

RESEARCH ARTICLE

Autoregressive With Slack Time Series Model for Forecasting a Partially-Observed Dynamical Time Series

AKIFUMI OKUNO^{1,2}, YUYA MORISHITA³, AND YOH-ICHI MOTOTAKE⁴¹School of Statistical Thinking, The Institute of Statistical Mathematics, Tachikawa 190-8562, Japan²RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan³Department of Nuclear Engineering, Kyoto University, Kyoto 606-8501, Japan⁴Graduate School of Social and Data Sciences, Hitotsubashi University, Kunitachi 186-8601, Japan

Corresponding author: Yoh-Ichi Mototake (y.mototake@r.hit-u.ac.jp)

This work was supported by the Strategic Research Projects from the Research Organization of Information and Systems through the Project “The Statistical and Mathematical Modeling for Plasma Physics and Complementary Plasma Data,” under Grant 2022-SRP-13. The work of Akifumi Okuno was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 21K17718 and Grant 22H05106. The work of Yoh-Ichi Mototake was supported in part by the Japan Science and Technology Agency (JST) Precursory Research for Embryonic Science and Technology (PRESTO) under Grant JPMJPR212A, in part by JSPS KAKENHI under Grant 22K13979 and Grant 23H03460, and in part by the New Energy and Industrial Technology Development Organization (NEDO) under Grant JPNP22100843-0.

ABSTRACT This study delves into the domain of dynamical systems, specifically the forecasting of dynamical time series defined through an evolution function. Traditional approaches in this area predict the future behavior of dynamical systems by inferring the evolution function. However, these methods may confront obstacles due to the presence of missing variables, which are usually attributed to challenges in measurement and a partial understanding of the system of interest. To overcome this obstacle, we introduce the autoregressive with slack time series (ARS) model, that simultaneously estimates the evolution function and imputes missing variables as a slack time series. Assuming time-invariance and linearity in the (underlying) entire dynamical time series, our experiments demonstrate the ARS model’s capability to forecast future time series. From a theoretical perspective, we prove that a 2-dimensional time-invariant and linear system can be reconstructed by utilizing observations from a single, partially observed dimension of the system.


INDEX TERMS Autoregressive model, completely missing variables, dynamical system, slack time series.

I. INTRODUCTION

Notwithstanding its difficulty, forecasting of the evolution of intricate non-linear dynamical time series has been in the spotlight in various scientific fields [1], [2]. A plausible approach to forecasting the evolution is to isolate the non-linear estimation problem into (i) learning non-linear representations by applying highly non-linear functions such as deep neural networks [3], and (ii) estimating its evolution with simple linear models. One example is a reservoir computing (RC) [4], [5]. RC first randomly specifies a state in the reservoir layer in recurrent neural network [6], and optimizes the weights only in the output layer; RC corresponds to

non-linearly transform its input (in the reservoir layer) and trains a simple linear prediction model (in the output layer). It has been reported that such a simple combination of the non-linear representation learning and the linear estimation is effective to forecasting the evolution of intricate dynamical series [7]. Effectiveness of the simple combination is not limited to RC; applying a linear model to the non-linear representation in more general deep neural network is also regarded as a solid forecasting strategy [8].

Unfortunately, however, partial degrees of freedom corresponding to several state variables are not observed in some practical situations [9], [10]. There could be a variety of reasons for missing observations: it would be caused by the difficulty of measurement, it would be caused by the immature understanding of the system of interest, and so forth.

The associate editor coordinating the review of this manuscript and approving it for publication was Sajid Ali .

Generally speaking, it is quite difficult to find and identify all the related state variables in the system in real world situations. To address the issue, studies in line with dynamical systems have considered and substituted the delay embedding [11], [12], with the aid of delay embedding theorem (also known as Takens' theorem) [13]. While the delay embedding theorem indicates that the (topological) attractor structure of the entire dynamical system is recovered by the delay embedding of the partial observations, i.e., this theorem provides a "rough estimate" of the entire system, it does not intend to provide any accurate forecasting. Therefore, for more accurate forecasting purposes, [14] leverages the random feature maps of the delay embedding, [15] applies a deep neural network to the delay embedding, and [16] proposes estimating the missing variables themselves directly by a simple transformation of the observed variables. Therein, neural ordinal differential equation (neural ODE) [19] is used to model the evolution of the time series. In line with these approaches, forecasting of the evolution of a partially-observed dynamical time series with the estimation of the missing variables has been actively studied recently [10].

Recent research has leveraged deep neural networks to reconstruct the full dynamics of systems, but the complexity of training these networks highlights the appeal of simpler, more manageable models as alternatives. Consider a straightforward scenario where a 2-dimensional time series exhibits circular motion, yet only the first dimension (corresponding to a cosine curve) is observable. In such cases, while the complete dynamics adhere to a simple, time-invariant linear system at regular discrete time points (that are considered in many literature; see, e.g., [17] and [18]), the observed cosine function exhibits time-variance. This discrepancy suggests that partial observations may appear to follow ostensibly complex patterns, even when the underlying dynamics are fundamentally simpler. Motivated by this fact, we introduce the autoregressive with slack time series (ARS) model, designed to estimate the evolution function and impute missing variables using a slack time series simultaneously. By assuming the entire underlying dynamical time series is time-invariant and linear, our experiments validate the efficacy of the ARS model. Theoretically, we demonstrate that it is possible to reconstruct a 2-dimensional, time-invariant, and linear system using observations from just one partially observed dimension, as depicted in Figure 1.

Organization of this paper is as follows: Section I-A describes the symbols and notations used throughout this paper, Section II describes the preliminaries, Section III describes the proposed ARS model, Section IV describes the numerical experiments, and Section V describes the remaining discussions and conclusions. Particularly, the discussion section includes a theoretical guarantee for the ARS model prediction with a simple setting, and relations to classical state-space models and higher-order autoregressive (AR) models.

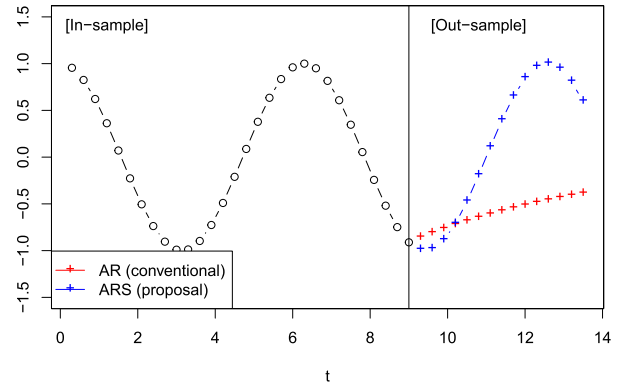


FIGURE 1. Time series forecasting with (red) conventional AR model shown in Equation (4) and (blue) proposed ARS model shown in Equation (5). See Section III-A for further details of the experiment.

A. SYMBOLS AND NOTATIONS

This section provides a summary of the symbols used throughout this paper. The symbol $t \geq 0$ represents time, and $x = x(t)$ signifies a state at time t , which takes on a value within the non-empty state space $\mathcal{X} \subset \mathbb{R}^d$. As long as $x(t)$ follows a dynamical system, the temporal state $x(t)$ is also referred to as the dynamical time series. Time t is discretized into small intervals $h > 0$, such that $t = jh$ for $j = 1, \dots, n$. At each discretized time point $t = jh$, the state $x(jh) \in \mathbb{R}^d$ is split into an observed component $z(jh) \in \mathbb{R}^r$ and a missing (hidden) component $z^\dagger(jh) \in \mathbb{R}^s$, resulting in $x(jh) = (z(jh), z^\dagger(jh))$. The sequence $\{z_j^\dagger\}_{j=1}^n$ represents a slack time series, where each element z_j^\dagger acts as a substitute for the missing component $z^\dagger(jh)$. This study aims to estimate the slack time series as \hat{z}_j^\dagger , with the completed time series expressed as $\hat{x}^\dagger(jh) = (z(jh), \hat{z}_j^\dagger)$ for $j = 1, 2, \dots, n$.

II. PRELIMINARIES

This section provides preliminaries. More specifically, Section II-A describes the dynamical system, Section II-B describes the autoregressive model, and Section II-C describes the problem setting considered in this study.

A. DYNAMICAL SYSTEM

Let $d \in \mathbb{N}$. A dynamical system is a pair (\mathcal{X}, φ) , where $\mathcal{X} \subset \mathbb{R}^d$ is a non-empty set called state space and $\varphi : \mathbb{R}_{\geq 0} \times \mathcal{X} \rightarrow \mathcal{X}$ denotes an evolution function satisfying

$$\varphi(t', \varphi(t, \mathbf{x})) = \varphi(t + t', \mathbf{x}), \quad \varphi(0, \mathbf{x}) = \mathbf{x}$$

for all $t, t' \geq 0$ and $\mathbf{x} \in \mathcal{X}$. Intuitively speaking, $\varphi(t, \mathbf{x}) \in \mathcal{X}$ represents the state evolved from $\mathbf{x} \in \mathcal{X}$ during the period of time $t \geq 0$. As $\mathbf{x} = \mathbf{x}(t) \in \mathcal{X}$ depends on the time $t \geq 0$, we call $\mathbf{x}(t)$ as a (dynamical) time series herein. Typically, the dynamical system can be specified by a differential equation

$$\frac{d\mathbf{x}(t)}{dt} = \frac{d\varphi(t', \mathbf{x}(t))}{dt'} \Big|_{t'=0} = f_t(\mathbf{x}(t))$$

equipped with the time derivative $f_t : \mathcal{X} \rightarrow \mathcal{T}$ defined with some set $\mathcal{T} \subset \mathbb{R}^d$. The time derivative f_t is called

time-invariant if f_t is independent of the time t (herein, f represents such a time-invariant derivative); an example of such a time-invariant dynamical system (with $d = 3$, $\mathbf{x} = (x_1, x_2, x_3)$) is the Lorenz system [20] $f(\mathbf{x}) = (-\alpha x_1 + \alpha x_2, -x_1 x_3 + \beta x_1 - x_2, x_1 x_2 - \gamma x_3)$ for some constants $\alpha, \beta, \gamma \in \mathbb{R}$. See, e.g., [1] for the long-standing history of the dynamical systems.

While the above dynamical system considers the evolution of the series in continuous time, in practice, the evolving states are observed only for the discrete-time $t = h, 2h, \dots, nh$; with sufficiently short period of time $h > 0$, the evolution of the time-invariant dynamical system can be approximated by a first-order Taylor expansion

$$\begin{aligned} \mathbf{x}(t+h) &= \varphi(h, \mathbf{x}(t)) \\ &= \varphi(0, \mathbf{x}(t)) + h \left. \frac{d\varphi(t', \mathbf{x}(t))}{dt'} \right|_{t'=0} + O(h^2) \\ &= \mathbf{x}(t) + hf(\mathbf{x}(t)) + O(h^2). \end{aligned} \tag{1}$$

$O(h^2)$ denotes the term smaller than h^2 in the limit $h \searrow 0$. By assuming the linearity in the time derivative f , i.e., $f(\mathbf{x}) = A\mathbf{x}$ for some matrix $A \in \mathbb{R}^{d \times d}$, the approximation shown in Equation (1) indicates that the state evolved for the short period of time $h \geq 0$ is approximated by a simple linear transformation

$$\mathbf{x}(t+h) = B\mathbf{x}(t) + O(h^2) \tag{2}$$

for some matrix $B = B(h) := I + hA \in \mathbb{R}^{d \times d}$ defined with the $d \times d$ identity matrix I . The short period approximation of the time-invariant dynamical system shown in Equation (2) leads to the autoregressive model described in the following Section II-B.

While the discussions below consider only the linear time derivative f for simplicity, dynamical systems equipped with even non-linear (and continuous) time derivative f can be approximated by a polynomial extension. Also see the discussion in Appendix A.

B. AUTOREGRESSIVE MODEL

With a positive integer n , assume that the evolving states $\mathbf{x}(t)$ are observed at the discrete timepoints $t = h, 2h, 3h, \dots, nh$. For modeling the short period approximation of the dynamical system shown in Equation (2), we may employ an autoregressive (AR) model of order 1 :

$$\hat{\mathbf{x}}((j+1)h) = \hat{B}\mathbf{x}(jh) \quad (j = 1, 2, \dots, n-1), \tag{3}$$

where $\hat{B} \in \mathbb{R}^{d \times d}$ is a matrix typically estimated by minimizing the loss function

$$\hat{B} := \arg \min_{B \in \mathbb{R}^{d \times d}} \sum_{j=1}^{n-1} \|\mathbf{x}((j+1)h) - B\mathbf{x}(jh)\|_2^2.$$

While this study considers the AR model of order $p = 1$ (typically denoted by AR(1)) for simplicity, the AR(1) model can be straightforwardly extended to the AR model of higher order $p \in \mathbb{N}$: $\hat{\mathbf{x}}((j+1)h) = \sum_{k=1}^p \hat{B}_k \mathbf{x}((j+1-k)h)$, with the

estimated matrices $\{\hat{B}_k\}_{k=1}^p \subset \mathbb{R}^{d \times d}$. See, e.g., [21] and [22] for details of the extensions, and Section V-C for further discussions.

C. PROBLEM SETTING

Herein, assume that the evolving time series $\{\mathbf{x}(jh)\}_{j=1,2,\dots,n}$ follows the time-invariant dynamical system (shown in Equation (2)) equipped with the linear time derivative f . The AR model shown in Equation (3) is expected to approximate the dynamics well. However, in some practical situations, not all the variables in the state $\mathbf{x} \in \mathcal{X}$ can be observed; we assume that the time series is divided into two time series of observed and missing (hidden) variables:

$$\mathbf{x}(jh) = (\mathbf{z}(jh), \mathbf{z}^\dagger(jh)) \quad (j = 1, 2, \dots, n),$$

where $\mathbf{z}(t) \in \mathbb{R}^r$ represents the time series of the observed variables (of interest) and $\mathbf{z}^\dagger(t) \in \mathbb{R}^s$ represents that of the missing variables, with $r, s \in \mathbb{N}$ satisfying $r + s = d$. As the time series $\mathbf{z}^\dagger(t)$ is missing, the conventional AR model can consider only the observed part in this setting, i.e.,

$$\hat{\mathbf{z}}((j+1)h) = \hat{C}\mathbf{z}(jh) \quad (j = 1, 2, \dots, n-1), \tag{4}$$

where $\hat{C} \in \mathbb{R}^{r \times r}$ is a matrix typically estimated by minimizing the loss function $\sum_{j=1}^{n-1} \|\mathbf{z}((j+1)h) - C\mathbf{z}(jh)\|_2^2$.

However, while the time-invariance of the system is considered important as also discussed in Section II-A, evolution of the partial observation $\mathbf{z}(t)$ may lose the time-invariance property, even if the entire dynamics $\mathbf{x}(t)$ is time-invariant. A simple example is the circular motion; see Example 1 and Figure 2.

Example 1: Consider the case that $\mathbf{x}(t)$ follows a circular motion in \mathbb{R}^2 , i.e., $\mathbf{x}(t) = (\cos t, \sin t)$, and assume that $r = s = 1$, i.e., $\mathbf{z}(t) = \cos t, \mathbf{z}^\dagger(t) = \sin t$. Then, the evolution of the partial observation $\mathbf{z}(t) = \cos t$ is time-variant while that of the entire time series $\mathbf{x}(t)$ is time-invariant (as $\mathbf{x}(t+h) = B\mathbf{x}(t)$ for some rotation matrix $B = B(h)$).

III. AUTOREGRESSIVE WITH SLACK TIME SERIES (ARS) MODEL

Section III-A describes the proposed ARS model, and Section III-B describes the estimation procedure of the ARS model.

A. ARS MODEL

To address the problems described in Section II-C, this study proposes autoregressive with slack time series (ARS) model, whose definition is based on a simple idea. In this model, we simply impute the missing variable $\{\mathbf{z}^\dagger(jh)\}_{j=1}^n$ in AR model by a slack time series \mathbf{z}_j^\dagger to be estimated from the (partial) observations. Using the completed variables $\mathbf{x}^\dagger(jh) = (\mathbf{z}(jh), \mathbf{z}_j^\dagger)$, ARS model is then defined by

$$\hat{\mathbf{z}}((j+1)h) = E_{r,s} \hat{B} \mathbf{x}^\dagger(jh), \quad \mathbf{x}^\dagger(jh) = \begin{pmatrix} \mathbf{z}(jh) \\ \mathbf{z}_j^\dagger \end{pmatrix}, \tag{5}$$

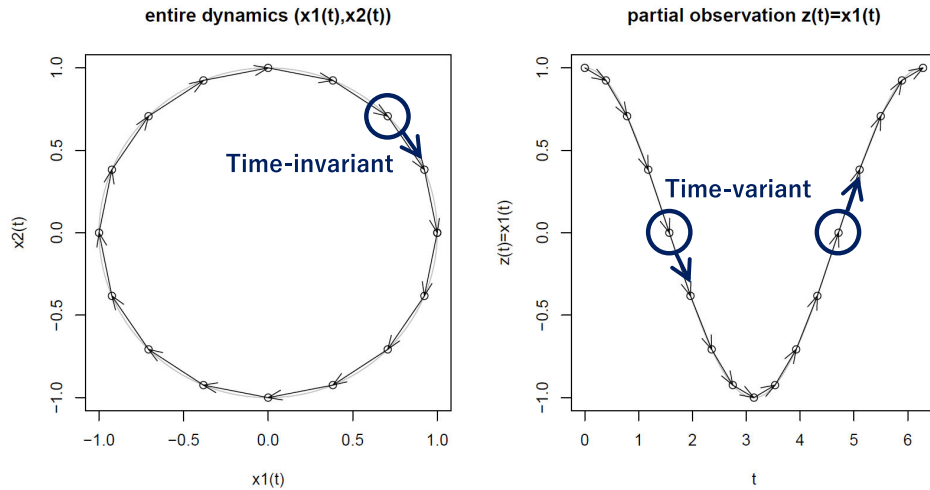


FIGURE 2. Circular motion. While (left) the entire dynamics $x(t) = (\cos t, \sin t)$ is time-invariant (as $x(t+h)$ can be identified as a function of $x(t)$), (right) the partial observation $z(t) = x_1(t) = \cos t$ is time-variant (as the next state of $x(t) = 0$ can be both $x(t+h) > 0$ and $x(t+h) < 0$ depending on the current time t).

where $E_{r,s} := (I_r, O_{r,s})$ is a $r \times d$ matrix defined with the $r \times r$ identity matrix I_r and the $r \times s$ zero matrix $O_{r,s}$. \hat{B} is a matrix and $\hat{z}_j^\dagger \in \mathbb{R}^{\tilde{s}}$ is a vector, where they are estimated by solving the following problem:

$$\begin{aligned}
 (\hat{B}, \{\hat{z}_j^\dagger\}_{j=1}^n) &= \arg \min_{(B, \{z_j^\dagger\}_{j=1}^n)} \ell(B, \{z_j^\dagger\}_{j=1}^n), \text{ where} \\
 \ell(B, \{z_j^\dagger\}_{j=1}^n) &:= \sum_{j=1}^{n-1} \|x^\dagger((j+1)h) - Bx^\dagger(jh)\|_2^2. \quad (6)
 \end{aligned}$$

$\{z_j^\dagger\}$ is especially called *slack time series*, and the dimension $s \in \mathbb{N}$ of the vector z_j^\dagger is a user-specified parameter. The above optimization problem is solved with the aid of the simple linear regression analysis; see Section III-B for details.

While the idea behind the ARS model (i.e., the time series $\{z^\dagger(jh)\}$ of missing variables is completed by the estimated slack time series $\{\hat{z}_j^\dagger\}$) seems intuitive and simple enough, one natural question here is whether the completion really works. Regarding this question, for simplicity, this study assumes the time-invariance (and linearity) of the entire dynamical system to recover the entire dynamics from the partial observations. Then, we demonstrate the ARS model by a simple numerical experiment with missing variables shown in Figure 1. Therein, we first compute the dynamical time series $x(t) = (z(t), z^\dagger(t)) = (\cos t, \sin t)$ following the circular motion and assume that $r = s = 1$ (i.e., $z(t) = \cos t$ is observed and $z^\dagger(t) = \sin t$ is missing) and $h = 0.3$, $n = 30$. The overall dynamics of the entire circular motion $x(t) = (\cos t, \sin t)$ is time-invariant, while the dynamics of the partial observation $z(t) = \cos t$ is time-variant. Then, we forecast the future time series by leveraging AR(1) and ARS models (with $\tilde{s} = 1$). Both AR(1) and ARS models are trained with only the observed time series $\{z(t)\}$ though ARS model additionally estimates the slack time

series $\{z^\dagger(t)\}$. In this experiment of the circular motion, the slack time series is initialized by a standard normal random numbers, and is optimized by `optim` function in R language. The affirmative experimental result is also proved by our proposition; Section V-A shows for the above simple case $d = 2, r = s = 1$ that the ARS model can recover the underlying true dynamics.

Historically speaking, it has been widely known that the time series of partially observed variables contain rich enough information to (partially) recover the entire dynamics. The well known Takens' theorem (also known as delay embedding theorem) [13] proves under some assumptions that the delay embedding $(z(t), z(t-h), z(t-2h), \dots, z(t-(k-1)h)) \in \mathbb{R}^k$ computed only from the partial observation $z(t)$ has the same (topological) attractor structure to the entire dynamical time series $\{x(t)\}$. Namely, the delay embedding at least roughly recovers the (topological) "shape" of the entire dynamical time series. While the classical Takens' theorem considers only the deterministic sequence, Takens' theorem can be further generalized to stochastic variants. See, e.g., [23].

Although Takens' theorem only provides the rough estimate of the entire dynamical system through the delay embedding (computed only from the partial observations), we find that directly estimating the missing variables can improve the forecasting accuracy. Therefore, compared to previous approaches based on Takens' theorem, the simpler ARS model holds the potential to offer a more general and comprehensible model.

We last note that the minimizer of the ARS loss function $\ell(B, \{z_j^\dagger\}_{j=1}^n)$ is not unique. For instance, if we multiply z_j^\dagger by any positive real number $\alpha > 0$, the matrix B_α , whose corresponding rows are also multiplied by $1/\alpha$, yields the same loss function value $\ell(B, \{z_j^\dagger\}_{j=1}^n) = \ell(B_\alpha, \{\alpha z_j^\dagger\}_{j=1}^n)$. There remain the freedom of the constant multiplication,

though this multiplication does not affect the forecast, i.e., $\hat{\mathbf{x}}^\dagger((j+1)h) = \hat{\mathbf{B}}\mathbf{x}^\dagger(jh) = \hat{\mathbf{B}}_\alpha\mathbf{x}_\alpha^\dagger(jh)$, where $\mathbf{x}_\alpha^\dagger(jh) = (\mathbf{z}(jh)^\top, \alpha\hat{\mathbf{z}}_j^\top)^\top$. Also see Section Section V-A for our proposition, indicating the prediction uniqueness of the ARS model.

B. ESTIMATION OF THE ARS MODEL

While ARS model needs to estimate the AR model parameter $B \in \mathbb{R}^{d \times d}$ and the slack time series $\{\mathbf{z}_j^\dagger\}_{j=1}^n$ simultaneously, we can skip the estimation of the parameter $B \in \mathbb{R}^{d \times d}$. We describe the procedure in the following.

With given vectors $\{\hat{\mathbf{z}}_j^\dagger\}$ (whereby we obtain $\mathbf{x}^\dagger(jh) = (\mathbf{z}(jh)^\top, \hat{\mathbf{z}}_j^\top)^\top$), the AR model parameter $B \in \mathbb{R}^{d \times d}$ is estimated by a simple matrix formulae used in the linear regression analysis:

$$\arg \min_{B \in \mathbb{R}^{d \times d}} \ell(B, \{\hat{\mathbf{z}}_j^\dagger\}_{j=1}^n) = (\hat{D}^\top \hat{D})^{-1} \hat{D}^\top \hat{D}_+,$$

where $\hat{D} \in \mathbb{R}^{(n-1) \times d}$ denotes the matrix concatenating the vectors $\mathbf{x}^\dagger(jh) \in \mathbb{R}^d$ for $j = 1, 2, \dots, n-1$, and $\hat{D}_+ \in \mathbb{R}^{(n-1) \times d}$ denotes the matrix of 1 step further of the time points, i.e., the matrix concatenating $\mathbf{x}^\dagger((j+1)h) \in \mathbb{R}^d$. Note that the matrices \hat{D}, \hat{D}_+ depend on the vectors $\{\hat{\mathbf{z}}_j^\dagger\}_{i=1}^n$. Then, the minimum loss function value is also obtained as

$$\min_{B \in \mathbb{R}^{d \times d}} \ell(B, \{\hat{\mathbf{z}}_j^\dagger\}_{j=1}^n) = \text{tr}\{\hat{D}_+^\top (I - \hat{H}) \hat{D}_+\}, \quad (7)$$

where $\hat{H} = \hat{D}(\hat{D}^\top \hat{D})^{-1} \hat{D}^\top$ is the hat matrix and $\text{tr}H = \sum_{j=1}^d h_{jj}$ denotes the trace of the matrix $H = (h_{jk})$, i.e., the sum of diagonal entries. Equation (7) is obtained by $\hat{\mathbf{x}}^\dagger = \hat{D}\hat{B} = \hat{D}(\hat{D}^\top \hat{D})^{-1} \hat{D}^\top \hat{D}_+ = \hat{H}\hat{D}_+$ and $\hat{H}^2 = \hat{H}$ (see, e.g., [24] for basic matrix formulae). As Equation (7) indicates that

$$\begin{aligned} \min_{(B, \{\hat{\mathbf{z}}_j^\dagger\}_{j=1}^n)} \ell(B, \{\hat{\mathbf{z}}_j^\dagger\}_{j=1}^n) &= \min_{\{\hat{\mathbf{z}}_j^\dagger\}_{j=1}^n} \min_{B \in \mathbb{R}^{d \times d}} \ell(B, \{\hat{\mathbf{z}}_j^\dagger\}_{j=1}^n) \\ &= \min_{\{\hat{\mathbf{z}}_j^\dagger\}_{j=1}^n} \text{tr}\{\hat{D}_+^\top (I - \hat{H}) \hat{D}_+\}, \end{aligned} \quad (8)$$

in practice, we may solve the minimization problem (8) by leveraging some general-purpose optimization functions. We use `optim` function in R language in our implementation. Appendix A provides a possible extension of the ARS model to consider the interaction effects.

IV. NUMERICAL EXPERIMENTS

We examine AR and the proposed ARS models using synthetic datasets. Particularly, the experimental settings and results are shown in Section IV-A and IV-B, respectively. R source codes to reproduce the experimental results are provided in <https://github.com/oknakfm/ARS>.

A. SETTINGS

1) SYNTHETIC DATASET GENERATION

we generate two different types of synthetic datasets, following (i) the circular motion and (ii) Lorenz dynamics. More specifically, (i) and (ii) are defined as follows. (i)

- 1) Circular motion: $\mathbf{x}(j) = (\cos(5 + j/20), \sin(5 + j/20))$.
- 2) Lorenz dynamics: define the evolution function $g : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ (of period of a constant length) for $\mathbf{x} = (x_1, x_2, x_3)$ as follows:

$$\begin{aligned} g(\mathbf{x}) &:= \left(\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \right. \\ &\quad \left. + \frac{1}{200} \begin{pmatrix} -\alpha & \alpha & 0 & 0 & 0 & 0 \\ \beta & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & -\gamma & 1 & 0 & 0 \end{pmatrix} \right) \\ &\quad \times (x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3)^\top, \end{aligned}$$

where $\alpha = 10, \beta = 28, \gamma = 8/3$. The above g is a first-order Taylor approximation of the original Lorenz dynamics [20] shown in Equations (10)–(12). With the function $g^{(m)}$, which is the composition of the function g of degree $m \in \mathbb{N}$, define $\mathbf{x}(0) = g^{(100)}(1/4, 1/4, 1/4)$ and $\mathbf{x}(j+1) = g(\mathbf{x}(j))$.

For the training set, we generate 10 instances of the sequence of length $n = 100$ for settings (i) and (ii). Independent normal errors with the standard deviations $\sigma = 0, 0.01$ are incorporated to the training sequences. For the test set (of forecasting), we compute the subsequent sequence of length $n = 30, 100$, for settings (i) and (ii), respectively.

2) MISSING MECHANISM

for the setting (i), we regard the first $r = 1$ entry as the observed variable: $\mathbf{z}(j) = x_1(j)$, and the remaining $s = 1$ entry as the missing variable: $\mathbf{z}^\dagger(j) = x_2(j)$. For the setting (ii), we regard the first $r = 2$ entry as the observed variable: $\mathbf{z}(j) = (x_1(j), x_2(j))$, and the remaining $s = 1$ entry as the missing variable: $\mathbf{z}^\dagger(j) = x_3(j)$.

3) METHODS TO BE COMPUTED

using the training sets (of observed variables), we compute AR and ARS models. For ARS models, we employ $\tilde{s} = 1$; we initialize the parameter by adding the standard Gaussian noise to the true missing parameter $\mathbf{z}^\dagger(j)$. We employ `optim` function with BFGS option in R package, to train the ARS models. While the slack time series here is initialized by adding random noise to the true missing parameters (for computational stability purposes), note that the series is initialized randomly by a standard normal distribution (without referencing the true parameters), in the forecasting of the circular motion shown in Figure 1.

B. RESULTS

Using the training instances, we train AR and ARS models, and forecast $\mathbf{z}(n+k)$, for $k = 5, 10, \dots, 25$. For the estimator $\hat{\mathbf{z}}(n+k)$ for each method, we compute the MSE $\hat{e}_k := \|\mathbf{z}(n+k) - \hat{\mathbf{z}}(n+k)\|_2^2/r$ for $k = 5, 10, \dots, 25$. We further compute the relative error of ARS models to the baseline (AR model), and compute the mean and the standard deviation. The results for the settings (i) and (ii) are shown in Tables 1–2. We can observe that all the relative errors are less than 1, i.e., the MSE

of ARS models are smaller than that of the conventional AR models.

For visualization, the training sets and the forecasts of AR/ARS models (using the training instance 1) are also plotted in Figures 3–6. We can examine that ARS models forecast the true points better than the conventional AR model.

V. DISCUSSIONS AND CONCLUSION

Section V-A, V-B and V-C provide discussion on theory, and the relation to state-space model and higher-order AR model, respectively, and Section V-D concludes this study.

A. DISCUSSION 1: ARS MODEL CAN RECOVER THE UNDERLYING TRUE DYNAMICAL TIME SERIES

The ARS model involves a significant number of parameters: $\{z_j^\dagger\}_{j=1}^n$ contains $n \cdot s$ parameters, and B contains d^2 parameters. Therefore, ARS model has a high degree of freedom (at least the ARS model is overparameterized, i.e., the number of parameters exceeds the sample size n), and it remains uncertain whether the ARS prediction is obtained correctly and uniquely.

In the specific scenario of $d = 2, r = s = 1$, the answer to this question is affirmative. This section provides theoretical evidence that the ARS model can successfully recover the underlying true dynamical time series, regardless of the ostensibly large number of parameters.

To rigorously describe the theory, we employ the following notations: let $h > 0$ be a fixed interval, and assume that $\mathbf{x}_j^{*\dagger} = (z_j^\top, z_j^{*\dagger\top})^\top$ with $z_j := z(jh)$ follows an underlying true linear dynamics

$$\mathbf{x}_{j+1}^{*\dagger} = B^* \mathbf{x}_j^{*\dagger}, \quad (j = 1, 2, \dots). \quad (9)$$

Then, Proposition 1 holds.

Proposition 1: Let $d = 2, r = s = 1$ and assume the identity (9) for the underlying true dynamics. With sufficiently large n , the future prediction via the ARS model: $\hat{z}_{n+k} = (1, 0)\hat{B}^k \hat{\mathbf{x}}_n^\dagger$ coincides with the underlying true z_{n+k} , for any $k \in \mathbb{N}$.

Proposition 1 proves for the 2-dimensional case (with 1-dimensional observation) that the ARS model can recover the underlying true dynamical time series. See Appendix B for the proof. While we skip generalizing the proposition for simplicity, we can expect that the similar holds for $d > 2$ by considering the numerical results shown in Section IV.

B. DISCUSSION 2: RELATION TO STATE-SPACE MODEL

While the aforementioned experiments consider a small random error (where $\sigma = 0, 0.01$), we may employ a state-space model to deal with more influential random errors. A state-space model for linear systems is defined as

$$\begin{aligned} z((j+1)h) &\sim N(E_{r,s} \mathbf{x}((j+1)h), \Sigma_Z), \\ \mathbf{x}((j+1)h) &\sim N(B\mathbf{x}(jh), \Sigma_X), \end{aligned}$$

where $\{\mathbf{x}(jh)\}_{j=1}^n \subset \mathbb{R}^d$ represents a series of latent states, $E_{r,s} \in \mathbb{R}^{r \times d}$ represents the observation matrix,

and $B \in \mathbb{R}^{d \times d}$ denotes the transition matrix. The proposed ARS model corresponds to the above state-space model defined with the observation matrix $E_{r,s} = (I_r, O_{r,s})$ where I_r denotes the $r \times r$ identity matrix and $O_{r,s}$ denotes the zero matrix of size $r \times s$. While the state-space model is defined as a general framework, in practical situations, most of studies estimates only the hidden states \mathbf{x} by preliminarily fixing the model parameter B . See, e.g., Bayesian filters including Kalman filters [25], [26]. The difference to our ARS model is that the model parameter B is usually fixed in most of the Bayesian filters, while ARS estimates both the unobserved variables and the model parameter B simultaneously.

From another perspective, hidden markov model (HMM) [27] has much in common with the state-space model. However, in most cases, HMM employs a finite set for the state space \mathcal{X} , and estimate the transition probability from a state $\mathbf{x} \in \mathcal{X}$ to another state $\mathbf{x}' \in \mathcal{X}$. Baum-Welch algorithm [28] (also known as EM algorithm) assumes a latent structure on the hidden state \mathbf{x} , and taking marginal with respect to the unobserved \mathbf{x} . Viterbi algorithm [29] is closer approach to ours; it estimates the actual value of hidden state \mathbf{x} by leveraging dynamic programming. While the state space \mathcal{X} in HMM can be extended to a set of countably many states [30], HMM still focuses on modeling the transition of discrete states, while this paper considers a set of uncountably many states.

C. DISCUSSION 3: RELATION TO HIGHER ORDER AR MODEL

While this study so far considers AR(1) model $\hat{z}((j+1)h) = \hat{c}z(jh)$, we may employ higher order AR models. For instance, AR(2) model $\hat{z}((j+1)h) = \hat{c}_1 z(jh) + \hat{c}_2 z((j-1)h)$ is expected to have higher expressive power than the AR(1) model. This observation is true. Intuitively speaking, AR(1) model can approximate first-order differential equation by considering the relation $dz(t)/dt|_{t=jh} \approx \{z((j+1)h) - z(jh)\}/h$ with small $h > 0$, while AR(2) model can approximate second-order one by considering the second-order central $d^2z(t)/dt^2|_{t=jh} \approx \{z((j+1)h) - 2z(jh) + z((j-1)h)\}/h^2$. However, even increasing the order $p \in \mathbb{N}$, AR model is limited to approximate the dynamics described by a linear differential equation $\sum_{k=0}^p c_k d^k z(t)/dt^k = 0$ (for some $c_0, c_1, \dots, c_p \in \mathbb{N}$). For instance, AR model cannot approximate the partial observation $x_1(t)$ of Lorenz system

$$dx_1(t)/dt = -\alpha x_1(t) + \alpha x_2(t), \quad (10)$$

$$dx_2(t)/dt = -x_1(t)x_3(t) + \beta x_1(t) - x_2(t), \quad (11)$$

$$dx_3(t)/dt = x_1(t)x_2(t) - \gamma x_3(t). \quad (12)$$

defined with parameters $\alpha, \beta, \gamma \geq 0$. This is because the partial observation $x_1(t)$ of the Lorenz system shown in Equations (10)–(12) follows a non-linear differential equation. See Proposition 2 whose proof is shown in Appendix C.

Proposition 2: Partial observation $x_1(t)$ of Lorenz dynamics (shown in Equations (10)–(12)) follows a non-linear ordinary differential equation. More specifically,

TABLE 1. Relative errors to AR model, with the setting (i) circular motion.

	$k = 5$	$k = 10$	$k = 15$	$k = 20$	$k = 25$	
$\sigma = 0$	6.35 ± 5.37	7.26 ± 6.20	8.37 ± 7.28	9.82 ± 8.72	1.19 ± 1.09	$(\times 10^{-6})$
$\sigma = 0.01$	2.33 ± 1.80	2.95 ± 3.13	3.66 ± 3.72	4.14 ± 4.13	4.52 ± 4.45	$(\times 10^{-1})$

TABLE 2. Relative errors to AR model, with the setting (ii) Lorenz dynamics.

	$k = 5$	$k = 10$	$k = 15$	$k = 20$	$k = 25$	
$\sigma = 0$	0.05 ± 0.03	0.14 ± 0.07	0.49 ± 0.27	2.05 ± 1.26	6.62 ± 3.86	$(\times 10^{-2})$
$\sigma = 0.01$	0.13 ± 0.06	0.15 ± 0.10	0.28 ± 0.20	0.77 ± 0.58	1.77 ± 1.47	$(\times 10^{-1})$

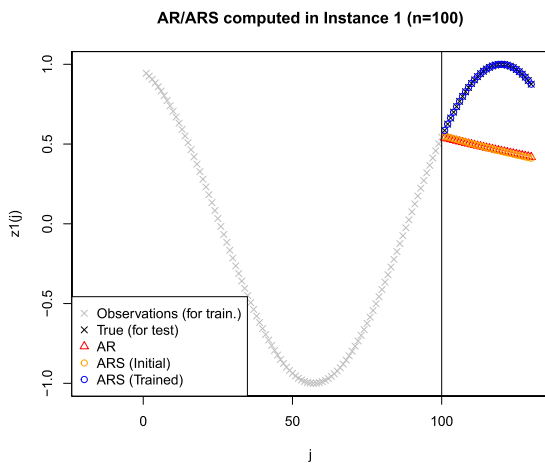


FIGURE 3. (i) Circular motion, $n = 100$, $\sigma = 0$.

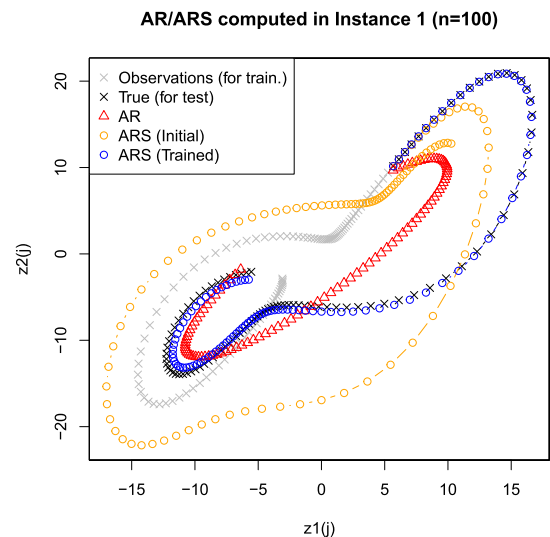


FIGURE 5. (ii) Lorenz, $n = 100$, $\sigma = 0$.

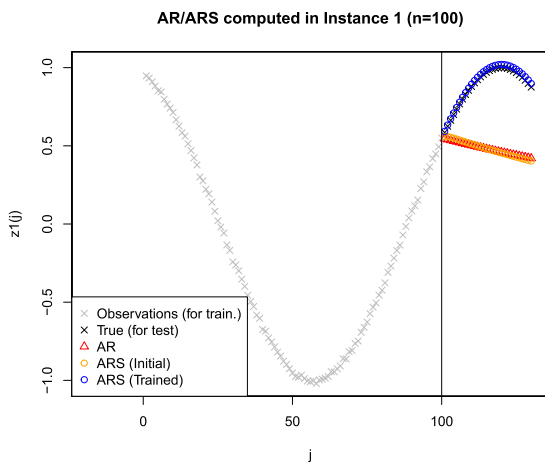


FIGURE 4. (i) Circular motion, $n = 100$, $\sigma = 0.01$.

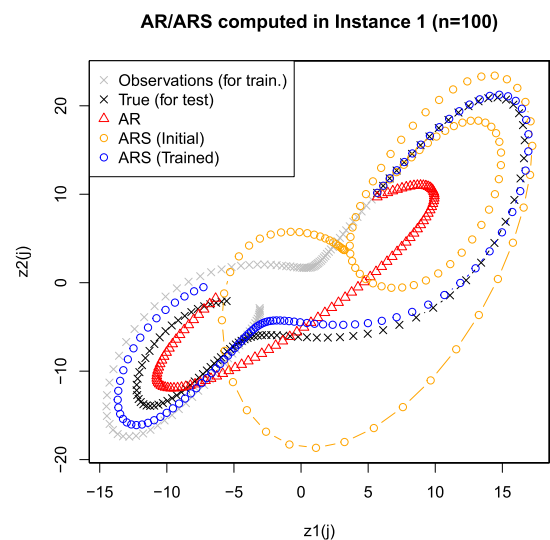


FIGURE 6. (ii) Lorenz, $n = 100$, $\sigma = 0.01$.

$x_1(t)$ satisfies $\sum_{k=0}^3 P_k(x_1(t))d^k x_1(t)/dt^k = 0$ for some functions P_1, P_2, P_3, P_4 and at least one of P_1, P_2, P_3, P_4 is a non-constant function.

Mathematically speaking, higher order AR model is a special case of the ARS model. Here, consider arbitrary time series $\{z(jh)\}_{j=1}^n$; this inclusion relation can be proved by substituting the delayed series $z(j-1)h$ into the slack time series \hat{z}_j^\dagger in ARS model. Then, the ARS model reduces to the

AR(2) model $\hat{z}((j+1)h) = \hat{c}_1 z(jh) + \hat{c}_2 z((j-1)h)$. General AR(p) model also can be implemented by substituting the delayed series $z((j-1)h), z((j-2)h), \dots, z((j-p+1)h)$ to the ARS model in the same way. Furthermore, the ARS

model (more rigorously speaking, extension of ARS model shown in Appendix A) is capable of representing the Lorenz equation while the higher-order AR model cannot describe Lorenz dynamics. We last note that the slack time-series can be incorporated to the higher-order AR models; we can easily define higher-order ARS models (though we skip the details in this paper for simplicity).

D. CONCLUSION

This study considered the setting that some variables in dynamical time series were missing; we extended the autoregressive (AR) model to propose AR with slack time series (ARS) model. The effectiveness of the ARS model was demonstrated by numerical experiments.

As this study provides the concept of the slack time series, there remain following limitations indicating potential future works.

- 1) The optimization procedure is currently lacking in efficiency. We may consider more efficient and stable optimization algorithms; one possible approach is to employ some parametric models (or some restrictions) for the slack time series $z^\dagger(t)$.
- 2) Random errors included in more practical dynamical time series should be removed. We may combine the concept of slack time series to the conventional filters for state space models.
- 3) The proposed ARS model still offers considerable scope for further theoretical exploration. It would be worthwhile to elucidate mathematical conditions that the proposed ARS approximates the underlying dynamics from the partial observations. Also the evaluation of the spatial and temporal complexity is preferred.
- 4) ARS model has a high expressive capability (at least ARS model generalizes general order AR model as described in Section V-C, and it includes n -s parameters therein). Given this complexity, the potential for overfitting warrants careful consideration. To address this, some strategies such as regularization appears essential for practical applications.
- 5) Lastly, the ARS model has been evaluated solely through experiments on synthetic datasets. We plan to estimate the proposed ARS to solve more practical problems. As an end-goal in mind, this study has been started to model the time evolution of the temperature/density profiles of fusion plasmas. This real-world application would be a demonstration for the potential utility of the proposed ARS model.

APPENDIX A EXTENSION: INTERACTION EFFECTS

As discussed in Section II-A and II-B, the AR model is derived from the short period approximation of the dynamical system (shown in Equation (1)) with the assumption $f(\mathbf{x}) = d\varphi(t, \mathbf{x})/dt = \mathbf{A}\mathbf{x}$ for some matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$.

Unfortunately, however, this assumption $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is restrictive. For instance, a dynamical system following a Lorenz equation $f(\mathbf{x}) = (-\alpha x_1 + \alpha x_2, -x_1 x_3 + \beta x_1 - x_2, x_1 x_2 - \gamma x_3)$ for some constants $\alpha, \beta, \gamma \in \mathbb{R}$ (see [1] for details) does not satisfy this assumption as interaction terms $x_1 x_2$ and $x_1 x_3$ are included therein.

Therefore, to deal with the interaction effects in ARS model, we consider a function which generates interaction terms:

$$\begin{aligned} \mathcal{J}(\mathbf{x}) &= (x_1, \dots, x_d, x_1 x_2, \dots, x_1 x_d, x_2 x_3, \dots, x_{d-1} x_d) \\ &: \mathbb{R}^d \rightarrow \mathbb{R}^{d(d+1)/2} \end{aligned} \quad (13)$$

for $\mathbf{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, and define an extended ARS model:

$$\hat{\mathbf{x}}^\dagger((j+1)h) = \hat{E}\mathcal{J}(\mathbf{x}^\dagger(jh)), \quad (14)$$

where $\mathbf{x}^\dagger(jh)$ is defined same as Equation (5). $\hat{E} \in \mathbb{R}^{d \times d(d+1)/2}$ is a matrix and $\hat{\mathbf{z}}_j^\dagger$ is a vector, where they are estimated by solving the following optimization problem:

$$\begin{aligned} (\hat{E}, \{\hat{\mathbf{z}}_j^\dagger\}_{j=1}^n) &= \arg \min_{(E, \{\mathbf{z}_j^\dagger\}_{j=1}^n)} \tilde{\ell}(E, \{\mathbf{z}_j^\dagger\}_{j=1}^n), \\ \tilde{\ell}(E, \{\mathbf{z}_j^\dagger\}_{j=1}^n) &:= \sum_{j=1}^{n-1} \|\mathbf{x}^\dagger((j+1)h) - \hat{E}\mathcal{J}(\mathbf{x}^\dagger(jh))\|_2^2. \end{aligned} \quad (15)$$

While this study considers only the interaction of order 2 for simplicity, the function (13) can be further generalized so as to include the interaction of order 3 (i.e., $x_1 x_2 x_3$) and so on. Note that polynomial functions of sufficiently high degree can approximate any continuous functions (Weierstrass' theorem; see, e.g., [31]); the extended ARS model equipped with the interaction terms of sufficiently high degree is expected to express any time-invariant dynamical system (equipped with continuous and non-linear time derivative f).

APPENDIX B PROOF OF PROPOSITION 1

In this proof, $\mathbf{e}_1 = (1, 0)^\top$, $\mathbf{e}_2 = (0, 1)^\top$ denote unit vectors. As the observations $\{\mathbf{z}_j\}_{j=1}^n$ are assumed to follow the linear dynamics $\mathbf{x}_{j+1}^\dagger = \mathbf{B}\mathbf{x}_j^\dagger$, we have the following identity:

$$\begin{aligned} 0 &= \min_{\mathbf{B}, \{\mathbf{z}_j^\dagger\}} \sum_{j=1}^{n-1} \|\mathbf{x}_{j+1}^\dagger - \mathbf{B}\mathbf{x}_j^\dagger\|_2^2 \\ &= \min_{\mathbf{B}, \mathbf{z}_1^\dagger} \sum_{j=1}^{n-1} \{\mathbf{z}_{j+1}^\dagger - \mathbf{e}_1^\top \mathbf{B}^j \mathbf{x}_1^\dagger\}^2, \end{aligned} \quad (16)$$

where the last equality is obtained by substituting $\mathbf{z}_{j+1}^\dagger = \mathbf{e}_2^\top \mathbf{B}^j \mathbf{x}_1^\dagger$. We prove the assertion by the following 2 steps.

Step 1: In this first step, we show that the ARS model $\mathbf{e}_1^\top \mathbf{B}^j \mathbf{x}_1^\dagger$ is specified by identifiable few parameters. To this end, we employ the eigendecomposition $\mathbf{B} = \mathbf{U}^{-1} \mathbf{\Lambda} \mathbf{U}$,

where Λ is a diagonal matrix whose diagonal entries are $\lambda_1, \lambda_2 \in \mathbb{C}$, and U is an eigen matrix. Then, we have

$$\begin{aligned} e_1^\top B^j x_1^\dagger &= e_1^\top U^{-1} \Lambda^j U x_1^\dagger \\ &= \frac{u_{22}(u_{11}z_1 + u_{12}z_1^\dagger)}{\Delta} \lambda_1^j - \frac{u_{21}(u_{21}z_1 + u_{22}z_1^\dagger)}{\Delta} \lambda_2^j \\ &= \alpha \lambda_1^j - \beta \lambda_2^j, \end{aligned}$$

where $\Delta := |U| = u_{11}u_{22} - u_{12}u_{21}$ and $\alpha := u_{22}(u_{11}z_1 + u_{12}z_1^\dagger)/\Delta$, $\beta := u_{21}(u_{21}z_1 + u_{22}z_1^\dagger)/\Delta$. Considering that $u_{11}, u_{12}, u_{21}, u_{22}$ are real numbers, and the eigenvalues λ_1, λ_2 of the matrix U are compatible with $\{(u_{11} + u_{22}) \pm \sqrt{(u_{11} - u_{22})^2 + 4u_{12}u_{21}}\}/2$, the following hold:

- 1) if $\lambda_1 \notin \mathbb{R}$, we have $(\lambda_1, \lambda_2) = (\eta \exp(i\theta), \eta \exp(-i\theta))$ for some $\eta > 0$ and $\theta \in (0, 2\pi) \setminus \{\pi\}$.
- 2) if $\lambda_1 \in \mathbb{R}$, we have $(\lambda_1, \lambda_2) = (\eta_1, \eta_2)$ for some real numbers $\eta_1, \eta_2 \in \mathbb{R}$.

For the case 1), $\hat{z}_{j+1} = e_1^\top B^j x_1^\dagger = 2\alpha\eta^j \cos(j\theta) =: f_j^{(1)}(\alpha, \eta, \theta)$ as the imaginary part of $\alpha\lambda_1^j - \beta\lambda_2^j$ should be 0.

For the remaining case 2), $\hat{z}_{j+1} = e_1^\top B^j x_1^\dagger = \alpha\eta_1^j - \beta\eta_2^j =: f_j^{(2)}(\alpha, \eta, \beta)$. Considering both cases 1) and 2), the essential parameters in the ARS prediction \hat{z}_{j+1} are identifiable, i.e., (i) $f_j^{(1)}(\alpha, \eta, \theta) = f_j^{(1)}(\alpha', \eta', \theta')$ holds for all j if and only if $(\alpha, \eta, \theta) = (\alpha', \eta', \theta')$, (ii) $f_j^{(1)}(\alpha, \eta, \theta) = f_j^{(2)}(\alpha, \eta, \beta)$ does not hold for all j , (iii) $f_j^{(2)}(\alpha, \eta, \beta) = f_j^{(2)}(\alpha', \eta', \beta')$ holds for all j if and only if $(\alpha, \eta, \beta) = (\alpha', \eta', \beta')$.

Step 2: True parameters B^* , $\{z_j^{*\dagger}\}$ satisfy the equality (16). As the above Step 1 indicates that the ARS model is specified by identifiable few parameters, all the solutions satisfying the equality (16) corresponds to the same (identifiable) parameters. Therefore, the prediction \hat{z}_{n+k} is also uniquely determined and is compatible with the underlying true dynamical time series z_{n+k} . Therefore, the assertion is proved. \square

APPENDIX C PROOF OF PROPOSITION 2

As Equation (10) indicates that

$$x_2(t) = x_1(t) + \frac{1}{\alpha} \frac{dx_1(t)}{dt}, \tag{17}$$

we have

$$\frac{dx_2(t)}{dt} = \frac{dx_1(t)}{dt} + \frac{1}{\alpha} \frac{d^2x_1(t)}{dt^2}. \tag{18}$$

Substituting the identities (17) and (18) into both sides of the equation (11) yields

$$\begin{aligned} \frac{dx_1(t)}{dt} + \frac{1}{\alpha} \frac{d^2x_1(t)}{dt^2} &= -x_1(t)x_3(t) + \beta x_1(t) \\ &\quad - \left\{ x_1(t) + \frac{1}{\alpha} \frac{dx_1(t)}{dt} \right\}. \end{aligned}$$

Arranging the obtained terms proves

$$x_3(t) = -\frac{1}{x_1(t)} \left\{ \frac{1}{\alpha} \frac{d^2x_1(t)}{dt^2} + \frac{1 + \alpha}{\alpha} \frac{dx_1(t)}{dt} + (1 - \beta)x_1(t) \right\}. \tag{19}$$

This identity proves

$$\begin{aligned} \frac{dx_3(t)}{dt} &= \frac{1}{x_1(t)^2} \left\{ \frac{1}{\alpha} \frac{d^2x_1(t)}{dt^2} + \frac{1 + \alpha}{\alpha} \frac{dx_1(t)}{dt} + (1 - \beta)x_1(t) \right\} \\ &\quad - \frac{1}{x_1(t)} \left\{ \frac{1}{\alpha} \frac{d^3x_1(t)}{dt^3} + \frac{1 + \alpha}{\alpha} \frac{d^2x_1(t)}{dt^2} + (1 - \beta) \frac{dx_1(t)}{dt} \right\}. \end{aligned} \tag{20}$$

As the Lorenz equation (12) indicates with Equations (18) and (19) that

$$\begin{aligned} \frac{dx_3(t)}{dt} &= x_1(t)x_2(t) - \gamma x_3(t) \\ &= x_1(t) \left\{ \frac{dx_1(t)}{dt} + \frac{1}{\alpha} \frac{d^2x_1(t)}{dt^2} \right\} + \gamma \frac{1}{x_1(t)} \\ &\quad \times \left\{ \frac{1}{\alpha} \frac{d^2x_1(t)}{dt^2} + \frac{1 + \alpha}{\alpha} \frac{dx_1(t)}{dt} + (1 - \beta)x_1(t) \right\}. \end{aligned} \tag{21}$$

Comparing equations (20) and (21), where both are multiplied by $x_1(t)^2$, yields

$$\begin{aligned} &\left\{ \frac{1}{\alpha} \frac{d^2x_1(t)}{dt^2} + \frac{1 + \alpha}{\alpha} \frac{dx_1(t)}{dt} + (1 - \beta)x_1(t) \right\} \\ &\quad - x_1(t) \left\{ \frac{1}{\alpha} \frac{d^3x_1(t)}{dt^3} + \frac{1 + \alpha}{\alpha} \frac{d^2x_1(t)}{dt^2} + (1 - \beta) \frac{dx_1(t)}{dt} \right\} \\ &= x_1(t)^3 \left\{ \frac{dx_1(t)}{dt} + \frac{1}{\alpha} \frac{d^2x_1(t)}{dt^2} \right\} \\ &\quad + \gamma x_1(t) \left\{ \frac{1}{\alpha} \frac{d^2x_1(t)}{dt^2} + \frac{1 + \alpha}{\alpha} \frac{dx_1(t)}{dt} + (1 - \beta)x_1(t) \right\}. \end{aligned}$$

Therefore, rearranging the terms proves the assertion $\sum_{k=0}^3 P_k(x_1(t)) \frac{d^k x_1(t)}{dt^k} = 0$, where

$$\begin{aligned} P_0(x) &= (1 - \beta) - (1 - \beta)\gamma x, \\ P_1(x) &= \frac{1 + \alpha}{\alpha} - \left(\frac{1 + \alpha}{\alpha} \gamma - (1 - \beta) \right) x - x^3, \\ P_2(x) &= \frac{1}{\alpha} - \frac{1 + \alpha + \gamma}{\alpha} x - \frac{1}{\alpha} x^3, \\ P_3(x) &= -\frac{1}{\alpha} x. \end{aligned}$$

\square

ACKNOWLEDGMENT

The authors would like to thank the editor, the AE, and four anonymous reviewers, for constructive comments and suggestions. They also thank Kohei Hattori for helpful discussions. They would like to thank Sadayoshi Murakami, Masayuki Yokoyama, and Naoki Kenmochi, for helpful discussions in this project. They also thank the first director Kazuhei Kikuchi, who passed away young after the first meeting, in 2022. They pray for his soul to rest in peace.

REFERENCES

- [1] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Studies in Nonlinearity), 2nd ed. Boca Raton, FL, USA: CRC Press, 2001.
- [2] T. Jackson and A. Radunskaya, *Applications of Dynamical Systems in Biology and Medicine*, vol. 158. New York, NY, USA: Springer, 2015.
- [3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>
- [4] H. Jaeger, "The 'echo state': Approach to analysing and training recurrent neural networks," German Nat. Res. Inst. Comput. Sci., Bonn, Germany, Tech. Rep. 148, 2001.
- [5] H. Jaeger, "Adaptive nonlinear system identification with echo state networks," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2002, pp. 609–616.
- [6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
- [7] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: A review," *Neural Netw.*, vol. 115, pp. 100–123, Jul. 2019.
- [8] B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep learning for universal linear embeddings of nonlinear dynamics," *Nature Commun.*, vol. 9, no. 1, p. 4950, Nov. 2018.
- [9] D. Lucor, A. Agrawal, and A. Sergent, "Simple computational strategies for more effective physics-informed neural networks modeling of turbulent natural convection," *J. Comput. Phys.*, vol. 456, May 2022, Art. no. 111022.
- [10] S. Cheng, C. Quilodrán-Casas, S. Ouala, A. Farchi, C. Liu, P. Tando, R. Fablet, D. Lucor, B. Iooss, J. Brajard, D. Xiao, T. Janjic, W. Ding, Y. Guo, A. Carrassi, M. Bocquet, and R. Arcucci, "Machine learning with data assimilation and uncertainty quantification for dynamical systems: A review," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 6, pp. 1361–1387, Jun. 2023.
- [11] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Phys. Rev. Lett.*, vol. 45, no. 9, pp. 712–716, Sep. 1980.
- [12] K. T. Alligood, T. D. Sauer, J. A. Yorke, and D. Chillingworth, *Chaos: An Introduction To Dynamical Systems*. New York, NY, USA: Springer, 1997.
- [13] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence* (Lecture Notes in Mathematics). New York, NY, USA: Springer, 1981, pp. 366–381.
- [14] G. A. Gottwald and S. Reich, "Combining machine learning and data assimilation to forecast dynamical systems from noisy partial observations," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 31, no. 10, Oct. 2021, Art. no. 101103.
- [15] S. Ouala, D. Nguyen, L. Drumetz, B. Chapron, A. Pascual, F. Collard, L. Gaultier, and R. Fablet, "Learning latent dynamics for partially observed chaotic systems," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 30, no. 10, Oct. 2020, Art. no. 103121.
- [16] S. Ouala, S. L. Brunton, B. Chapron, A. Pascual, F. Collard, L. Gaultier, and R. Fablet, "Bounded nonlinear forecasts of partially observed geophysical systems with physics-constrained deep learning," *Phys. D, Nonlinear Phenomena*, vol. 446, Apr. 2023, Art. no. 133630.
- [17] T. Satheesh, R. Sakthivel, S. Harshavarthini, R. Manikandan, and D. J. Almkhles, "Antiwindup compensator-based control for periodic piecewise delayed systems with input saturations," *J. Dyn. Syst., Meas., Control*, vol. 145, no. 2, Feb. 2023, Art. no. 021003.
- [18] T. Satheesh, R. Sakthivel, N. Aravindh, and M. Chadli, "Design of proportional integral observer-based resilient control for periodic piecewise time-varying systems: The finite-time case," *Int. J. Robust Nonlinear Control*, vol. 34, no. 2, pp. 1169–1195, Jan. 2024.
- [19] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," in *Advances in Neural Information Processing Systems*, vol. 31. Red Hook, NY, USA: Curran Associates, 2018.
- [20] E. N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.*, vol. 20, no. 2, pp. 130–141, 1963.
- [21] T. C. Mills, *Time Series Techniques for Economists*. Cambridge, U.K.: Cambridge Univ. Press, 1990.
- [22] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1994.
- [23] K. Barański, Y. Gutman, and A. Śpiwak, "A probabilistic Takens theorem," *Nonlinearity*, vol. 33, no. 9, pp. 4940–4966, Sep. 2020.
- [24] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, document Version 20121115, Nov. 2012.
- [25] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960.
- [26] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *J. Basic Eng.*, vol. 83, no. 1, pp. 95–108, Mar. 1961.
- [27] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Statist.*, vol. 37, no. 6, pp. 1554–1563, Dec. 1966.
- [28] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, USA, Tech. Rep. CRG-TR-96-2, Feb. 1996.
- [29] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 2, pp. 260–269, Apr. 1967.
- [30] M. Beal, Z. Ghahramani, and C. Rasmussen, "The infinite hidden Markov model," in *Advances in Neural Information Processing Systems*, vol. 14. Cambridge, MA, USA: MIT Press, 2001.
- [31] H. Jeffreys, "'Weierstrass's theorem on approximation by polynomials' and 'extension of Weierstrass's approximation theory,'" in *Methods of Mathematical Physics*. Hoboken, NJ, USA: Wiley, 1988, pp. 446–448.



AKIFUMI OKUNO received the Ph.D. degree in informatics from Kyoto University, Japan, in 2020. He is currently an Assistant Professor with The Institute of Statistical Mathematics. His research interests include mathematical statistics, statistical machine learning, and their applications to science.



YUYA MORISHITA received the Ph.D. degree in engineering from Kyoto University, Japan, in 2023. He is currently an Assistant Professor with the Department of Nuclear Engineering, Kyoto University. His research interests include computational modeling and control of fusion plasmas.



YOH-ICHI MOTOTAKE received the Ph.D. degree from The University of Tokyo, Japan, in 2016. He is currently an Associate Professor with the Department of Social Data Science, Hitotsubashi University. His research interests include physics in pattern dynamics, machine learning, and interpretable AI.

...