Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 2389–2394

Data-Enabled Identification of Nonlinear Dynamics of Water Systems using Sparse Regression Technique ⋆

Faegheh Moazeni ∗ **and Javad Khazaei** ∗

∗ *Lehigh University, Bethlehem, PA 18015 USA (e-mails: moazeni@lehigh.edu, khazaei@lehigh.edu).*

Abstract: The complex, multi-variable, highly nonlinear and strong coupling characteristics of water distribution systems (WDSs) has significantly limited the capability of model-based approaches for control purposes in such systems. With the emerging application of high-resolution metering devices and historical data, model-free identification of WDSs can facilitate the control design without tedious modeling complexities. This paper develops a data-driven framework to facilitate the identification of nonlinear models of WDSs using available data. A quadruple tank system that represents the nonlinear and strong coupling nature of WDSs is considered as the test system and sparse identification of nonlinear dynamics (SINDy) is utilized to identify the nonlinear dynamics from the data. Unlike existing modeling approaches that either heavily rely on knowing the detailed dynamics of the system (model-based) or designs that relay on large historical data and are not interpretable (data-driven approaches), the proposed model-free identification framework is parsimonious, which can accurately capture the dynamics of the quadruple tank process with available measurements suitable for control problems. The effectiveness of the proposed approach is validated using time-domain simulations in MATLAB.

Keywords: Water Distribution Systems, Sparse Identification of Nonlinear Dynamics (SINDy), Data-driven System Identification, Quadruple Tank Process.

1. INTRODUCTION

Dynamics of the water distribution system (WDS) is of growing importance due to role of modeling in efficiently controlling the water system assets and responding to the increased demand for clean water Zhang et al. (2020). Existing practices for modeling WDSs mainly focus on physics-based modeling based on governing physical laws (i.e., hydraulic laws and mass balance) in these systems Abdelbaki et al. (2017); Kara et al. (2016); Zamenian et al. (2017); Creaco et al. (2017). For example, a combined modeling technique was proposed in Abdelbaki et al. (2017) to utilize the map information from the GIS data in the hydraulic modeling for water management in WDSs, or in Zamenian et al. (2017), a random-parameter negative-binomial approach was utilized along with the WDS model to estimate a system-wide monthly frequency of water main breaks in WDSs. However, due to multivariate, nonlinear, complex, and highly coupled dynamics of assets in WDS, modeling can be challenging Tung et al. (2020).

The increasing availability of data from improved monitoring technologies such as smart metering devices and wireless sensors provide an opportunity to discover enhanced models for WDSs that could account for nonlinearities and complex nature of these systems. Several studies have focused on data-driven modeling for water distribution assets or treatment processes Seo et al. (2015); Liu et al. (2016); Seyedzadeh et al. (2020); Zhang et al. (2018); Ahmed et al. (2019). For example, in Seo et al. (2015), a wavelet-based artificial neural network approach was proposed to forecast the water level on a daily basis for WDSs. In Liu et al. (2016), the water quality forecast on an hourly basis was carried out using a multitask multi-view learning methods to capture the spatio-temporal and spatial correlations between water quality treatment facilities. Other existing data-driven approaches include estimating the discharge of drip tape irrigation using temperature and pressure measurements Seyedzadeh et al. (2020), identifying the reservoir operation using artificial neural network and support vector regression Zhang et al. (2018), predicting the water quality using various machine learning methods Ahmed et al. (2019), and reservoir water inflow forecast using neural networks Yang et al. (2017). While the existing research shows the significant potential of data-driven approaches for predicting the behavior of water system assets or the quality of water, none of the existing studies have focused on utilizing data-driven techniques for identifying nonlinear dynamics of WDS at a system level. In addition, most existing studies focus on machine learning-based methods, which heavily rely on training data and are not interpretable. Such models would require a re-training with system expansions and might not be suitable for system-level control in WDS.

To address the existing challenges of the data-driven approaches for identifying the complex dynamics in

⋆ This research was in part under support from the National Science Foundation under Grant NSF-EPCN 2221784.

10.1016/j.ifacol.2023.10.1212

water distribution systems, this paper investigates the application of sparse identification theory. Sparse regression techniques have demonstrated great potential for precisely modeling the nonlinear dynamics of unknown systems Brunton et al. (2016a,b); Khazaei and Blum (2022), However, there is currently no reported data-driven modeling of the nonlinear dynamics of WDSs using sparsity promoting techniques. The identification of WDS is challenging due to its complex, highly nonlinear and strong coupling characteristics. The novelty of this paper lies in investigating the potential usage of sparse regression in such systems. The main contribution of this work is to identifying nonlinear dynamics of quadruple water tank process that closely represent the nonlinear dynamics of WDSs using data-driven sparse regression technique. In addition, to deal with noisy measurements, this paper proposes the use of data denoising via Savitzky Golay filtering approach that simplifies the SINDy process and has not been reported in any work. Using time-domain simulations, we will validate the effectiveness of the proposed data-driven approach in accurately identifying the dynamics of the quadruple tank system using available measurements.

The rest of the paper is organized as follows: Section II formulates the dynamics of multi-tank system. Sparse identification of dynamics is included in Section III. Time-domain simulations are included in Section IV and Section V concludes the paper.

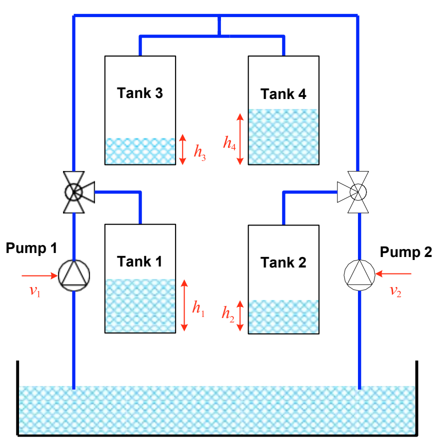

Fig. 1. Schematic of a quadruple tank process.

1.1 Dynamic Model of Quadruple Tank Process

The quadruple water tank process in Fig. 1 can be represented by the following set of differential equations Johansson (2000):

$$\frac{dh_1}{dt} = -\frac{a_1}{A_1}\sqrt{2gh_1} + \frac{a_3}{A_1}\sqrt{2gh_3} + \frac{\gamma_1 k_1}{A_1}v_1 \quad (1)$$

$$\frac{dh_2}{dt} = -\frac{a_2}{A_2}\sqrt{2gh_2} + \frac{a_4}{A_2}\sqrt{2gh_4} + \frac{\gamma_2 k_2}{A_2}v_2 \quad (2)$$

$$\frac{dh_3}{dt} = -\frac{a_3}{A_3}\sqrt{2gh_3} + \frac{(1-\gamma_2)k_2}{A_3}v_2 \quad (3)$$

$$\frac{dh_4}{dt} = -\frac{a_4}{A_4}\sqrt{2gh_4} + \frac{(1-\gamma_1)k_1}{A_4}v_1 \quad (4)$$

where A_i is the cross-sectional area of tank i, a_i is a cross-sectional area of the outlet hole in tank i, h_i is the water level at tank i, v_i is the voltage applied to pump i with a corresponding flow $k_i v_i$. Parameters $\gamma_i \in [0, 1]$ are determined from the settings of the valves. The water flow to tank 1 is $\gamma_1 k_1 v_1$, and the water flow to tank 4 is $(1-\gamma_1)k_1 v_1$. Similarly, the water flow to tank 2 is $\gamma_2 k_2 v_2$, and the water flow to tank 3 is $(1-\gamma_2)k_2 v_2$. Furthermore, the acceleration of gravity is denoted by g. By re-arranging equation (1) and writing the model in state-space form,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}, \quad \mathbf{y} = \mathbf{h}(\mathbf{x}) \quad (5)$$

where

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} -\frac{a_1}{A_1}\sqrt{2g}\sqrt{x_1} + \frac{a_3}{A_1}\sqrt{2g}\sqrt{x_3} \\ -\frac{a_2}{A_2}\sqrt{2g}\sqrt{x_2} + \frac{a_4}{A_2}\sqrt{2g}\sqrt{x_4} \\ -\frac{a_3}{A_3}\sqrt{2g}\sqrt{x_3} \\ -\frac{a_4}{A_4}\sqrt{2g}\sqrt{x_4} \end{bmatrix} \quad (6)$$

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} \frac{\gamma_1 k_1}{A_1} & 0 \\ 0 & \frac{\gamma_2 k_2}{A_2} \\ 0 & \frac{(1-\gamma_2)k_2}{A_3} \\ \frac{(1-\gamma_1)k_1}{A_4} & 0 \end{bmatrix}, \quad \mathbf{h}(\mathbf{x}) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad (7)$$

In equations (5)-(7), $\mathbf{x} = [h_1\ h_2\ h_3\ h_4]^T$ is the state vector, $\mathbf{y} = \mathbf{x}$ is the output vector, and $\mathbf{u} = [v_1\ v_2]^T$ is the input vector of the system, which includes the voltage applied to the pumps.

1.2 Operating Points

Assuming detailed information about Quadruple tank process system and its parameters are not available, the objective is to identify the dynamic model presented in equations (5)-(7) from available measurements of the states. Two separate operating points will be studied to validate the effectiveness of the proposed model identification framework. These two different operating points were selected specifically as the system shows a minimum phase (MP) characteristics in one of the operating points and non-minimum phase (NMP) behavior at the other one. These two operating points are listed in the following table.

Table 1. Operating points of quadruple tank.

Parameter	MP operating point	NMP operating point
h_1, h_2 [cm]	12.4, 12.7	12.6, 13
h_3, h_4 [cm]	1.8, 1.4	4.8, 4.9
v_1, v_2 [V]	3, 3	3.15, 3.15
k_1, k_2	3.33, 3.35	3.14, 3.29
γ_1, γ_2	0.7, 0.6	0.43, 0.34

Simulations were carried out to depict the operation of the system in these two operating points. Fig. 3 depicts the trajectory of the states at two operating points.

2. DATA-DRIVEN IDENTIFICATION OF DYNAMICS

We will utilize sparse identification of nonlinear dynamics to identify equations (5)-(7) from measurements. An

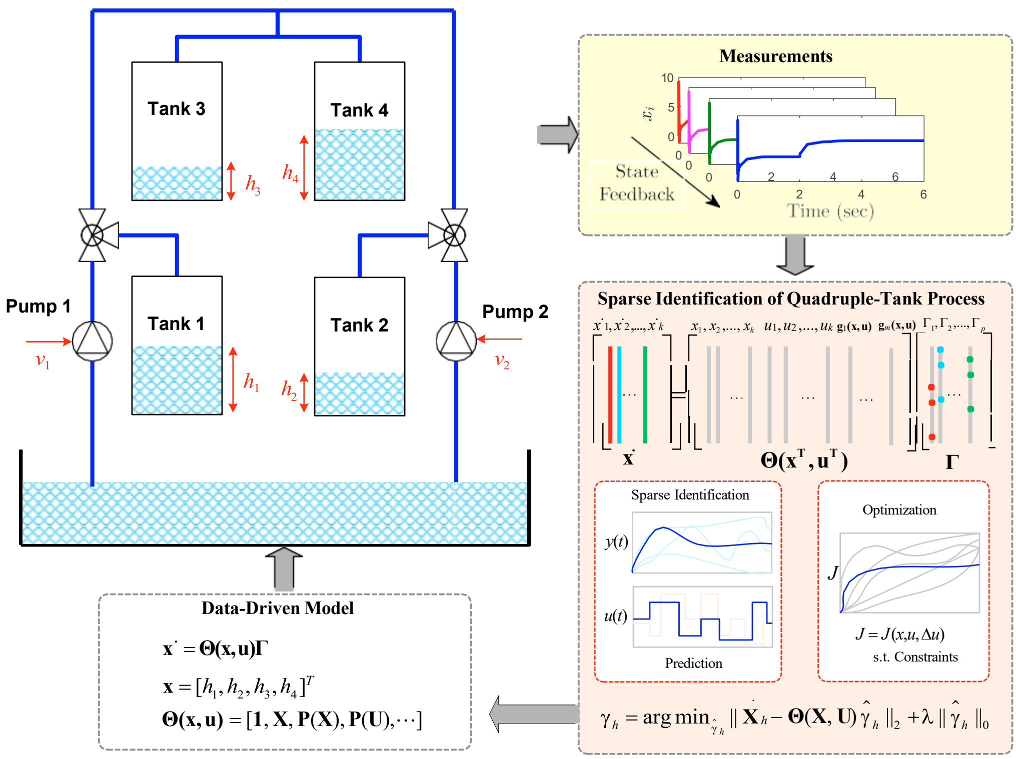

Fig. 2. Proposed data-driven framework for identifying the dynamics of a quadruple tank process.

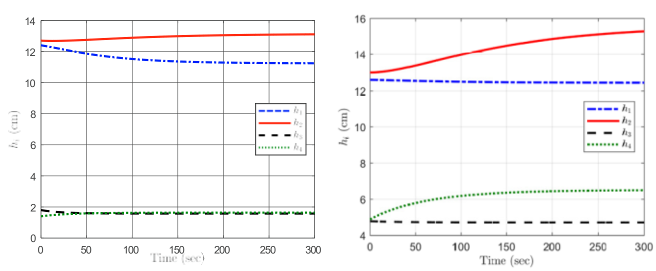

Fig. 3. Trajectory of the states in MP operating point (Left), and trajectory of the states in NMP operating point (right).

overview of the proposed data-driven model identification approach for the quadruple tank process is depicted in Fig. 2. By taking the measurements from the water tank levels, h_i, and the inputs applied to pumps, v_i, and constructing a library of functions that could represent the quadruple tank dynamics, a sequentially thresholded least-square optimization problem is solved to obtain the nonlinear dynamics using sparse regression technique. The process is explained in detail in the following.

To identify nonlinear dynamics of the dynamical systems from measurements, the first step is to estimate the derivatives of the states ($\dot{x}$) from measurements and then construct a library of candidate functions ($\theta_i(x)$) that describe how state variables vary with time. If no information about the dynamics of the system is known, an extended basis of candidate functions can be selected to cover all possible functions. Since most dynamical systems have few nonlinear terms in the dynamics, sparsity promoting techniques can identify the candidate functions with a major impact on forming the system dynamics.

This method is known as sparse identification of nonlinear dynamics (SINDy), which was originally proposed in Brunton et al. (2016a), and is explained step-by-step in the following.

2.1 Measurements

SINDy uses symbolic regression and sparse representations to determine the dynamics of the system. This approach relies on the fact that most dynamical systems that are represented by differential equations to the form $\dot{x} = f(x, u)$ have a relatively few terms on the right hand side. The actual dynamics of a quadruple tank process is represented by $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$, where $\mathbf{x}(t) \in \mathbb{R}^n$ is the state vector, $\mathbf{u}(t) \in \mathbb{R}^q$ is the input or control vector, and $\mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)) : \mathbb{R}^n \times \mathbb{R}^q \rightarrow \mathbb{R}^n$. By collecting m measurement samples from the water tank levels and pump inputs, the quadruple tank process dynamics can be identified by a library of candidate functions, $\mathbf{\Theta} \in \mathbb{R}^{m \times p}$. To identify the governing equations of the system in (5)-(7), we collect a time-history of the tank levels (state vector) $\mathbf{x}(t)$, pump inputs $\mathbf{u}(t)$, and derivatives of the states $\dot{\mathbf{x}}(t)$. Since only $\mathbf{x}(t)$ and $\mathbf{u}(t)$ might be available in most real-world systems, the derivative measurements $\dot{\mathbf{x}}(t)$ must be estimated first. This can be done through numerical derivative calculations from measurements of the states. To do so, first, the measurement data is sampled at m intervals $t_1, t_2, \ldots, t_m$ and arranged into:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^T(t_1) \\ \mathbf{x}^T(t_2) \\ \vdots \\ \mathbf{x}^T(t_m) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \ldots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \ldots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_m) & x_2(t_m) & \ldots & x_n(t_m) \end{bmatrix} \quad (8)$$

and inputs for t_m samples are written into a matrix $\mathbf{U}$,

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}^T(t_1) \\ \mathbf{u}^T(t_2) \\ \vdots \\ \mathbf{u}^T(t_m) \end{bmatrix} = \begin{bmatrix} u_1(t_1) & u_2(t_1) & \dots & u_n(t_1) \\ u_1(t_2) & u_2(t_2) & \dots & u_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ u_1(t_m) & u_2(t_m) & \dots & u_n(t_m) \end{bmatrix} \quad (9)$$

the measurements for derivatives can be approximated numerically from $\mathbf{X}$ following the procedure.

2.2 Estimating the Derivatives, $\dot{\mathbf{X}}$

By using difference approximation, ordinary differential equations and partial differential equations can be numerically solved. At specified mesh points, the derivatives of a smooth function can be approximated using Taylor series expansion. This paper uses the central difference approximation because it is more accurate for smooth functions. Consequently, $\dot{\mathbf{X}}$ can be approximated by Larsson and Thom´ee (2003):

$$\dot{\mathbf{X}} \approx \frac{\mathbf{X}(i+1) - \mathbf{X}(i-1)}{2h} \quad (10)$$

where $\mathbf{X}(i+1)$ is the measured data at sample $i+1$ and h is the sampling time of the simulation or data collection platform. Utilizing the measurements from the states and

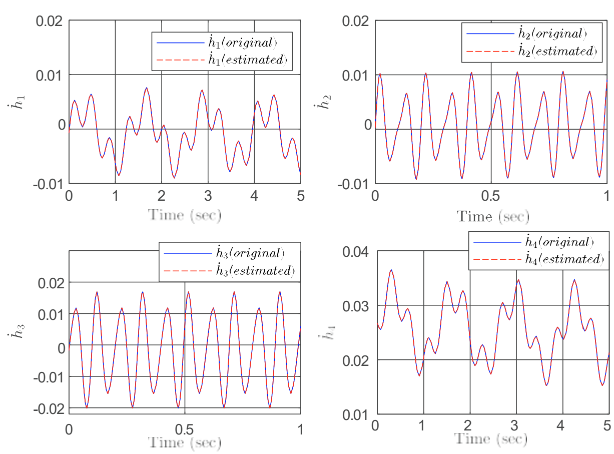

Fig. 4. Derivative estimation for MP operation point.

perturbing the inputs (pump voltages, v_i) with a sinusoidal signal (to excite the dynamic modes), the results of the estimated derivatives for minimum phase (MP) operating points of the quadruple tank process are shown in Fig. 4. As it can be seen, the estimated derivative can accurately represent the measured derivative.

2.3 Sparse Identification of System Dynamics

By utilizing the measured data $\mathbf{X} \in \mathbb{R}^{m\times n}$ to obtain the derivatives of the states, the derivative data is a linear combination of columns from the candidate function (e.g., polynomials, or sinusoids) library expressed by entries of the matrix $\Xi \in \mathbb{R}^{p\times n}$ such that Brunton et al. (2016a):

$$\dot{\mathbf{X}} = \mathbf{\Theta}(\mathbf{X}, \mathbf{U})\mathbf{\Xi}. \quad (11)$$

Having estimated $\dot{\mathbf{X}}$, $\mathbf{\Theta}(\mathbf{X}, \mathbf{U})$ can be constructed by linear and nonlinear functions of the columns of $\mathbf{X}$ and $\mathbf{U}$. Typically, the candidate functions include monomials and trigonometric functions for nonlinear systems. An example of such function is represented in equation (12), where $\mathbf{M_i}(\mathbf{X}, \mathbf{U})$ denotes a nonlinear combination of the i-order monomials of $\mathbf{X}$ and $\mathbf{U}$. For example, $\mathbf{M_{0.5}}(\mathbf{X}, \mathbf{U})$ includes square-root functions that exist in the quadruple tank system, or $\mathbf{P_2}(\mathbf{X}, \mathbf{U})$ involves polynomials up to the second order. Having estimated $\dot{\mathbf{X}}$ and finding $\mathbf{\Theta}(\mathbf{X}, \mathbf{U})$ using available $\mathbf{X}, \mathbf{U}$, we can solve for $\dot{\mathbf{X}} = \mathbf{\Theta}(\mathbf{X}, \mathbf{U})\mathbf{\Xi}$ by solving for the sparse vectors of coefficients in $\mathbf{\Xi}$ that decide what terms are active in the $\dot{\mathbf{X}}$ dynamics. This is achieved by solving an optimization of the form:

$$\boldsymbol{\xi}_h = \arg\min_{\hat{\boldsymbol{\xi}}_h} ||\dot{\mathbf{X}}_h - \mathbf{\Theta}(\mathbf{X}, \mathbf{U})\hat{\boldsymbol{\xi}}_h||_2 + \lambda||\hat{\boldsymbol{\xi}}_h||_0 \quad (13)$$

where $\boldsymbol{\xi}_h$ is the h-th column of $\boldsymbol{\xi}$ represented by $\boldsymbol{\xi}_h = [\xi_1\ \xi_2\ \dots\ \xi_p]^T$ and $\dot{\mathbf{X}}_h$ represents the h-th column of $\dot{\mathbf{X}}$. The objective function in (13) utilizes the L2 norm $||.||_2$ to minimize the error between the derivatives $\dot{\mathbf{X}}$ and estimated derivatives using calculated $\boldsymbol{\xi}$ through a least-squares problem and the L0 norm, $||.||_0$ minimizes the number of nonzero elements in $\boldsymbol{\xi}_h$ to promote sparsity in the coefficients matrix $\boldsymbol{\xi}$. In addition, λ is the sparsity-promoting hyperparameter.

The minimization problem of (13) is solved by the sequentially thresholded least squares method, which is an iterative algorithm defined by Zhang and Schaeffer (2019):

$$S^k = \left\{ j \in [p] : |\xi_j^k| \geq \lambda \right\}, \quad k \geq 0 \quad (14)$$

$$\hat{\boldsymbol{\xi}}_h^0 = \mathbf{\Theta}(\mathbf{X}, \mathbf{U})^\dagger \dot{\mathbf{X}}_h \quad (15)$$

$$\boldsymbol{\xi}^{k+1} = \underset{\hat{\boldsymbol{\xi}}_h \in \mathbb{R}^p : \mathrm{supp}(\boldsymbol{\xi}_h) \subseteq S^k}{\operatorname{argmin}} \|\dot{\mathbf{X}}_h - \mathbf{\Theta}(\mathbf{X}, \mathbf{U})\hat{\boldsymbol{\xi}}_h\|_2, \quad (16)$$

where k is the iteration number, $\mathbf{\Theta}(\mathbf{X}, \mathbf{U})^\dagger$ is the pseudo inverse of $\mathbf{\Theta}(\mathbf{X}, \mathbf{U})$ defined as:

$$\mathbf{\Theta}(\mathbf{X}, \mathbf{U})^\dagger := [\mathbf{\Theta}(\mathbf{X}, \mathbf{U})^T \mathbf{\Theta}(\mathbf{X}, \mathbf{U})]^{-1} \mathbf{\Theta}(\mathbf{X}, \mathbf{U})^T \quad (17)$$

and the support set of $\boldsymbol{\xi}_h$ is defined by $\mathrm{supp}(\boldsymbol{\xi}_h) := \{ j \in [p] : \xi_j = 0 \}$. The coefficients $\boldsymbol{\xi}_h$ can be found using the sparse regression formulation presented in **Algorithm 1**. If the intent is to identify the signal $\mathbf{U}$ for feedback control, i.e., $\mathbf{U} = G(s)\mathbf{X}$, where $G(s)$ is the transfer function of the controller, the matrix of inputs can be identified using $\mathbf{U} = \Theta(\mathbf{X})\mathbf{\Gamma_u}$, where $\Theta(\mathbf{X})$ is the matrix of candidate functions with the terms corresponding to $\mathbf{U}$ have been removed from $\Theta(\mathbf{X}, \mathbf{U})$ and $\mathbf{\Gamma_u}$ can be found using the sparse regression algorithm similar to $\mathbf{\Xi}$.

Algorithm 1 Sparse Regression Algorithm

Input: Measurements $\mathbf{X}, \mathbf{U}$
Input: Estimated derivatives $\dot{\mathbf{X}}$
procedure Sparsity Promoting Algorithm
$\Gamma = \Theta \backslash \dot{\mathbf{X}}$ (least-square solution)
for k = 1 : 10 **do** (number of iterations)
Set λ (sparsity knob)
$|\Xi| < \lambda \longrightarrow ind_{small}$
$\Xi(ind_{small}) \longrightarrow 0$
for k = 1 : n **do** (n dimension of state $\mathbf{X}$)
$ind_{big} = \sim ind_{small}(:, k)$
$\Xi(ind_{big}, k) = \Theta(:, ind_{big}) \backslash \dot{\mathbf{X}}(:, k)$
end for
end for
Output: sparse matrix Ξ

$$\Theta(\mathbf{X},\mathbf{U}) = \left[\mathbf{M}_{0.5}(\mathbf{X},\mathbf{U})\ \ \mathbf{X}\ \ \mathbf{U}\ \ \mathbf{M}_2(\mathbf{X},\mathbf{U})\ \dots\ \sin(\mathbf{X},\mathbf{U})\ \ \cos(\mathbf{X},\mathbf{U})\ \ \sin(\mathbf{2}(\mathbf{X},\mathbf{U}))\ \dots \right] \quad (12)$$

Table 2. Parameter identification using SINDy.

Dynamics	Term	Term	Term	Term	Term	Term
$\dot{x}_1$	$-\frac{a_1\sqrt{2g}}{A_1}\sqrt{x_1}$	$0\sqrt{x_2}$	$\frac{a_3\sqrt{2g}}{A_1}\sqrt{x_3}$	$0\sqrt{x_4}$	$\frac{\gamma_1 k_1}{A_1}v_1$	$0v_2$
Physical	−0.112	0	0.112	0	0.118	0
Identified	−0.111	0	0.110	0	0.118	0
$\dot{x}_2$	$0\sqrt{x_1}$	$-\frac{a_2\sqrt{2g}}{A_2}\sqrt{x_2}$	$0\sqrt{x_3}$	$\frac{a_4\sqrt{2g}}{A_2}\sqrt{x_4}$	$0v_1$	$\frac{\gamma_2 k_2}{A_2}v_2$
Physical	0	−0.078	0	0.078	0	0.104
Identified	0	-0.073	0	0.074	0	0.1
$\dot{x}_3$	$0\sqrt{x_1}$	$0\sqrt{x_2}$	$-\frac{a_3\sqrt{2g}}{A_3}\sqrt{x_3}$	$0\sqrt{x_4}$	$0v_1$	$\frac{(1-\gamma_2)k_2}{A_3}v_2$
Physical	0	0	−0.112	0	0	0.119
Identified	0	0	−0.110	0	0	0.122
$\dot{x}_4$	$0\sqrt{x_1}$	$0\sqrt{x_2}$	$0\sqrt{x_3}$	$-\frac{a_4\sqrt{2g}}{A_4}\sqrt{x_4}$	$\frac{(1-\gamma_1)k_1}{A_4}v_1$	$0v_2$
Physical	0	0	0	0	-0.078	0.118
Identified	0	0	0	0	-0.074	0.123

3. TIME-DOMAIN SIMULATIONS

To validate the effectiveness of the proposed data-enabled model-free identification of quadruple tank process dynamics, time-domain simulations using MATLAB was carried out and the states were observed by perturbing the inputs of the pumps. The models were run for 300 seconds with 50μsec sampling time and the training data set includes a state matrix $\mathbf{X} \in \mathbb{R}^{4\times6000}$ and input matrix size of $\mathbf{U} \in \mathbb{R}^{2\times6000}$. The Hyperparameter λ was tuned manually and chosen as 0.01 for sparse model identification. The proposed framework has also been validated with comparisons with physical models.

3.1 Model Identification

First, data was collected on states and input of a physical quadruple tank system that was simulated in MATLAB using the parameters provided in Johansson (2000). Since it is assumed that the original model of the system is not known, the candidate terms for $\Theta(\mathbf{X}, \mathbf{U})$ included square-root functions, polynomials up to degree 2, and sinusoidal functions, i.e., u_i, x_i, $\sqrt{x_i}$, $x_i x_j$, x_i^2, $x_i \cos x_j$, $x_i \sin x_j$, $u_i \cos x_j$, $u_i \sin x_j$, respectively.
Sparse regression was then carried out to identify the sparse matrix of coefficients, $\boldsymbol{\Xi}$. The identified $\boldsymbol{\Xi}$ for the studied model were used to develop a data-driven quadruple tank model in MATLAB. A comparison between the parameters of the physical model and the identified model is shown in Table. 2, which confirms the identified data-driven model accurately represents the dynamics of the physical model.

3.2 Time-domain Validation

In the second case study, the obtained SINDy model of the quadruple tank process was compared with the physical model in several scenarios. A step change from 3V to 4V was applied to the pump 1's voltage v_1 at 25 seconds and another step change was applied at 100 seconds to reduce v_1 from 4 to 2 V. In addition, a step change was applied to the pump 2's voltage v_2 at 50 seconds to increase it from 3 to 4V and another step change was applied at 150 seconds to reduce v_1 from 4 to 2 V. The physical system was compared with the identified data-driven model using SINDy as shown in Fig. 5. The results indicate the effectiveness of the proposed data-driven model identification approach (SINDy) for accurately identifying the nonlinear dynamics of quadruple tank process system. Such model identification can significantly simplify the complex physics-based modeling of large-scale water distribution systems.

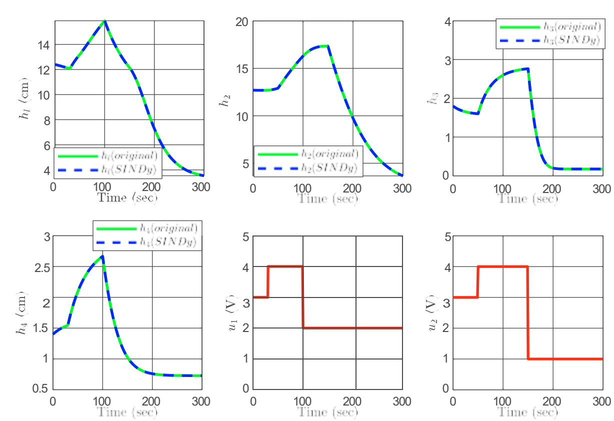

Fig. 5. Time-domain validation of the identified model.

3.3 Noisy Measurements

In this case, the impact of measurement noise on dynamics identification is studied. The state measurements were augmented with Gaussian-distributed random noise, with mean of 0 and variance of 0.1. The collected noisy data was then utilized in the identification process. In order to minimize the impact of measurement noise on the identification process, a filtering process using Savitzky

Golay filter Savitzky and Golay (1964) was first carried out. The physical noisy system was compared with the identified data-driven model using SINDy as shown in Fig. 6. The results exhibit the effectiveness and accuracy of the proposed approach for identifying the nonlinear dynamics of quadruple tank system from noisy measurements.

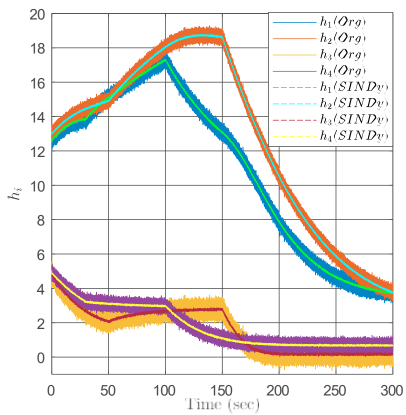

Fig. 6. The identified model with noisy measurement.

4. CONCLUSION

In this paper, a model-free data-driven identification of dynamics was proposed for a quadruple tank process that incorporates the nonlinear behavior of water distribution networks. Using sparse identification of nonlinear dynamics with control and utilizing the available measurements, we predicted the nonlinear dynamics of the multi-tank system was through a library of candidate functions. The learned dynamics demonstrated the effectiveness of the sparse identification theory for data-driven model identification of nonlinear water distribution systems. Time-domain simulations validated the close tracking of system states using the data-driven model. Future research will focus on expanding the analysis to a large-scale water distribution system model identification using sparse regression.

REFERENCES

Abdelbaki, C., Benchaib, M.M., Benziada, S., Mahmoudi, H., and Goosen, M. (2017). Management of a water distribution network by coupling gis and hydraulic modeling: a case study of chetouane in algeria. *Applied Water Science*, 7(3), 1561–1567.

Ahmed, A.N., Othman, F.B., Afan, H.A., Ibrahim, R.K., Fai, C.M., Hossain, M.S., Ehteram, M., and Elshafie, A. (2019). Machine learning methods for better water quality prediction. *Journal of Hydrology*, 578, 124084.

Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016a). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15), 3932–3937.

Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016b). Sparse identification of nonlinear dynamics with control (sindyc). *IFAC-PapersOnLine*, 49(18), 710–715.

Creaco, E., Campisano, A., Franchini, M., Modica, C., et al. (2017). Unsteady flow modeling of pressure real-time control in water distribution networks. *Journal of Water Resources Planning and Management*, 143(9), 04017056.

Johansson, K.H. (2000). The quadruple-tank process: A multivariable laboratory process with an adjustable zero. *IEEE Transactions on control systems technology*, 8(3), 456–465.

Kara, S., Karadirek, I.E., Muhammetoglu, A., and Muhammetoglu, H. (2016). Hydraulic modeling of a water distribution network in a tourism area with highly varying characteristics. *Procedia Engineering*, 162, 521–529.

Khazaei, J. and Blum, R.S. (2022). Model-free distributed control of dynamical systems. *International Journal of Information and Communication Engineering*, 16(10), 475–480.

Larsson, S. and Thom´ee, V. (2003). *Partial differential equations with numerical methods*, volume 45. Springer.

Liu, Y., Zheng, Y., Liang, Y., Liu, S., and Rosenblum, D.S. (2016). Urban water quality prediction based on multi-task multi-view learning. In *Proceedings of the 25th international joint conference on artificial intelligence*.

Savitzky, A. and Golay, M.J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8), 1627–1639.

Seo, Y., Kim, S., Kisi, O., and Singh, V.P. (2015). Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology*, 520, 224–243.

Seyedzadeh, A., Maroufpoor, S., Maroufpoor, E., Shiri, J., Bozorg-Haddad, O., and Gavazi, F. (2020). Artificial intelligence approach to estimate discharge of drip tape irrigation based on temperature and pressure. *Agricultural Water Management*, 228, 105905.

Tung, T.M., Yaseen, Z.M., et al. (2020). A survey on river water quality modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology*, 585, 124670.

Yang, T., Asanjan, A.A., Welles, E., Gao, X., Sorooshian, S., and Liu, X. (2017). Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. *Water Resources Research*, 53(4), 2786–2812.

Zamenian, H., Mannering, F.L., Abraham, D.M., and Iseley, T. (2017). Modeling the frequency of water main breaks in water distribution systems: Random-parameters negative-binomial approach. *Journal of Infrastructure Systems*, 23(2), 04016035.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X., and Zhuang, J. (2018). Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology*, 565, 720–736.

Zhang, L. and Schaeffer, H. (2019). On the convergence of the sindy algorithm. *Multiscale Modeling & Simulation*, 17(3), 948–972. doi:10.1137/18M1189828.

Zhang, Q., Zheng, F., Chen, Q., Kapelan, Z., Diao, K., Zhang, K., and Huang, Y. (2020). Improving the resilience of postdisaster water distribution systems using dynamic optimization framework. *Journal of Water Resources Planning and Management*, 146(2), 04019075.