect of decoherence when QRCs are implemented on current quantum machines, we simulate the selected QRCs under dephasing and generalized amplitude damping (GAD) channels as in [30].

For all detailed numerical settings, see [27].

### A. Nonlinear quantum optics

This example demonstrates that RCs can act as nonlinear Wiener filters to extract the signal component of a highly noisy time series from nonlinear quantum optics. We consider a low-photon-number Kerr cavity with two input-output ports (Fig. 2(a) or [31, Fig. 2(a)]), whose internal mode is governed by a Hamiltonian $H_0 = \Delta a^\dagger a + \chi(a^\dagger)^2 a^2$, where $a$ is the annihilation operator, $\Delta = 100$ is the detuning from a reference frequency, and $\chi = -5$ governs the nonlinearity. The cavity is coupled to two incoming fields $\alpha_{\mathrm{in}j}(t)$ via operators $L_j = -\iota\sqrt{\kappa_j}a$, where $\iota = \sqrt{-1}$ and $\kappa_j = 25$ for $j = 1, 2$. Here, $\alpha_{\mathrm{in}1}(t)$ is a coherent field with a constant amplitude $\eta = 21.5$ and $\alpha_{\mathrm{in}2}(t)$ is in the vacuum state.

We obtain a discretized trajectory $\beta_k = \mathrm{Tr}((L_2 + L_2^\dagger)\rho_k) + \eta_k$ from homodyne measurement on the second output port with sampling time $10^{-3}$ on Qutip [32]. Here, $\rho_k$ is the state at time $k$ and $\eta_k$ is a quantum Gaussian white noise. Our
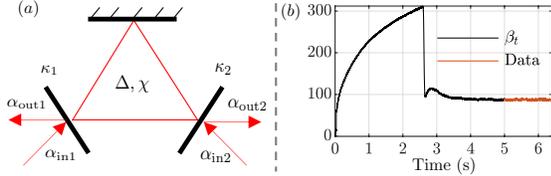
Fig. 2. (a) Kerr-nonlinear cavity with two input-output ports. The top mirror is fully reflective (without loss) and the other mirrors are partially transmitting. (b) The simulated trajectory of $\beta_k$ and the data employed.
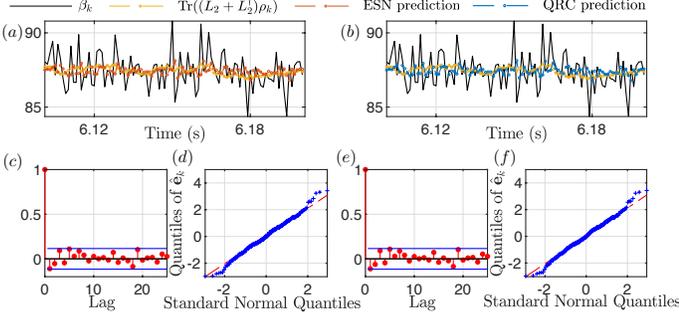


Fig. 3. The noisy time series $\beta_k$, the signal $\alpha_k = \text{Tr}((L_2 + L_2^\dagger)\rho_k)$, (a) the ESN prediction and (b) the QRC prediction on the first 100 validation data. (c) The ESN residual sample ACF (horizontal blue lines show the 95% CI). (d) The ESN residual Q-Q plot. (e) The QRC residual sample ACF. (e) The QRC residual Q-Q plot.

TABLE I
DIMENSION (OR QUBIT NUMBER) $N$, $p$-VALUE, RMSE AND AVERAGE RMSE ($\overline{\text{RMSE}}$) OF RCs FOR STOCHASTIC MODELING.

|  | Optics | | Meteorology | | Coupled electric drive | |
|---|---|---|---|---|---|---|
|  | ESN | QRC | ESN | QRC | ESN | QRC |
| $N$ | 2 | 2 | 2 | 2 | 2 | 2 |
| $p$-value | 0.5 | 0.5 | 0.5 | 0.23 | 0.5 | 0.44 |
| RMSE | 1.12 | 1.12 | 0.047 | 0.049 | 0.10 | 0.11 |
| $\overline{\text{RMSE}}$ | 1.12 | 1.12 | 0.055 | 0.058 | 0.12 | 0.14 |

goal is to employ RCs to separate the "signal part" $\alpha_k = \text{Tr}((L_2 + L_2^\dagger)\rho_k)$ from the highly noisy $\beta_k$. We simulate $\beta_k$ starting in the vacuum state on a truncated Hilbert space of dimension 1000 for 6.4s and employ the data after 5s until $\rho_k$ reaches a steady state, see Fig. 2(b).

We set $L_w = 99$, $L_t = 1000$ and $L_v = 300$. ESN and QRC with $N = 2$ achieve comparable RMSE and average RMSE, their residual sample ACFs show no autocorrelation within 95% CI and pass the Lilliefors test; see Table I and Fig. 3. We report 0.5 for p-values$\geq$ 0.5. In Fig. 3(a)(b), we observe that the ESN and QRC predictions follow the signal $\alpha_k$ with a normalized root mean-squared error between them of $\sqrt{\frac{\sum_{k=L_2+1}^{L}(\alpha_k - \hat{\mathbf{y}}_k)^2}{\sum_{k=L_2+1}^{L}\alpha_k^2}} = 0.0047$ and 0.0048, respectively. These suggest that QRC can act as a nonlinear Wiener filter, separating the signal $\alpha_k$ from the noisy time series $\beta_k$.

For all decoherence strengths, QRC obtains RMSE = 1.12 as the noiseless QRC, and its residuals show no autocorrelation and pass the Lilliefors test with $p$-value$\geq$ 0.5. QRC under decoherence can still effectively extract $\alpha_k$ from $\beta_k$.
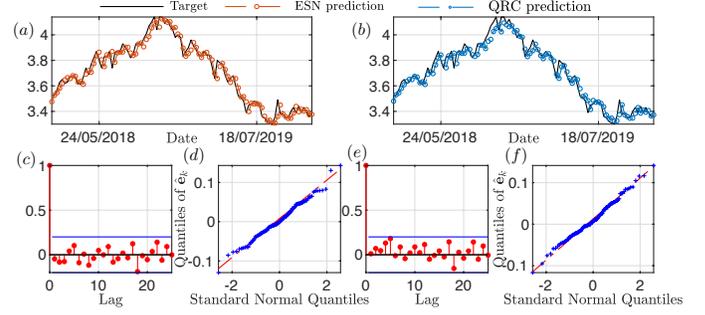


Fig. 4. The finance time series, (a) the ESN prediction and (b) the QRC prediction on validation data. (c) The ESN residual sample ACF. (d) The ESN residual Q-Q plot. (e) The QRC residual sample ACF. (f) The QRC residual Q-Q plot.

### B. Finance

This time series describes weekly 5/1-year adjustable rate mortgage average (2005-20) in the US [33]. After removing trend and seasonal components using the "mstl" R command [24], the data is highly correlated up to 300 lags. This example tests RCs' ability to model highly correlated data. We set $L_w = 100$, $L_t = 580$ and $L_v = 100$. All RCs achieve comparable RMSE. The QRC average RMSE is similar to the selected QRC's RMSE, whereas the difference for ESN is more pronounced. All RCs achieve uncorrelated residuals (within 95% CI) and pass the Lilliefors test; see Table I and Fig. 4. These suggest that both RCs are capable of modeling this highly correlated time series.

QRC under GAD experiences an increased RMSE. Despite this, QRC residuals under decoherence show no autocorrelation within 95% CI and pass the Lilliefors test with $p$-value$\geq$ 0.35. This suggests that decoherence does not significantly impact QRC's ability to model highly correlated data.

### C. Coupled electric drive system

We employ RCs on modeling a single-input (i.e., $u_k \in \mathbb{R}$) single-output nonlinear system consists of two electric motors driving a pulley using a flexible belt [34]. Input data is a pseudo-random binary sequence (persistently exciting) with amplitude 0.5 and only $L = 500$ data are available, presenting a challenge for RCs.

We exploit spatial multiplexing, see [14], where outputs of two distinct and non-interacting RC members are combined linearly. The first member processes both $\mathbf{y}_k$ and $\mathbf{u}_k$ as described in Sec. IV, and the second only processes $\mathbf{u}_k$. We label each member's parameters with subscripts $1, 2$. For simplicity, we set $N = N_1 = N_2$ so that dimensions of both members (or numbers of qubits) are the same, with $N = 2$ preferred by FPE for all RCs. For the second ESN member, we set $C_2 = 0$ and $\sigma_{\text{m}}(A_2) = 0.7$, where the latter ensures convergence. For the multiplexed QRC, we set $\epsilon_1 = 0.5$ and $\epsilon_2 = 0.9$, and set the second QRC's CPTP map $T^{(2)}(u_k)$ as

$$T^{(2)}(u_k) = g(u_k)T_1^{(2)} + (1 - g(u_k))T_2^{(2)},$$

where $T_l^{(2)}(\rho_k) = U_l^{(2)}\rho_k(U_l^{(2)})^\dagger$ for some arbitrary but fixed unitaries $U_l^{(2)}$ for $l = 1, 2$; see [27] for the details. The second

QRC is convergent by construction. For multiplexed RCs, output parameters of their second member are not constrained.
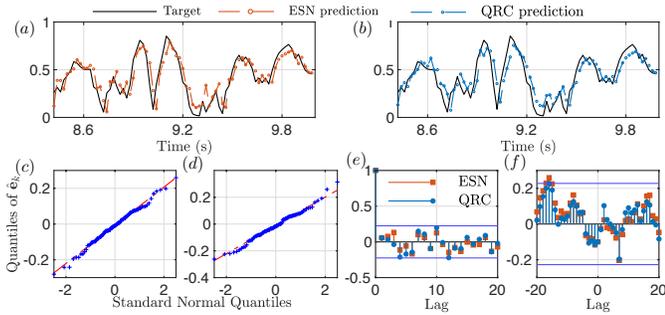


Fig. 5. The target output, (a) the ESN prediction and (b) the QRC prediction on validation data. (c) The ESN residual Q-Q plot. (d) The QRC residual Q-Q plot. (e) The ESN (blue) and QRC (red) sample ACF. (f) The ESN (blue) and QRC (red) sample cross-correlation between inputs and residuals.

We set $L_w = 20$, $L_t = 400$ and $L_v = 79$. All RCs achieve comparable RMSE and average RMSE, and they pass the Lilliefors test. Residuals of ESN and QRC are uncorrelated and independent of inputs (within 95% CI); see Table I and Fig. 5. Under both decoherence channels, residuals of QRC show autocorrelation at lag 1. Despite this, the increase in the QRC RMSE is small, by at most 0.01 compared to the noiseless QRC.

## VI. CONCLUSION

We have introduced convergent reservoir computers with output feedback as stationary and ergodic NARX($\infty$) models. Our approach can harness nonlinear dynamical systems for temporal information processing, making them versatile for nonlinear stochastic modeling. Numerical experiments demonstrate that these reservoir computers with a few tunable parameters are adequate in modeling diverse data sets.

Many exciting problems remain open for future research, such as improving the modeling performance through reservoir design and developing model selection methods for the proposed scheme. With this view, this work opens the potential for reservoir computing paradigm to tackle traditional challenges encountered in control and time series modeling, further bridging these scientific disciplines.

## REFERENCES

[1] J. Fan and Q. Yao, *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media, 2008.
[2] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.
[3] S. Boyd and L. Chua, "Fading memory and the problem of approximating nonlinear operators with volterra series," *IEEE Transactions on circuits and systems*, vol. 32, no. 11, pp. 1150–1161, 1985.
[4] S. N. Kumpati *et al.*, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on neural networks*, vol. 1, no. 1, pp. 4–27, 1990.
[5] I. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systems part ii: stochastic non-linear systems," *International journal of control*, vol. 41, no. 2, pp. 329–344, 1985.
[6] M. Schoukens and K. Tiels, "Identification of block-oriented nonlinear systems starting from linear approximations: A survey," *Automatica*, vol. 85, pp. 272–292, 2017.
[7] G. Tanaka *et al.*, "Recent advances in physical reservoir computing: A review," *Neural Networks*, vol. 115, pp. 100–123, 2019.
[8] K. Nakajima and I. Fischer, *Reservoir Computing: Theory, Physical Implementations, and Applications*. Springer Singapore, 2021.
[9] P. Billingsley, *Probability and measure*. John Wiley & Sons, 2008.
[10] J. Pathak *et al.*, "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach," *Physical review letters*, vol. 120, no. 2, p. 024102, 2018.
[11] T. Kim and B. R. King, "Time series prediction using deep echo state networks," *Neural Computing and Applications*, vol. 32, no. 23, pp. 17 769–17 787, 2020.
[12] K. Fujii and K. Nakajima, "Harnessing disordered-ensemble quantum dynamics for machine learning," *Physical Review Applied*, vol. 8, no. 2, p. 024030, 2017.
[13] J. Chen, H. I. Nurdin, and N. Yamamoto, "Towards single-input single-output nonlinear system identification and signal processing on near-term quantum computers," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 401–406.
[14] ——, "Temporal information processing on noisy quantum computers," *Phys. Rev. Applied*, vol. 14, p. 024065, Aug 2020.
[15] D. Marković and J. Grollier, "Quantum neuromorphic computing," *Applied Physics Letters*, vol. 117, no. 15, p. 150501, 2020.
[16] A. Pavlov and N. van de Wouw, "Convergent discrete-time nonlinear systems: the case of PWA systems," in *2008 American Control Conference*. IEEE, 2008, pp. 3452–3457.
[17] D. N. Tran, B. S. Rüffer, and C. M. Kellett, "Convergence properties for discrete-time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 8, pp. 3415–3422, 2018.
[18] L. Grigoryeva and J.-P. Ortega, "Echo state networks are universal," *Neural Networks*, vol. 108, pp. 495–508, 2018.
[19] L. Grigoryeva, J. Henriques, and J.-P. Ortega, "Reservoir computing: information processing of stationary signals," in *Joint 2016 CSE, EUC and DCABES*. IEEE, 2016, pp. 496–503.
[20] L. Gonon and J.-P. Ortega, "Reservoir computing universality with stochastic inputs," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 1, pp. 100–112, 2019.
[21] E. Bollt, "On explaining the surprising success of reservoir computing forecaster of chaos? The universal machine learning dynamical system with contrast to VAR and DMD," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 31, no. 1, p. 013108, 2021.
[22] Y. Lin, E. D. Sontag, and Y. Wang, "A smooth converse Lyapunov theorem for robust stability," *SIAM Journal on Control and Optimization*, vol. 34, no. 1, pp. 124–160, 1996.
[23] J. Chen and H. I. Nurdin, "Learning nonlinear input–output maps with dissipative quantum systems," *Quantum Information Processing*, vol. 18, no. 7, p. 198, 2019.
[24] R. Hyndman *et al.*, *forecast: Forecasting functions for time series and linear models*, 2020, R package version 8.12. [Online]. Available: http://pkg.robjhyndman.com/forecast
[25] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.
[26] J. Löfberg, "YALMIP : A Toolbox for Modeling and Optimization in MATLAB," in *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.
[27] J. Chen and H. I. Nurdin, "Nonlinear autoregression with convergent dynamics on novel computational platforms," *arXiv preprint*, 2021.
[28] S. Konishi and G. Kitagawa, *Information criteria and statistical modeling*. Springer Science & Business Media, 2008.
[29] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
[30] IBM, *Device backend noise model simulations*, 2021. [Online]. Available: https://qiskit.org/documentation/tutorials/simulators/
[31] C. Santori *et al.*, "Quantum noise in large-scale coherent nonlinear photonic circuits," *Physical Review Applied*, vol. 1, no. 5, p. 054005, 2014.
[32] J. R. Johansson, P. D. Nation, and F. Nori, "QuTiP: An open-source Python framework for the dynamics of open quantum systems," *Computer Physics Communications*, vol. 183, no. 8, pp. 1760–1772, 2012.
[33] *5/1-Year adjustable rate mortgage average in the United States*, 2020. [Online]. Available: https://fred.stlouisfed.org/series/MORTGAGE5US
[34] T. Wigren and M. Schoukens, "Coupled electric drives data set and reference models," *Technical Report, Department of Information Technology, Uppsala University*, 2017.