DATA-DRIVEN CONTROL OF THE CHEMOSTAT USING THE KOOPMAN OPERATOR THEORY

Benaissa DEKHICI[1*], Boumediene BENYAHIA[1], Brahim CHERKI[1]

The chemostat is widely used as a laboratory pilot for bioprocess studies. Chemostat models are nonlinear and rarely used in modern control experiments. For a data-driven control strategy, we use the Koopman operator approach to derive a linear model for a simple chemostat with one substrate and one biomass, using only the chemostat's input-output data. For chemostat control, we use the linear Koopman model to develop a MPC controller. The linear Koopman model best fits chemostat data compared to the local linearization-based model. In addition, the MPC based on the Koopman model gives very satisfying results compard with a linear MPC controller when applied to control the chemostat. The results are gained for a large space of initial conditions when chemostat control is usually limited.

Keywords: Chemostat, Model predictive control, Data-driven control design, Linear model, Koopman operator theory.

1. Introduction

The chemostat is a laboratory instrument used to study microbial ecology, develop microbial species in a controlled environment, and emulate bioprocesses like the Continuously Stirred Tank Reactor (CSTR) used for wastewater treatment [1]. The mass balance rule is used to describe biological responses between substrates and biomasses in a chemostat. In the literature, the simplest chemostat model with one substrate and one biomass is well known [1, 2]. It is updated to use existing control methods based on the input substrate concentration and the dilution rate [3, 4]. Models of bioprocesses especially of the chemostat remain strongly nonlinear and uncertain, which hinders experimental implementation of current controllers.

Data-driven modeling and control solutions are prevalent in engineering, biology, and physics. These techniques, especially those based on the Koopman operator theory [5], avoid modeling uncertainty and high nonlinearity, and provide linear models for dynamical systems. The infinite-dimensional Koopman operator captures nonlinear dynamic systems' activities. It projects a finite-dimensional nonlinear system onto an infinite-dimensional linear space, from which the linear model can be recovered. In recent years, many scientific

[1] Automatic Laboratory of Tlemcen, Faculty of technology, University of Tlemcen, Algeria, *e-mail: benaissa.dekhici@univ-tlemcen.dz

research on the Koopman operator coupled with Dynamic Mode Decomposition method have been published [6, 7, 8].

In this paper, we adopt a Koopman operator extension to control a chemostat with one substrate and one biomass. We employ the chemostat's nonlinear model as a virtual system on which we apply control sequences to generate data (just the input-output data), from which we deduce a linear model using the Extended Dynamic Mode Decomposition (EDMD); a data-driven approximation of the Koopman operator [9, 10]. To highlight the capability of the linear Koopman model, we use it to perform a Model Predictive Control (MPC) [11] for the initial system (chemostat) , and we compare results with other conventional models like the local linearization-based model. We control the chemostat in the large space of the initial conditions which is the novelty of this work. We will be able to impose linear inequality constraints on state and control inputs, as well as nonlinear state constraints, using the proposed Koopman MPC framework. The suggested technique, which avoids the difficult and complex solution of non-convex optimization issues in standard nonlinear MPC schemes [12], may be simply applied for predictive control of the nonlinear chemostat system using improved and very efficient linear MPC solvers.

The paper is organized as follows: In Section 2, we outline the nonlinear system (chemostat) that we are interested in. The Koopman model of the chemostat is then presented in Section 3 for both full-state and input-output measurements. Section 4 then builds a linear Koopman-MPC controller for chemostat's input-output data. Finally, in Section 5, simulation findings are displayed and analyzed before some conclusions and views are reached.

2. **Nonlinear dynamical model of the chemostat**

The chemostat is a small bioreactor used in the study of the growth of microorganisms or the emulation of biological reactions. The chemostat can function in CSTR mode (Figure 1), where it is operated at a constant volume V with an equal outflow and inflow rate F, and it is fed with an inlet substrate concentration S_{in}.

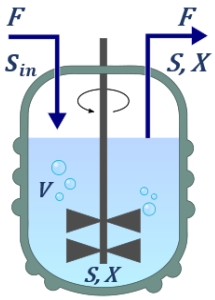

FIGURE 1. Continuous Stirred Tank Reactor (CSTR). [1]

Here, we consider a chemostat with a simple structure with only one limiting substrate S and one population of microorganisms X (bacteria). The dynamics of such system is described by a set of nonlinear differential equations derived from the law of mass balance [13].

$$\begin{cases} \dot{S} = D(S_{in} - S) - \mu(S)X \\ \dot{X} = (\mu(S) - D)X \end{cases} \tag{1}$$

S and X are the substrate and the biomass concentrations, respectively, S_{in} is the input substrate concentration, D is the dilution rate given by $D = \frac{F}{V}$, $\mu(S)$ is the specific growth rate of the biomass on the substrate (the kinetics), which can be modeled by different functions, but the one we use here is the Monod type function given by [2]:

$$\mu(S) = \mu_{max}\frac{S}{S+K} \tag{2}$$

where μ_{max} is the maximum growth rate and K is the half-saturation constant ($\mu(K) = \frac{\mu_{max}}{2}$). The specific growth rate $\mu(S)$ is a function of the limiting substrate concentration S.

In this work, we suppose that the initial conditions S_0 et X_0 of system (1) are in the box $[S_0^-, S_0^+] \times [X_0^-, X_0^+] \subset \Sigma$ where the set Σ is defined by :

$$\Sigma = \{(S, X) \in \mathbb{R}_+^2 : 0 < S < S_{in} \text{ and } X > 0\}$$

It is well known that the state variables will remain positive for all future times [1]. We assume the measured output of system (1) is the substrate $y = S$, and we consider $D \in [D_{Min}, D_{Max}]$ as the control variable.

Substituting (2) by (1), the system is rewritten as follows:

$$\begin{cases} \dot{S} = D(S_{in} - S) - \mu_{max}\frac{S}{S+K}X \\ \dot{X} = (\mu_{max}\frac{S}{S+K} - D)X \end{cases} \tag{3}$$

The chemostat model in (3) has at most two equilibrium points : the undesired washout equilibrium where the chemostat is washed out by bacteria, and the desired positive equilibrium where the chemostat is normally functioning (i.e S is degraded by X). Stability of equilibria depends on operating parameters D and S_{in}. Figure 2 displays the behavior of system (3), when it is functioning around the positive equilibrium. For more details on analysis of equilibria and their stability, the reader can refer to [14]. In this paper, the continuous-time system (3) is turned into discrete-time dynamics model in order to make numerical control and data storage easier. Denote $x = [S, X]^\top$ the state vector, hence nonlinear system (3) can be characterized as $\dot{x} = f(x, u)$. Utilizing the 4^{th} order Runge-Kutta technique (RK4) with a specified sampling time T_d as in [22], the discrete-time model of the chemostat is given by:

$$x_{k+1} = f(x_k, u_k), \tag{4}$$

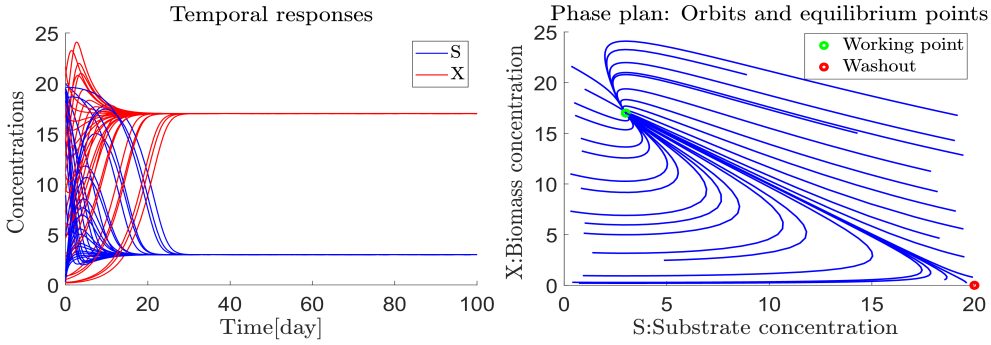

FIGURE 2. Numerical simulation of system (3) around the positive equilibrium. Right: the phase plan $X(t)$ with respect to $S(t)$ for different initial conditions., Left: the solutions $S(t)$ and $X(t)$ with respect to time

where $x_k = [S_k, X_k]^\top$, $x_k \in \Sigma \subset \mathbb{R}^2_+$ and $u \in \mathcal{U} \subset \mathbb{R}^1$. The main contribution of this study is the development of a data-driven control approach for the chemostat that is purely based on input-output data.

3. Koopman model of the chemostat

The Koopman operator theory is a data-driven approach that provides an infinite-dimensional linear operator which can transform a nonlinear dynamical system to a linear one [5, 8]. For the intent of this article, we focus more on the functional development of the idea behind the Koopman operator. Nevertheless, for a large review on the Koopman operator theory and its applications, the reader is referred to [15]. In this section, we will first present the Koopman model for the chemostat when having access to the full-state measurements of the system. The case of input-output Koopman model of the chemostat is addressed in the second subsection where we have restrictions in terms of measuring the full-state of the system.

3.1. Koopman model of the chemostat with full-state measurements

For the dynamic system (4), first we can define an observable function $g \in \mathbb{G}$ where $g : E \to \mathbb{R}$ and E is the Cartesian product of the original state-space $\mathbb{R}^n$ and the space of all input sequences $\ell(\mathcal{U}) = \{(u_k)_{k=0}^{\infty} | u_k \in \mathcal{U}\}$, i.e., $E = \mathbb{R}^n \times \ell(\mathcal{U})$. Now, we can define the Koopman operator $\mathcal{K} : \mathbb{G} \to \mathbb{G}$:

$$[\mathcal{K}g](x, (u_k)_{k=0}^{\infty}) = g(f(x, u_0), (u_k)_{k=1}^{\infty}). \tag{5}$$

The Koopman operator $\mathcal{K}$ in (5) is an infinite-dimensional linear operator, that acts on observables g defined on state space and maps them into new function $\mathcal{K}g$ also defined on state space ($\mathcal{K}$ updates g based on changes in state space trajectories). Thus, the nonlinear dynamical system (4) can be equally

represented by a linear operator of infinite dimensions, since the set of all observables form an infinite-dimensional vector space. Implicitly, the Koopman operator (5) supposes that $\mathbb{G}$ is invariant under its actions, hence, $\mathbb{G}$ usually can also include functions based on $(u_k)_{k=0}^{\infty}$ even in the controlled setting. For a detailed explanation of the Koopman operator in the extended state-space, see [9]. We underline that the Koopman operator represents the nonlinear system in a linear fashion with a generalization of the local linearization around equilibria; this linear representation is globally accurate as demonstrated in [16]. In this paper, we are focusing on time domain forecast of the trajectories of (4). For that reason, we are using a finite-dimensional approximation of the Koopman operator $\mathcal{K}$ by applying a data-driven technique called EDMD [9, 10]. The procedure is done in two main steps:

(1) The chemostat's measured states, including substrate and biomass concentrations, are uplifted to the space of observables, but the control inputs are not lifted;
(2) An approximation of the Koopman operator is obtained by applying a least-square regression on the uplifted data.i.e., we are seeking to approximate the nonlinear dynamics of the chemostat by linear time-invariant model given by:

$$\begin{aligned} s_{k+1} &= As_k + Bu_k, \\ \hat{x} &= Cs_k, \end{aligned} \tag{6}$$

where the state $s_k \subset \mathbb{R}^{N+2}$ is the lifted state, and the first two elements of s_k are represented by the substrate and the biomass concentrations respectively; $u_k \subset \mathbb{R}^1$ is the control input which is not lifted; $\hat{x} \subset \mathbb{R}^2$ is the Koopman model output or the prediction of x.

Consider the collection of data generated from the nonlinear system (4) in the form of:

$$\boldsymbol{X} = [x_1, .., x_K], \;\; \boldsymbol{X'} = [x_2, .., x_{K+1}], \;\; \boldsymbol{U} = [u_1, .., u_K] \tag{7}$$

where K is the number of collected data points. The data sets in (7) should be obtained using the same sampling time. However, it is not necessary to have temporal ordering in the data and is not generated from the same trajectory of (4).

First, the data $\boldsymbol{X}, \boldsymbol{X'}$ and $\boldsymbol{U}$ presented in (7) is lifted to a higher-dimensional space with the help of the lifting function $\phi : \mathbb{R}^2 \to \mathbb{R}^{N+2}$ where ϕ is defined as

$$\boldsymbol{\phi}(x) = [x^\top, \; \phi_1(x), \ldots, \phi_N(x)]^\top, \tag{8}$$

where the original state $x = [S \quad X]^\top$ is included; $\{\phi_i\}_{i=1}^{N}$ denote a vector of user-defined (possibly nonlinear) basis functions, such as sinusoids, exponentials, monomials and radial basis functions. By combining (7) with (8), we can form a new set of lifted data snapshot matrices given by:

$$\boldsymbol{X_{lift}} = [\boldsymbol{\phi}(x_1), .., \boldsymbol{\phi}(x_K)], \;\; \boldsymbol{X'_{lift}} = [\boldsymbol{\phi}(x_2), .., \boldsymbol{\phi}(x_{K+1})], \;\; \boldsymbol{U} = [u_1, .., u_K] \tag{9}$$

These new data matrices are the lifted coordinates of the system (4) in the lifted space (the space of lifting functions). The components of the matrix U have not been lifted to maintain the Koopman model's linear dependency on the original input of the system as in [9]. In addition, the link between the original state x_k and the lifted state s_k is presented by:

$$s_k = \boldsymbol{\phi}(x_k) = [x_k^\top, \phi_1(x_k), .., \phi_N(x_k)]^\top, \tag{10}$$

The matrix $A \in \mathbb{R}^{(N+2)\times(N+2)}, B \in \mathbb{R}^{(N+2)\times 1}$ and $C \in 2 \times \mathbb{R}^{(N+2)}$ in (6) are given as the solution to the following optimization problems:

$$\min_{A,B} \parallel \boldsymbol{X'_{lift}} - A\boldsymbol{X_{lift}} - B\boldsymbol{U} \parallel_F, \quad \min_{C} \parallel \boldsymbol{X} - C\boldsymbol{X_{lift}} \parallel_F, \tag{11}$$

With the symbol $\parallel . \parallel_F$ as the Frobenius norm [1]of a matrix. The optimization problems presented in (11) are linear least-squares problems, where A and B are calculated as the best one-step linear model in the lifted space as a least square sense and C is calculated as the best linear least squares prediction of $\boldsymbol{X}$ given $\boldsymbol{X_{lift}}$. For more details on the analytical solution and the application of the Koopman operator theory for the data-driven forecast of the chemostat with full-state measurements, see [17].

3.2. Koopman model for the input-output chemostat

Now, we illustrate how the technique may be adapted to the scenario where the measurements of the full-state are unavailable and just a single output is observed. ($y = S$, only the measurement of the substrate concentration is available, in practice, the measurement of biomass is always limited). The discrete-time representation of the nonlinear input-output dynamical system of the chemostat is considered as follows :

$$\begin{aligned} x_{k+1} &= f(x_k, u_k), \\ y_k &= h(x_k), \end{aligned} \tag{12}$$

where $x_k = [S_k \quad X_k]^\top$ is the state of the system with x_{k+1} as its successor, $x_k \in \Sigma \subset \mathbb{R}^2_+$, $u_k = D$ is the control input, f is the transition mapping and y_k is the measured output which in our case $h(x_k) = S_k$ so $h : \mathbb{R}^2_+ \to \mathbb{R}_+$. We are specifically searching for a basic model with a linear structure that is suitable for linear control design approaches such as MPC [11]. The Koopman model is assumed to be of the forms of a controlled linear dynamical system

$$\begin{aligned} s_{k+1} &= As_k + Bu_k, \\ \hat{y}_k &= Cs_k, \end{aligned} \tag{13}$$

With $\hat{y_k}$ being the prediction of the output y_k in (12). Notice that, the control input u_k of (13) stays the same as in (12) so that linear constraints can be applied linearly on the control inputs. In order to construct such a valid linear model for the chemostat given in (12), we should follow all the steps

[1]The Frobenius norm of a matrix A is given by $\parallel A \parallel_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}$

of the Koopman operator framework presented in Section 3.**1**. However, the one distinction is that we will choose the lifting functions depending on the actual measured output and also many previous measured outputs and inputs at the same time. This is known as the time-delayed embedding (embedding into a single data point many successive measurements of the output) which is a classical methodology extremely powerful in system identification theory (see, e.g., [19]), but recently the context of the Koopman operator has been including this technique in its approximations (see, e.g., [20, 8]).

Let us consider the following set of data matrices as in (7):

$$\tilde{\boldsymbol{X}} = [\boldsymbol{\xi}_1, .., \boldsymbol{\xi}_K], \;\; \tilde{\boldsymbol{X}}^{+} = [\boldsymbol{\xi}_1^{+}, .., \boldsymbol{\xi}_K^{+}], \;\; \boldsymbol{U} = [u_1, .., u_I] \tag{14}$$

The only difference is that the matrices include a string of samples of series with $n_d + 1$ length (n_d is a number of delays), where

$$\begin{aligned} \boldsymbol{\xi}_i &= [y_{i,n_d}^{\top} \quad \hat{u}_{i,n_d-1}^{\top} \quad y_{i,n_d-1}^{\top} \quad \dots \quad \hat{u}_{i,0}^{\top} \quad y_{i,0}^{\top}]^{\top} \in \mathbb{R}^{(n_d+1)n_h+n_d} \\ \boldsymbol{\xi}_i^{+} &= [y_{i,n_d+1}^{\top} \quad \hat{u}_{i,n_d}^{\top} \quad y_{i,n_d}^{\top} \quad \dots \quad \hat{u}_{i,1}^{\top} \quad y_{i,1}^{\top}]^{\top} \in \mathbb{R}^{(n_d+1)n_h+n_d} \\ u_i &= \hat{u}_{i,n_d} \end{aligned} \tag{15}$$

$(\hat{u}_{i,j})_{j=0}^{n_d}$ is a series of inputs generating a vector $(y_{i,j})_{j=0}^{n_d+1}$ of consecutive output measurements.

As in (9), the data matrices $\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}^{+}$ and $\boldsymbol{U}$ given in (14) are lifted with the help of the following lifting function:

$$\boldsymbol{\phi}(\boldsymbol{\xi}) = [\phi_1(\boldsymbol{\xi}), \dots, \phi_N(\boldsymbol{\xi})]^{\top}, \tag{16}$$

Then, new sets of lifted data are obtained as follows:

$$\tilde{\boldsymbol{X}}_{\boldsymbol{lift}} = [\boldsymbol{\phi}(\boldsymbol{\xi}_1), .., \boldsymbol{\phi}(\boldsymbol{\xi}_K)], \; \tilde{\boldsymbol{X}}_{\boldsymbol{lift}}^{+} = [\boldsymbol{\phi}(\boldsymbol{\xi}_1^{+}), .., \boldsymbol{\phi}(\boldsymbol{\xi}_K^{+})], \; \boldsymbol{U} = [u_1, .., u_K] \tag{17}$$

By having $\boldsymbol{X}_{\boldsymbol{lift}}, \boldsymbol{X}_{\boldsymbol{lift}}^{+}$ and the input matrix $\boldsymbol{U}$, we can find the linear matrices A, B and C of (13) by solving the following least squares problems:

$$\min_{A,B} \parallel \tilde{\boldsymbol{X}}_{\boldsymbol{lift}}^{+} - A\tilde{\boldsymbol{X}}_{\boldsymbol{lift}} - B\boldsymbol{U} \parallel_F, \quad \min_{C} \parallel [y_{1,n_d} \dots y_{K,n_d}] - C\tilde{\boldsymbol{X}}_{\boldsymbol{lift}} \parallel_F, \tag{18}$$

The predictor (13) starts from the initial condition:

$$s_0 = \boldsymbol{\phi}(\boldsymbol{\xi_0}) \tag{19}$$

Where

$$\boldsymbol{\xi_0} = [y_0^{\top} \;\; \hat{u}_{-1}^{\top} \;\; y_{-1}^{\top} \; \dots \; \hat{u}_{-n_d}^{\top} \, y_{-n_d}^{\top}]^{\top} \tag{20}$$

is the vector of n_d input measurements and n_d+1 are the latest measurements of the output.

4. Koopman MPC Framework

In the last section, we presented a methodology that allows us to build a model of the chemostat in the form of a linear dynamical system (13). In this section, we will use MPC to this linear model to control the original nonlinear system of the chemostat; this concept is depicted in Figure 3.

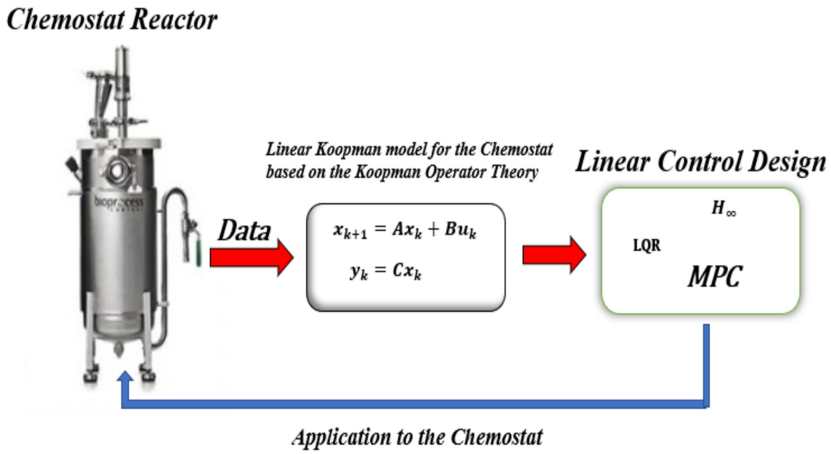

FIGURE 3. Koopman MPC Framework [9, 21].

MPC is used for constrained, feedback control of multivariable systems. In MPC, the altered input is determined by minimizing a user-defined cost function across a time horizon at each closed-loop sampling instant. Linear MPC solves a convex quadratic program (QP) to quickly assess the control input. Whereas Nonlinear MPC solves a significantly more sophisticated non-convex optimization problem every time step, making the solution exceedingly expensive to calculate. Local optimization approaches are therefore required. For an overview of MPC and Nonlinear MPC, the reader is referred to [11, 12].

The Koopman MPC uses the receding horizon control approach; first, the linear model (13) is used to forecast the evolution of the system during the prediction horizon, next, these forecasts are used to calculate the optimal control input sequence that minimizes the given cost function on this horizon, finally, we apply only the first component of the calculated control input sequence to the real system, thus generating a new output value and repeating the whole process. We will present, in the following, some notations and mathematical background on this technique. The discern property when using the Koopman linear model (13) is that the derived MPC problem is now a convex QP although we have a nonlinear dynamical system at the beginning. Since, the so-called *dense form* [9] is used , the solution of complex quadratic problem is independent of the possible very large dimension of the s, now, the QP can be easily solved with the help of optimized and highly efficient existing solvers for linear MPC.

The optimization problem solved by the Koopman MPC at each time instant of the closed loop process is as follows:

$$\begin{aligned}(u_i^\star)_{i=0}^{N_h-1}, (y_i^\star)_{i=0}^{N_h} = &\arg\min J((u_i)_{i=0}^{N_h-1}, (y_i)_{i=0}^{N_h}) \\ \text{s.t.} \quad & s_{i+1} = As_i + Bu_i, \qquad i = 0, ..., N_h \\ & y_i = Cs_i \\ & W_i^y y_i + W_i^u u_i \leq v_i, \qquad i = 0, ..., N_h - 1 \\ & W_{N_h} y_{N_h} \leq v_{N_h} \\ & s_0 = \boldsymbol{\phi}(\boldsymbol{\xi_k}),\end{aligned} \tag{21}$$

where N_h is the prediction horizon, and $(u_i)_{i=0}^{N_h-1}, (y_i)_{i=0}^{N_h}$ are the sequence of input and output value over N_h,the matrices $W^y_{i=0,...,N_h-1}, W^u_{i=0,...,N_h-1}$ W_{N_h} and the vector v_i are the state and input polyhedral constraints. The current state $\boldsymbol{\xi_k}$ of the delayed-state $\boldsymbol{\xi}$ at a time instant k is as follows:

$$\boldsymbol{\xi_k} = [y_k^\top \;\; u_{k-1}^\top \;\; y_{k-1}^\top \;\; \dots \;\; u_{k-n_d}^\top \; y_{k-n_d}^\top]^\top \tag{22}$$

The cost function J has a convex quadratic form given by:

$$\begin{aligned}J((u_i)_{i=0}^{N_h-1}, (y_i)_{i=0}^{N_h}) = &\; y_{N_h}^\top Q_{N_h} y_{N_h} + q^\top y_{N_h} \\ &+ \sum_{i=1}^{N_h-1} (y_i)^\top Q_i y_i) + u_i^\top R_i u_i + q_i^\top y_i + r_i^\top u_i \\ &+ u_0^\top R u_0 + r_0^\top u_0,\end{aligned} \tag{23}$$

with $R_{i=0,...,N_h-1}, Q_{i=0,...,N_h}$ are cost function matrices (real symmetric positive semi-definite), the cost function J can be utilized to formulate lots of control objectives, for instance, time-varying reference tracking. Furthermore, nonlinear constrains and objectives in the original state x_k can be handled by integrating nonlinear functions in the embedding variables (see [9],[21]). When the optimal input sequence $(u_i^\star)_{i=0}^{N_h-1}$ is calculated, its first component $u_0^\star$ is applied to the system to generate a new value of the measured output $(h(x_k))$ which refreshes the current state $\boldsymbol{\xi_k}$ and the whole procedure is then repeated in a receding horizon way. Algorithm 1 summarizes the Koopman MPC approach's closed-loop operation.

Algorithm 1 Closed-loop operation of Koopman MPC

1: **for** k=0,1,... **do**
2: Set $s_0 = \boldsymbol{\phi}(\boldsymbol{\xi_k})$
3: Solve (21) to get an optimal solution $(u_i^\star)_{i=1}^{N_h}$
4: Apply $u_1^\star$ to the nonlinear system of the chemostat (12)
5: **end for**

5. **Simulation results and discussions**

In this section we extend the proposed data-driven technique to control the nonlinear input-output dynamical system of the chemostat (12). The control of the chemostat dynamics is always restricted in the large space of initial conditions. Rather, it is usually controlled in the invariant manifold (i.e., the trajectories begin from initial conditions $x_1(0), x_2(0)$ generated randomly with uniform distribution where $x_1(0) + x_2(0) = S_{in}$). We are going to construct a linear Koopman model and use it to control the chemostat in the general and difficult case (control in a large space of the initial conditions). The parameters of the chemostat dynamics are presented in Table 1. In terms of collecting

TABLE 1. Parameters of the chemostat dynamics

Parameters	Values	Units
S_{in}	20	mg/l
μ_{Max}	0.6	d^{-1}
K	3	mg/l
D	$[0.01, 0.5]$	$d^{-}1$

the set of data as in (14), we will consider the dynamical system of the CSTR in (12) as a virtual bioreactor generating this data (The data can be collected from a real chemostat system, if available) and we use it for constructing the Koopman model (13). To collect the data in (14), the discretization period of the RK4 method is chosen to be $T_d = 0.1$ day (since the dynamics of the chemostat are evolving slowly due to the consumption of the substrate by the biomass) and we simulate 200 trajectories over 1000 sampling periods with a random control input ($u = D$) signal equally distributed. There are many basis functions that can be used for the construction of the linear Koopman model of the chemostat. In this paper, we choose the Inverse Multi-Quadric Radial Basis Function (IMQ-RBF) [23] given by:

$$\phi(x) = \frac{1}{\sqrt{1+ \parallel x - x_c \parallel^2}} \tag{24}$$

where x_c is the center. The IMQ-RBF above is one of the most effective and commonly-used basis functions in the approximations of the Koopman operator from data, as recommended in [10]. Furthermore, we choose that the trajectories of the system (12) begin from initial conditions $x_1(0), x_2(0)$ randomly generated with a uniform distribution, where $x_1(0) \in [10, 20]$ and $x_2(0) \in [2, 10]$ (it is obvious that in general $x_1(0)+x_2(0) \neq S_{in}$). We choose the number of delays $n_d = 1$ and the lifting functions $\boldsymbol{\phi}_i = S_k$ taken to form the time-delayed vector $\boldsymbol{\xi} \in \mathbb{R}^3$ in (15). We used 60 IMQ-RBFs (24) with centers randomly selected with a consistent distribution over $[0, 20]^3$. Therefore, the dimension of the lifted space is $N = 63$.

5.1. Model prediction comparison

First, we compare the prediction accuracy of the output of the constructed linear Koopman model (13) with the output of the true dynamics of the chemostat and a model based on local linearization of the chemostat at a given initial condition x_0 (the prediction accuracy is evalutated using the Relative Mean Squared Error (RMSE) (25)).

$$RMSE = 100 \times \sqrt{\frac{\Sigma_k \parallel x_{Predicted}(kT_d) - x_{Real}(kT_d) \parallel_2^2}{\Sigma_k \parallel x_{Real}(kT_d) \parallel_2^2}} \tag{25}$$

Figure 4 shows the accuracy of the output forecast of the Koopman model (13) constructed only from input-output data generated from two randomly chosen initial conditions ($x_0^1 = [10, 9]^\top, x_0^2 = [12, 5]^\top$), and how it fits the real system of the chemostat compared to the local linearization-based model, in Table 2 we demonstrate the superiority of the Koopman model prediction accuracy over the linerization-based model for longer prediction times and several initial conditions. This is done by the RMSE (25) averaged over 30 days forecast horizon over 100 randomly selected initial conditions (in both Figure 4 and Table 2, we have applied a pseudo-random binary control signal $u = [0.1, 0.35]$ for each initial condition anew).

TABLE 2. Forecast comparison using RMSE (25).

Model	**Average RMSE**
Koopman Model	**11.68 %**
Local linearization at x_0	*4.72* $\times 10^4$ *%*

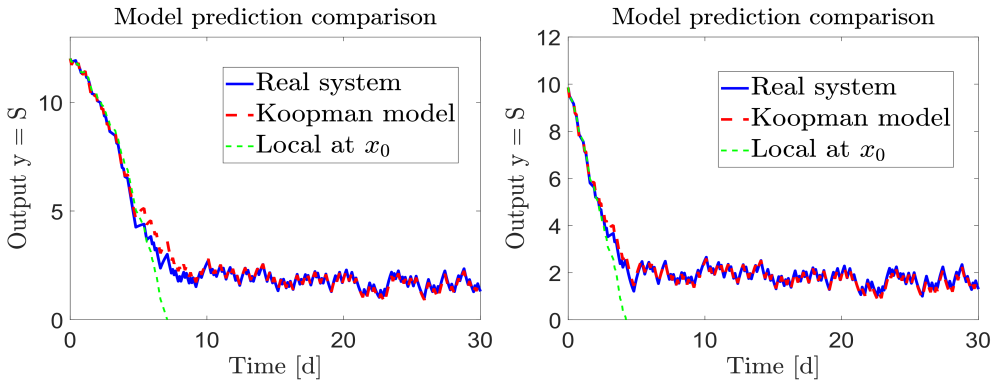

FIGURE 4. Control in larger space of the initial condition - Model prediction comparison of the chemostat - $y_k = S_k$. Right: initial condition $x_0[10, 9]^\top$. Left: initial condition $x_0[12, 5]^\top$

5.2. Feedback control of the chemostat

Now, we apply the Koopman-MPC controller proposed in Section 4 for the feedback control of the chemostat in the large space of initial condition. The control objective is to track a given substrate concentration reference y_{Ref}, which means that we need to minimize the following objective function of the MPC problem:

$$\begin{aligned} J = &(y_{N_h} - y_{Ref})^\top Q_{N_h}(y_{N_h} - y_{Ref}) \\ &+ \sum_{i=1}^{N_h-1} (y_i - y_{Ref})^\top Q(y_i - y_{Ref}) + u_i^\top R u_i, \end{aligned} \tag{26}$$

Where cost functions matrices were chosen as $Q = Q_{N_h} = 10$ and $R = 0.01$ with a prediction horizon $N_h = 10$(i.e.,one day). We establish a comparison between MPC controller based on the Koopman model presented in this study (Koopman-MPC) and a local linearization-based MPC (Linear-MPC) in two cases. First, we track a reference constant ($y_{Ref} = S_k^{Constant}$) where we want to minimize the output which is the substrate concentration S_k. With no constraints imposed on the output but we have imposed constraints on the control input $u \in [0.2, 3]$. In the second case, we track a time-varying function $y_{Ref}(t) = 17cos(2\pi t/300)$ with constraints imposed on the output where $y \in [2, 15]$. Simulation results are presented in Figure 5, we can see that the tracking performance in the first case is almost identical. Both Koopman-MPC and Linear-MPC reach the desired output in the same time and manner. However, the Koopman-MPC controller has the advantage of being entirely *data-driven* with only *output measurements* being required. For the second case, we can clearly see a good reference tracking without any violation of the output constraints with a constrained control input applied ($u \in [0.2, 3]$) for the Koopman-MPC controller, however, the Linear-MPC controller turns out to be infeasible and therefore it stops before continuing the entire simulation period (in predictive control, infeasibility is a common issue that occurs as a result of different trial-and-error or theoretically backed-up methods, for more details, see [11, 12]). The Linear-MPC controller is infeasible due to the local linearization-based model's inaccurate long-term predictions. The Koopman-MPC controller is data-driven and can overcome simulation infeasibility. All the defining data are pre-computed offline with the Koopman-MPC controller, but with the Linear-MPC controller, they are re-computed at each iteration. Koopman-MPC is thus faster than Linear-MPC. These results are gained for a large space of initial conditions when chemostat control is limited.

6. Conclusion

Mathematical models can be used to describe the chemostat's behavior. However, these models remain complex and uncertain, which makes system control difficult. Our research shows how we might use data-driven strategies

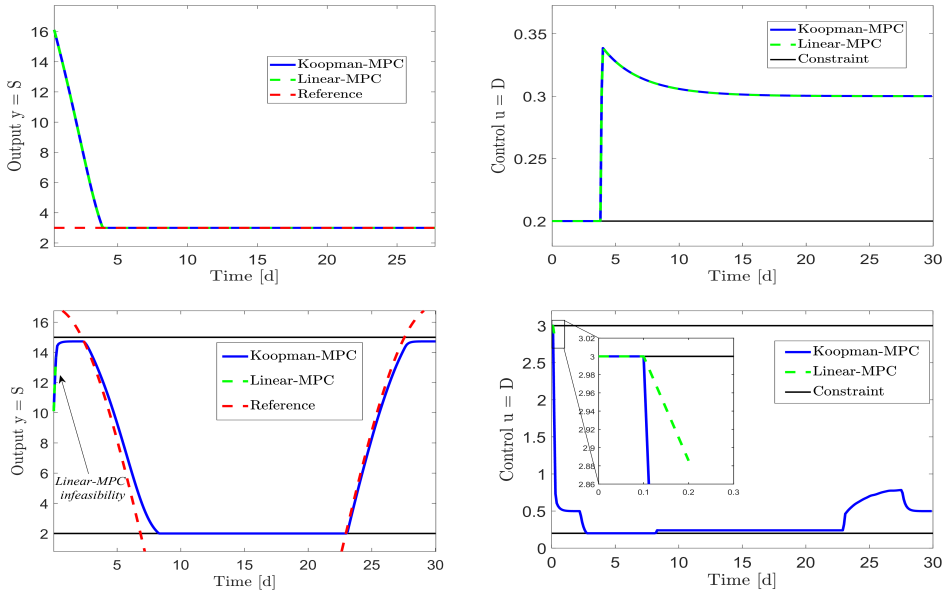

FIGURE 5. Feedback control of the chemostat in a large space of the initial conditions. Top: Constant reference tracking with no state constraints and constrained control input $u \in [0.2, 3]$, $y_{Ref} = 3$ and initial condition $x_0 = [18, 7]^\top$. Bottom: Time-varying reference tracking with constraints imposed on the output ($y \in [2, 15]$) and constrained control input ($u \in [0.2, 3]$) , initial condition $x_0 = [10, 9]^\top$

to create models from experimental data for control reasons.The Koopman-MPC technique [9] was utilized to build a global linear model for state estimation and control of the nonlinear chemostat model. The trick is to project the nonlinear model into a higher-dimensional space to make it appear linear. The linear Koopman model generated from the data is used to perform a linear MPC, which is applied to the original nonlinear chemostat model. Importantly, the whole control design process is internally data-driven, and only the measurements of input and output are required. The suggested chemostat modeling and control method outperforms the standard model-based MPC method in substrate concentration tracking accuracy and overall control performance. The ideal choice of lifting functions (observables) can reveal crucial information about the chemostat's dynamical system, especially when we use substrates with complicated growth rate functions. This will increase control and Koopman linear model accuracy.

Acknowledgements

This work is supported by the General Directorate for Scientific Research and Technological Development (DGRSDT), Algeria.

REFERENCES

[1] *H. Smith, P. Waltman* , The theory of the chemostat: Dynamics of Microbial Competition, **13** , 1995.

[2] *J. Monod*, La technique de culture continue: theorie et application, Ann. Inst.Pasteur,**79**, (1950) 390–410.

[3] *P. Leenheer, H. Smith*, Feedback control for chemostat models, Journal of Mathematical Biology, **46**, (2003) 48–70.

[4] *P.Skupin, M.Metzger*, Pi control for a continuous fermentation process with a delayed product inhibition, Journal of Process Control, **72**, (2018), 30–38.

[5] *B. Koopman, J. von Neuman*, Dynamical systems of continuous spectra, Proceedings of the National Academy of Sciences of the United States of America, (1932).

[6] *S. E. Otto, C. W. Rowley*, Koopman operators for estimation and control of dynamical systems, Annual Review of Control, Robotics and Autonomous Systems ,**4:1** ,(2021).

[7] *J. N. Kutz, S. L. Brunton, B. W. Brunton, J. L. Proctor*, Dynamic mode decomposition data-driven modeling of complex systems, SIAM, 2016.

[8] *A. Mauroy, I. Mezic, Y. Susuki*, The Koopman operator in systems and control concepts, methodologies, and applications, Springer, 2020.

[9] *M. Korda, I. Mezic*, Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control, Automatica, **93**, (2018), 149–160. `arXiv:1611.03537 [math.OC]`

[10] *M. O. Williams, I. G. Kevrekidis and C. W. Rowley*, A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition., Journal of Nonlinear Science ,(2015).

[11] *L. Grune, J. Pannek*, Nonlinear Model Predictive Control: Theory and Algorithms, **1**, 2011.

[12] *D. Q. Mayne, J. B. Rawlings, C. V. Rao and P. O. Scokaert*, Constrained model predictive control: Stability and optimality, Automatica, **36**, (2000), 789–814.

[13] *G. Bastin, D. Dochain*, On-line estimation and adaptive control of bioreactors, **1**, 1990.

[14] *J. Harmand, C. Lobry, A. Rapaport and T. Sari*, The chemostat: mathematical theory of microorganism cultures, **1**, 2017.

[15] *M. budicic, R. Mohr, I. Mezic*, Applied koopmanism, Chaos: An Interdisciplinary, Journal of Nonlinear Science, **22**, (2012) .

[16] *I. Mezic*, Koopman operator spectrum and data analysis, arXiv: Chaotic Dynamics, (2017).

[17] *Benaissa Dekhici , Boumediene Benyahiya , Brahim Cherki* , Forecast of chemostat dynamics using data-driven approach, 2021 International Conference on Control, Automation and Diagnosis (ICCAD), 2021.

[18] *M. Korda, I. Mezic*, On convergence of extended dynamic mode decomposition to the koopman operator, Journal of Nonlinear Science, **28**,(2018), 687–710.

[19] *L. Ljung*, System identification. In Signal Analysis and Prediction, 1998.

[20] *S. L. Brunton, B. W. Brunton, J. L. P, E. Kaiser, J. N. Kutz*, Chaos as an intermittently forced linear system, Nature communications, **8**, (2017).

[21] *H. Arbabi and M. Korda and I. Mezić*, A Data-Driven Koopman Model Predictive Control Framework for Nonlinear Partial Differential Equations, 2018 IEEE Conference on Decision and Control (CDC), 2018.

[22] *Chen, Bin and Huang, Zhiwu and Zhang, Rui and Liu, Weirong and Li, Heng and Wang, Jing and Fan, Yunsheng and Peng, Jun*, Data-Driven Koopman Model Predictive Control for Optimal Operation of High-Speed Trains, IEEE Access, 2021.

[23] *M. E. Biancolini*, Fast Radial Basis Functions for Engineering Applications, Springer, 2017.