IFAC Journal of Systems and Control 35 (2026) 100384

Contents lists available at ScienceDirect

IFAC Journal of Systems and Control

journal homepage: www.elsevier.com/locate/ifacsc

Full length article

Data-driven modeling with prior system knowledge ☆

Fritz A. Engeln [a,*], Jan-Willem van Wingerden [a], Timm Faulwasser [b]

[a] Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands
[b] Institute of Control Systems, Hamburg University of Technology, Hamburg, Germany

ARTICLE INFO

Article history:
Received 25 October 2025
Received in revised form 10 January 2026
Accepted 31 January 2026

Keywords:
Data-driven control
System identification
Filtering
Persistency of excitation
Physics-informed learning

ABSTRACT

The behavior of a linear time-invariant system can be characterized entirely by measured input–output data that spans the vector space of all possible trajectories of the system relying on the fundamental lemma by Willems et al. However, useful *a priori* knowledge of the system is usually neglected. We propose a novel method for incorporating prior knowledge, specifically, known pole and zero locations, into a data-driven representation by constructing filters that pre-process the measured input–output data. To this end, a physics-informed data-driven predictor is introduced, where trajectories are obtained as linear combinations of the columns of a filtered block-Hankel matrix. We explicitly derive the output prediction error and show how leveraging prior knowledge reduces the impact of future noise realizations on output predictions and improves the accuracy of the initial state that is inferred from past data.

1. Introduction

In recent years, research attention in systems and control shifts from model-based methods towards data-driven approaches (Dörfler, 2023). Classical predictive control techniques such as model predictive control (MPC) aim to compute an optimal control policy based on output feedback, while minimizing an objective function and satisfying input–output constraints (Rawlings et al., 2017). Their performance depends critically on the deployment of an accurate model of the underlying system. Such model can be obtained from first principles, e.g., by modeling an electrical circuit using the constitutive relations to obtain differential equations. However, due to system complexity and uncertainties in many practical applications, obtaining such model from first principles is costly, or even infeasible. The fundamental problem of finding a model from measured data, known as system identification, was first introduced by Zadeh (1956) and has become a well-established research field (Ljung, 1999). Instead of relying on detailed knowledge of the underlying physics, system identification methods generate models that match measured input–output data. More recently, there has been growing interest in learning controllers directly from data without relying on an explicit parametric model. Instead, a data-driven framework is employed, which characterizes all possible trajectories of an unknown system using measured input–output data. The framework has its origin in the behavioral systems theory for discrete-time linear time-invariant (LTI) systems and is known as the *fundamental lemma* by Willems et al. (2005). From the behavioral perspective, a system is not described by a difference equation relating inputs and outputs, but rather by the linear subspace of all input–output trajectories it can generate. This data-driven system representation is well-suited for the design of data-driven predictive control (DPC) schemes (Coulson et al., 2019; Yang & Li, 2015), where the system's dynamics are constrained to the vector space of all input–output trajectories that belong to the system.

While data-driven methods circumvent the need for an accurate model, partial *a priori* knowledge is often neglected. Prior knowledge, such as pole and zero locations, can be obtained through empirical observations or derived from first principles. In particular, pole locations are directly associated with the system's dynamic behavior.

Several extensions of the *fundamental lemma* have been proposed, some of which leverage prior knowledge of the system. Berberich and Allgöwer (2020) use knowledge of basis functions to extend the *fundamental lemma* for certain classes of nonlinear systems. Zieglmeier et al. (2025) introduce a semi-DPC scheme, which is constrained by a parametric submodel in parallel with a data-driven submodel, capturing the known and unknown dynamics, respectively. However, separate initialization of each model remains challenging in practice. Li et al. (2024) address this by assuming a parametric model that, while imperfect, approximates the entire input–output behavior of the system rather than only a subset of its dynamics.

☆ This article is part of a Special issue entitled: 'IFAC WC 2026 - TC1.1' published in IFAC Journal of Systems and Control.

* Corresponding author.
E-mail addresses: f.a.engeln@tudelft.nl (F.A. Engeln), j.w.vanwingerden@tudelft.nl (J.-W. van Wingerden), timm.faulwasser@ieee.org (T. Faulwasser).

https://doi.org/10.1016/j.ifacsc.2026.100384

The main contribution of this paper is the design of filters based on prior knowledge, such as known pole and zero locations, and the integration of this knowledge into a data-driven system representation by applying the filters directly to the available data. In contrast to traditional filtering methods, which apply the same filter to both input and output signals to preserve the input–output relationship (Ljung, 1999), our approach intentionally modifies this relationship to embed known system properties. We formulate a data-driven predictor (DDP) based on the *fundamental lemma* and provide an explicit derivation of the resulting prediction errors. To the best of our knowledge, this is the first work to introduce such an approach.

The remainder of this paper is organized as follows. Section 2 introduces the model structure, notation, and essential theoretical concepts. A method to incorporate prior knowledge into data-driven system representations is presented in Section 3, which is followed in Section 4 by a derivation and explanation of the output prediction error. To further highlight the impact of prior knowledge, Section 5 illustrates simulated output predictions and errors. Section 6 provides some final conclusions and directions for future work.

Notation: Standard symbols are used throughout: the zero matrix is denoted by $\mathbf{0}$, the identity matrix by I, and a sequence of zeros of length T by $\underline{0}_T$. The ℓ_2-norm is written as $\|\cdot\|_2$, the Frobenius norm as $\|\cdot\|_F$, and the Moore–Penrose pseudoinverse as $(\cdot)^\dagger$.

2. Preliminaries

Following Ljung (1999), we consider a discrete-time multiple-input, multiple-output (MIMO) LTI system in innovation form

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Ke_k \\ y_k &= Cx_k + Du_k + e_k, \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^{n_u}$, $e_k \in \mathbb{R}^{n_y}$, and $y_k \in \mathbb{R}^{n_y}$ are the state, input, innovation noise and output vector, respectively. The discrete-time index is denoted by $k \in \mathbb{Z}_{\geq 0}$. The matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times n_u}$, $C \in \mathbb{R}^{n_y\times n}$, and $D \in \mathbb{R}^{n_y\times n_u}$ represent the system, input, output, and direct feedthrough matrix, respectively. The innovation e_k is a Gaussian random variable with zero-mean. The Kalman gain $K \in \mathbb{R}^{n\times n_y}$ and the covariance Σ_e of the innovation are obtained from the solution of the stationary discrete-time algebraic Riccati equation (DARE).

By introducing $\tilde{A} = A - KC$ and $\tilde{B} = B - KD$, the innovation form in (1) is transformed into predictor form:

$$\begin{aligned} x_{k+1} &= \tilde{A}x_k + \tilde{B}u_k + Ky_k \\ y_k &= Cx_k + Du_k + e_k, \end{aligned} \tag{2}$$

where the state does not explicitly depend on innovations.

We denote a stacked vector of T consecutive input samples starting at time $\hat{i}$ by $\underline{u}_{\hat{i},T} \in \mathbb{R}^{n_u T}$ and a block-Hankel matrix of depth L constructed from the sequence $\underline{u}_{\hat{i},T}$ by

$$U_L = \begin{bmatrix} U_p \\ U_f \end{bmatrix} = \begin{bmatrix} u_{\hat{i}} & \cdots & u_{\hat{i}+T-L} \\ \vdots & \ddots & \vdots \\ u_{\hat{i}+L-1} & \cdots & u_{\hat{i}+T-1} \end{bmatrix} \in \mathbb{R}^{n_u L\times(T-L+1)},$$

with $p,\ f \in \mathbb{Z}_{>0}$, referred to as the past and future horizons, respectively, such that $L = p + f$. Similarly, the matrices $Y_L \in \mathbb{R}^{n_y L\times(T-L+1)}$ and $E_L \in \mathbb{R}^{n_y L\times(T-L+1)}$ are constructed from the sequences $\underline{y}_{\hat{i},T}$ and $\underline{e}_{\hat{i},T}$.

Definition 1. We say that a sequence $\underline{u}_{\hat{i},T}$ is persistently exciting (PE) of order L if $\operatorname{rank}(U_L) = n_u L$.

Consider the block-Toeplitz matrix

$$\mathcal{T}_{(B,D)} = \begin{bmatrix} D & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ CB & D & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}B & CA^{f-3}B & CA^{f-4}B & \cdots & D \end{bmatrix},$$

with $\mathcal{T}_{(B,D)} \in \mathbb{R}^{n_y f\times n_u f}$, and similarly $\mathcal{T}_{(K,I)} \in \mathbb{R}^{n_y f\times n_y f}$. The extended controllability matrix is defined as

$$\mathcal{K}_{(\tilde{B})} = \begin{bmatrix} \tilde{A}^{p-1}\tilde{B} & \tilde{A}^{p-2}\tilde{B} & \cdots & \tilde{B} \end{bmatrix} \in \mathbb{R}^{n\times n_u p},$$

and similarly $\mathcal{K}_{(K)} \in \mathbb{R}^{n\times n_y p}$. The notation is simplified by introducing $\mathcal{K} = \begin{bmatrix} \mathcal{K}_{(\tilde{B})} & \mathcal{K}_{(K)} \end{bmatrix} \in \mathbb{R}^{n\times(n_u+n_y)p}$. Furthermore, the extended observability matrix is given by

$$\Gamma = \begin{bmatrix} C^\top & (CA)^\top & \cdots & (CA^{f-1})^\top \end{bmatrix}^\top \in \mathbb{R}^{n_y f\times n}.$$

2.1. The data-driven predictor

Let an nth-order minimal system be given by the innovation form in (1). We derive a DDP, following the steps outlined by van Wingerden et al. (2022).

Assumption 2. The output data y is noise-free.

Assumption 3. The past horizon length p is sufficiently large such that $\|\tilde{A}^p\|_F \approx 0$.

Note, that in the noise-free setting, there exist a deadbeat observer gain K_{db}, such that $\|\tilde{A}^p\|_F = 0$ for all $p \geq n$. With Assumptions 2 and 3, the system satisfies the well-known data equations

$$\begin{aligned} \underline{y}_{i+p,f} &= \Gamma\mathcal{K}\begin{bmatrix} \underline{u}_{i,p} \\ \underline{y}_{i,p} \end{bmatrix} + \mathcal{T}_{(B,D)}\underline{u}_{i+p,f} \\ Y_f &= \Gamma\mathcal{K}\begin{bmatrix} U_p \\ Y_p \end{bmatrix} + \mathcal{T}_{(B,D)}U_f, \end{aligned} \tag{3}$$

in which the data sequences start at time i and $\hat{i}$, respectively. The initial state is obtained by multiplying past input–output data with the extended controllability matrices $\mathcal{K}$. The multiplication with Γ and $\mathcal{T}_{(B,D)}$ gives the zero-input and zero-state response, respectively.

By moving the outputs $\underline{y}_{i+p,f}$ and Y_f in (3) to the right-hand side and introducing a column selector $g \in \mathbb{R}^{T-L+1}$, we write

$$\begin{bmatrix} \Gamma\mathcal{K} & \mathcal{T}_{(B,D)} & -I \end{bmatrix}\left(\begin{bmatrix} U_p \\ Y_p \\ U_f \\ Y_f \end{bmatrix} g - \begin{bmatrix} \underline{u}_{i,p} \\ \underline{y}_{i,p} \\ \underline{u}_{i+p,f} \\ \underline{y}_{i+p,f} \end{bmatrix}\right) = \underline{0}_f. \tag{4}$$

Lemma 4 (*Willems et al., 2005*)**.** *Consider recorded input–output trajectories $\underline{u}_{\hat{i},T}$ and $\underline{y}_{\hat{i},T}$ of an nth-order controllable LTI system, where the input data $\underline{u}_{\hat{i},T}$ is PE of order $L+n$ and suppose U_L and Y_L to be constructed from that data. Then, $\underline{u}_{i,L}$ and $\underline{y}_{i,L}$ are trajectories of the system if and only if there exists a g such that*

$$\begin{bmatrix} U_p \\ U_f \\ Y_p \\ Y_f \end{bmatrix} g = \begin{bmatrix} \underline{u}_{i,p} \\ \underline{u}_{i+p,f} \\ \underline{y}_{i,p} \\ \underline{y}_{i+p,f} \end{bmatrix}. \tag{5}$$

Lemma 4 is known as *Willems' fundamental lemma* and implies that the data matrix has $n_u L+n$ linearly independent columns that span the trajectory space of the system and (4) simplifies to (5).

If the past input–output sequences $\underline{u}_{i,p}$ and $\underline{y}_{i,p}$ are known, (5) gives a unique relationship between future input and output

trajectories. Therefore, we will refer to (5) as implicit DDP. Furthermore, for a given future input sequence $\underline{u}_{i+p,f}$ the output can be explicitly calculated by first solving for the column selector

$$g = \begin{bmatrix} U_p \\ U_f \\ Y_p \end{bmatrix}^{\dagger} \begin{bmatrix} \underline{u}_{i,p} \\ \underline{u}_{i+p,f} \\ \underline{y}_{i,p} \end{bmatrix}. \tag{6}$$

Since the data matrix is usually wide, i.e., it has more columns than rows, the column selector g is generally not unique and the Moore–Penrose inverse, denoted by $(\cdot)^{\dagger}$, yields the solution with minimum ℓ_2-norm. Next, the future output trajectory is reconstructed from linear combinations of the columns of the data matrix

$$\underline{y}_{i+p,f} = Y_f g.$$

In the implementation of DPC, the system dynamics are typically enforced by the implicit DDP defined in (5), making an explicit output prediction unnecessary (Coulson et al., 2019). However, the explicit prediction can be useful for analyzing the accuracy of the DDP.

3. Encoding prior knowledge

In the noise-free setting, there exists a deadbeat observer gain K_{db} that places all poles of $\tilde{A}$ at the origin. By propagating n steps backwards in time through the predictor form in (2), the autoregressive model

$$y_k = \sum_{i=1}^{n} \left(\Xi_i^{(y)} y_{k-i} + \Xi_i^{(u)} u_{k-i} \right) + \Xi_0^{(u)} u_k \tag{7}$$

is derived, with the coefficient matrices $\Xi_i^{(y)} = C\tilde{A}^{i-1}K_{db}$, $\Xi_i^{(u)} = C\tilde{A}^{i-1}\tilde{B}$, and $\Xi_0^{(u)} = D$. The deadbeat observer guarantees that $\|\tilde{A}^p\|_F = 0$ for all $p \geq n$ and the current output is expressed in terms of the current input and past n inputs and outputs.

Lemma 5 (*Coefficient Structure*)**.** *An LTI system can be represented by (7), such that*

(i) *the output coefficient $\Xi_i^{(y)}$ is a diagonal matrix and the vth diagonal entry corresponds to the negative ith coefficient of the characteristic polynomial (CP) of the vth output channel for $i = 1 \ldots n$ and $v = 1 \ldots n_y$, and*

(ii) *for single-input, single-output (SISO) systems, the coefficient $\Xi_i^{(u)}$ corresponds to the ith numerator coefficient of the transfer function*

Proof. We consider a SISO LTI system with the CP

$$\chi(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n.$$

Without loss of generality, we assume that the system is given in observable canonical form by the matrices

$$A_o = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & \cdots & 0 & -a_{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_1 \end{bmatrix}, \quad C_o = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

The observer gain K_{db} that places all poles of $\tilde{A}$ at the origin is

$$K_{db} = \begin{bmatrix} -a_n & -a_{n-1} & \cdots & -a_1 \end{bmatrix}^{\top}$$

and we write the coefficient matrix $\Xi_i^{(y)} = C_o \tilde{A}^{i-1} K_{db}$ as

$$\Xi_i^{(y)} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}^{i-1}}_{\tilde{A} = A_o - K_{db} C_o} \begin{bmatrix} -a_n \\ -a_{n-1} \\ \vdots \\ -a_1 \end{bmatrix}.$$

With the shift property of $\tilde{A}$ it follows that $\Xi_i^{(y)} = -a_i$. By representing the system in observable block-companion form, the proof can be extended to MIMO systems, and the diagonal structure of $\Xi_i^{(y)}$ follows naturally. □

We introduce the shift operator q, such that $y_{k-1} = q^{-1} y_k$, and rewrite the autoregressive model in (7) as

$$\left(I - \sum_{i=1}^{n} \Xi_i^{(y)} q^{-i} \right) y_k = \left(\Xi_0^{(u)} + \sum_{i=1}^{n} \Xi_i^{(u)} q^{-i} \right) u_k.$$

The matrix polynomials defined by $\Xi_i^{(y)}$ and $\Xi_i^{(u)}$ determine the poles and zeros of the system. We assume that a subset of s_p system poles and s_z zeros is known *a priori* and factorize the corresponding matrix polynomials into known and unknown parts

$$\begin{aligned} &\left(I - \sum_{i=1}^{n-s_p} \overline{\Xi}_i^{(y)} q^{-i} \right) \left(I - \sum_{i=1}^{s_p} \tilde{\Xi}_i^{(y)} q^{-i} \right) y_k \\ &= \left(\overline{\Xi}_0^{(u)} + \sum_{i=1}^{n-s_z} \overline{\Xi}_i^{(u)} q^{-i} \right) \left(I + \sum_{i=1}^{s_z} \tilde{\Xi}_i^{(u)} q^{-i} \right) u_k. \end{aligned} \tag{8}$$

Here, $\tilde{\Xi}_i$ and $\overline{\Xi}_i$ represent the known and unknown coefficients, respectively. By Lemma 5, the known output coefficients $\tilde{\Xi}_i^{(y)}$ are diagonal matrices and contain the coefficients of the CP of the known poles. For SISO systems, the input coefficients $\tilde{\Xi}_i^{(u)}$ relate to known zero locations. The known part of the system in (8) can be applied as filters to the input–output data and the filtered data, denoted by $(\cdot)'$, is given by

$$\begin{aligned} u'_k &= \left(I + \sum_{i=1}^{s_z} \tilde{\Xi}_i^{(u)} q^{-i} \right) u_k \\ y'_k &= \left(I - \sum_{i=1}^{s_p} \tilde{\Xi}_i^{(y)} q^{-i} \right) y_k. \end{aligned} \tag{9}$$

To initialize the input and output filters, it requires s_z and s_p past input and output samples, respectively. Furthermore, we define the sparse block-Toeplitz filter matrices

$$Q^{(u)} = \begin{bmatrix} \tilde{\Xi}_{s_z}^{(u)} & \cdots & \tilde{\Xi}_1^{(u)} & I & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tilde{\Xi}_{s_z}^{(u)} & \cdots & \tilde{\Xi}_1^{(u)} & I & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{\Xi}_{s_z}^{(u)} & \cdots & \tilde{\Xi}_1^{(u)} & I \end{bmatrix}$$

and

$$Q^{(y)} = \begin{bmatrix} -\tilde{\Xi}_{s_p}^{(y)} & \cdots & -\tilde{\Xi}_1^{(y)} & I & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\tilde{\Xi}_{s_p}^{(y)} & \cdots & -\tilde{\Xi}_1^{(y)} & I & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\tilde{\Xi}_{s_p}^{(y)} & \cdots & -\tilde{\Xi}_1^{(y)} & I \end{bmatrix}.$$

Equivalently to (9), filtered input–output trajectories are obtained by

$$\begin{aligned} \underline{u}'_{i+s_z, L-s_z} &= Q^{(u)} \underline{u}_{i,L} \\ \underline{y}'_{i+s_p, L-s_p} &= Q^{(y)} \underline{y}_{i,L}, \end{aligned}$$

where $Q^{(u)} \in \mathbb{R}^{n_u(L-s_z)\times n_u L}$ and $Q^{(y)} \in \mathbb{R}^{n_y(L-s_p)\times n_y L}$ have s_z and s_p less block-rows than columns, respectively, which reflects the initialization of the filters. Depending on the length of data trajectories, the size of the filter matrices may vary.

By applying the filters to the implicit predictor in (5), we introduce the physics-informed data-driven predictor (PI-DDP)

$$\begin{bmatrix} Q^{(u)} & \mathbf{0} \\ \mathbf{0} & Q^{(y)} \end{bmatrix} \begin{bmatrix} U_p \\ U_f \\ Y_p \\ Y_f \end{bmatrix} g = \begin{bmatrix} Q^{(u)} & \mathbf{0} \\ \mathbf{0} & Q^{(y)} \end{bmatrix} \begin{bmatrix} \underline{u}_{i,p} \\ \underline{u}_{i+p,f} \\ \underline{y}_{i,p} \\ \underline{y}_{i+p,f} \end{bmatrix}$$

or equivalently with the filtered past data matrices $U'_{p-s_z} \in \mathbb{R}^{n_u(p-s_z)\times(T-L+1)}$ and $Y'_{p-s_p} \in \mathbb{R}^{n_y(p-s_p)\times(T-L+1)}$

$$\begin{bmatrix} U'_{p-s_z} \\ U'_f \\ Y'_{p-s_p} \\ Y'_f \end{bmatrix} g = \begin{bmatrix} \underline{u}'_{i+s_z,p-s_z} \\ \underline{u}'_{i+p,f} \\ \underline{y}'_{i+s_p,p-s_p} \\ \underline{y}'_{i+p,f} \end{bmatrix}. \tag{10}$$

The key difference to the standard DDP in (5) is that future data sequences $\underline{u}'_{i+p,f}$ and $\underline{y}'_{i+p,f}$ of the PI-DDP in (10) are coupled through the filter matrices with past data sequences. Given past input–output data, the implicit PI-DDP gives a unique relationship between future filtered inputs and outputs. Since the remaining system has a reduced order of $n - s_p$, the PE requirement on the filtered input data also reduces to order $p + f + n - s_p$.

4. Output prediction errors

In Sections 2 and 3, we derived exact DDP and PI-DDP under Assumptions 2 and 3. However, practical settings often violate these assumptions. In this section, we extend our analysis by explicitly deriving the output prediction error that arises when these assumptions are relaxed. Considering noise and short past horizons, the data equations in (3) are extended to

$$\begin{aligned} y^*_{i+p,f} &= \underline{y}_{i+p,f} + \Gamma\tilde{A}^p x_i + \mathcal{T}_{(K,I)}\underline{e}_{i+p,f} \\ Y^*_f &= Y_f + \Gamma\tilde{A}^p X_{\hat{i}} + \mathcal{T}_{(K,I)}E_f, \end{aligned} \tag{11}$$

in which $(\cdot)^*$ indicates that Assumptions 2 and 3 no longer hold, x_i is the system's state at time i, and $X_{\hat{i}} = \begin{bmatrix} x_{\hat{i}} & \cdots & x_{\hat{i}+T-L+1} \end{bmatrix}$ contains the initial states of the columns of the data matrix.

4.1. Standard predictions

First, we consider the standard case without prior knowledge and rewrite the implicit DDP from (5) in terms of the data equations in (11):

$$\begin{bmatrix} U_p \\ U_f \\ Y^*_p \\ Y^*_f - \Gamma\tilde{A}^p X_{\hat{i}} - \mathcal{T}_{(K,I)}E_f \end{bmatrix} g = \begin{bmatrix} \underline{u}_{i,p} \\ \underline{u}_{i+p,f} \\ \underline{y}^*_{i,p} \\ \underline{y}^*_{i+p,f} - \Gamma\tilde{A}^p x_i \end{bmatrix}.$$

In order to predict the noise-free output, we assume that $\underline{e}_{i+p,f} = \underline{0}_f$. The matrix Y^*_f contains the available output data for which Assumptions 2 and 3 are relaxed and $\underline{y}^*_{i+p,f}$ corresponds to the true future output. The terms from future noise and convergence of the state observer contribute to the prediction error and are therefore moved to the right-hand side. The explicit output prediction

$$Y^*_f g = \underline{y}^*_{i+p,f} + \underbrace{\mathcal{T}_{(K,I)}E_f g + \Gamma\tilde{A}^p(X_{\hat{i}}g - x_i)}_{\text{error}} \tag{12}$$

is obtained from linear combinations of the output data. The prediction consists of the true output $\underline{y}^*_{i+p,f}$ and two error terms. The first one arises from future innovation noise and relates directly to Assumption 2. The multiplication with the block-Toeplitz matrix $\mathcal{T}_{(K,I)}$ gives the zero-state response from future noise. Since E_f contains zero-mean Gaussian innovation noise, the expected value of the error energy is minimized together with the ℓ_2-norm of g. For the implementation of a DPC scheme, our result suggests an ℓ_2-norm regularization of the column selector g. The effect of such regularization on the implicit DDP is further discussed by Klädtke and Darup (2023).

The second error term arises if the past horizon is too short for the observer to converge, i.e., $\|\tilde{A}^p\|_F \neq 0$, and relates to Assumption 3. The multiplication with the extended observability matrix Γ gives the zero-input response from the state observer error after p time steps.

4.2. Physics-informed predictions

Similarly to the extended data equations in (11), we express future filtered outputs in terms of the outputs for which Assumptions 2 and 3 hold and terms that arise by relaxing these assumptions

$$\begin{aligned} \underline{y}'^*_{i+p,f} &= \underline{y}'_{i+p,f} + Q^{(y)}\Gamma A^{p-s_p}x_i + Q^{(y)}\overline{\mathcal{T}}_{(K,I)}\begin{bmatrix} \underline{e}_{i,p} \\ \underline{0}_f \end{bmatrix} \\ Y'^*_f &= Y'_f + Q^{(y)}\Gamma A^{p-s_p}X_{\hat{i}} + Q^{(y)}\overline{\mathcal{T}}_{(K,I)}\begin{bmatrix} E_p \\ E_f \end{bmatrix}. \end{aligned} \tag{13}$$

By truncating the first $p - s_p$ block-rows of $\mathcal{T}_{(K,I)}$, we define

$$\overline{\mathcal{T}}_{(K,I)} = \begin{bmatrix} CA^{p-s_p-1}K & \cdots & I & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ CA^{f+p-2}K & \cdots & CA^{f+s_p-2}K & \cdots & I \end{bmatrix}.$$

Since the filter matrix $Q^{(y)}$ initializes future filtered outputs with the past s_p filtered outputs, past noise influences the future output predictions. Next, we replace the outputs in the PI-DDP in (10) by the outputs in (13):

$$\begin{aligned} &\begin{bmatrix} U'_{p-s_z} \\ U'_f \\ Y'^*_{p-s_p} \\ Y'^*_f - Q^{(y)}\left(\Gamma A^{p-s_p}X_{\hat{i}} + \overline{\mathcal{T}}_{(K,I)}\begin{bmatrix} E_p \\ E_f \end{bmatrix}\right) \end{bmatrix} g \\ &= \begin{bmatrix} \underline{u}'_{i+s_z,p-s_z} \\ \underline{u}'_{i+p,f} \\ \underline{y}'^*_{i+s_p,p-s_p} \\ \underline{y}'^*_{i+p,f} - Q^{(y)}\left(\Gamma A^{p-s_p}x_i + \overline{\mathcal{T}}_{(K,I)}\begin{bmatrix} \underline{e}_{i,p} \\ \underline{0}_f \end{bmatrix}\right) \end{bmatrix} \end{aligned}$$

and the filtered output prediction follows to:

$$\begin{aligned} Y'^*_f g = \underline{y}'^*_{i+p,f} &+ Q^{(y)}\Gamma A^{p-s_p}(X_{\hat{i}}g - x_i) \\ &+ Q^{(y)}\overline{\mathcal{T}}_{(K,I)}\left(\begin{bmatrix} E_p \\ E_f \end{bmatrix} g - \begin{bmatrix} \underline{e}_{i,p} \\ \underline{0}_f \end{bmatrix}\right). \end{aligned} \tag{14}$$

To unfilter the output prediction, we rewrite the output filter in (9) as

$$\underline{y}'_{i+p,f} = \underline{y}_{i+p,f} - \phi_{up}\underline{y}_{i+p-s_p,s_p} - \phi_{lo}\underline{y}_{i+p,f}, \tag{15}$$

with the upper triangular matrix $\phi_{up} \in \mathbb{R}^{n_y f \times n_y s_p}$ and the lower triangular matrix $\phi_{lo} \in \mathbb{R}^{n_y f \times n_y f}$ given by

$$\phi_{up} = \begin{bmatrix} \tilde{\Xi}_{s_p}^{(y)} & \cdots & \tilde{\Xi}_1^{(y)} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \tilde{\Xi}_{s_p}^{(y)} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \quad \phi_{lo} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \tilde{\Xi}_1^{(y)} & \ddots & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \tilde{\Xi}_{s_p}^{(y)} & \ddots & \tilde{\Xi}_1^{(y)} & \ddots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \tilde{\Xi}_{s_p}^{(y)} & \cdots & \tilde{\Xi}_1^{(y)} & \mathbf{0} \end{bmatrix}.$$

The unfiltered output, derived from (15) is

$$\underline{y}_{i+p,f} = (I - \phi_{lo})^{-1} \underline{y}'_{i+p,f} + (I - \phi_{lo})^{-1} \phi_{up} \underline{y}_{i+p-s_p,s_p}$$

and depends on the filtered output and the past s_p unfiltered outputs for initialization. The unfiltered PI-DDP, denoted by $\hat{(\cdot)}$, is obtained by using the past outputs $\underline{y}^*_{i+p-s_p,s_p}$ for initialization and applying (15) to the output prediction in (14):

$$\begin{aligned} \hat{\underline{y}}_{i+p,f} = \underline{y}^*_{i+p,f} &+ \mathcal{T}_{(K,I)} E_f g + \Gamma \mathcal{K}_{(\tilde{K})} (E_p g - \underline{e}_{i,p}) \\ &+ (I - \phi_{lo})^{-1} Q^{(y)} \Gamma A^{p-s_p} (X_{\hat{i}} g - x_i) \end{aligned} \tag{16}$$

in which $\mathcal{K}_{(\tilde{K})} \in \mathbb{R}^{n \times p n_y}$ is defined, such that

$$(I - \phi_{lo})^{-1} Q^{(y)} \overline{\mathcal{T}}_{(K,I)} = \begin{bmatrix} \Gamma \mathcal{K}_{(\tilde{K})} & \mathcal{T}_{(K,I)} \end{bmatrix}.$$

Due to space limitations, the full simplification is excluded. The prediction in (16) consists of the true output $\underline{y}^*_{i+p,f}$ and three error contributions from future noise, past noise, and the convergence of the state observer.

The error arising from future noise equals that of the standard DDP in (12) and is minimized together with the ℓ_2-norm of the column selector g. However, the PI-DDP in (10) admits smaller norm solutions for g compared to the DDP in (5). This follows from the observation that every solution g to the standard DDP also satisfies the PI-DDP. Moreover, the null space of the filtered data matrix is larger than that of the unfiltered data matrix, which provides additional flexibility to further reduce the norm of the solution g in the PI-DDP.

In contrast to the standard DDP, the prediction of the PI-DDP is affected by past noise realizations. The matrix $\mathcal{K}_{(\tilde{K})}$ determines how past noise is affecting the error of the state observer and multiplying with Γ gives the zero-input response from that error.

The third error contribution accounts for the convergence of the state observer.

Lemma 6 (*Orthogonality of Γ and $Q^{(y)}$*). *Let $Q^{(y)}$ be an output filter matrix designed with a subset of $s_p \leq n$ known poles of an nth-order LTI system. Then, the row space of $Q^{(y)}$ is orthogonal to a s_p-dimensional subspace of Γ, which forms a basis of the free responses associated with the known poles.*

Proof. Consider an nth-order SISO LTI system and the output filter matrix $Q^{(y)}$, derived from $s_p = n$ poles. Since every row of Γ can be expressed by the first row multiplied with a power of the matrix A, it is sufficient to prove that the first row of $Q^{(y)}$ is orthogonal to Γ, i.e.,

$$\begin{aligned} \begin{bmatrix} -\tilde{\Xi}_n^{(y)} & \cdots & -\tilde{\Xi}_1^{(y)} & I & 0 & \cdots & 0 \end{bmatrix} \Gamma = 0 \\ -\tilde{\Xi}_n^{(y)} C - \cdots - \tilde{\Xi}_1^{(y)} C A^{n-1} + C A^n = 0. \end{aligned}$$

With Lemma 5 we write

$$a_n C + \cdots + a_1 C A^{n-1} + C A^n = 0$$

and by the *Cayley–Hamilton Theorem*, every square matrix satisfies its own characteristic polynomial

$$C \left(a_n I + \cdots + a_1 A^{n-1} + A^n \right) = 0.$$

The proof can be extended to MIMO systems and partial information of the poles by representing the system matrix A in Jordan normal form. □

Since the product of $Q^{(y)}$ and Γ occurs in the error in (16), Lemma 6 implies that the convergence error associated with the states of the known poles is mitigated. The space of zero-input responses from those states is orthogonal to the row space of $Q^{(y)}$. If all pole locations are known, it follows that $Q^{(y)}\Gamma = 0$, and the third error term is completely eliminated. As a trade-off, past noise realizations affect the initialization of the output prediction. This trade-off is similar to the difference between a Kalman observer and a deadbeat observer. If the system is given in innovation form, the state estimate of the Kalman observer is independent of past noise realizations similar to the DDP in (12). However, for poles close to the unit circle it may require large past horizons to converge. In contrast, a deadbeat observer has a finite settling time but the state estimate explicitly depends on past noise realizations.

The error expressions in (12) and (16) suggest that prior knowledge is beneficial, if the known poles are located close to the unit circle and the Kalman observer would need long past horizons to converge. Furthermore, the impact of future noise is reduced whenever the PI-DDP admits smaller norm solutions of g than the regular DDP. This is the case when the system is excited close to the eigenfrequencies of the known poles, i.e., when a large part of the input–output trajectories resides in the null space of the filter matrix. While prior knowledge of zero locations, encoded in the filter matrix $Q^{(u)}$, does not explicitly affect the output prediction error, it may lead to smaller ℓ_2-norm solutions of g, thereby reducing the prediction error.

5. Numerical example

We consider the 5th-order model of a floating wind turbine given by Hegazy et al. (2025), which relates the pitch angle of the rotor blades to the rotor speed

$$G(z) = \frac{-0.052(z-\zeta_1)(z-\zeta_2)(z-\zeta_3)(z-\zeta_4)}{(z-\lambda_1)(z-\lambda_2)(z-\lambda_3)(z-\lambda_4)(z-\lambda_5)},$$

with the following poles and zeros:

$$\begin{aligned} \lambda_{1,2} &= 0.98 \pm 0.16i, \quad \lambda_{3,4} = -0.82 \pm 0.45i, \quad \lambda_5 = 0.72, \\ \zeta_{1,2} &= 0.99 \pm 0.16i, \quad \zeta_{3,4} = -0.84 \pm 0.47i, \end{aligned}$$

and a minimal representation (1) given by A, B, and C. The Kalman gain K and the covariance Σ_e are calculate from the DARE with a process noise covariance of 10^{-2} and a measurement noise covariance of 10^{-1}. It is reasonable to assume prior knowledge of the dominant poles at low frequencies, which can be obtained through first-principles modeling. Fig. 1 shows the mean and standard deviation of the DDP and PI-DDP with prior knowledge of the conjugated pole pair $\lambda_{1,2}$ across 1000 noise realizations, sampled from a Gaussian distribution, in blue and red, respectively. The filter matrix $Q^{(y)}$ is calculated from the CP of the known poles with $\tilde{\Xi}_1^{(y)} = 1.96$ and $\tilde{\Xi}_2^{(y)} = -0.986$. The column selector g is calculated by the Moore–Penrose inverse similar to (6).

The true output is represented by a dashed yellow line. The first $p = 8$ data samples are used to initialize the system and the future horizon length is $f = 20$. The data matrices are constructed from $T = 500$ input data, uniformly sampled from the interval $[-100 \;\; 100]$. At the time of prediction, the system is excited within the range of the data at a frequency close to the eigenfrequency of the known poles, i.e, 40^{-1} of the sampling frequency.

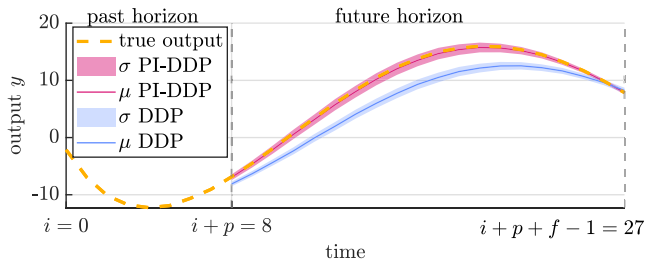

Fig. 1. Mean and standard deviation for output predictions of a floating wind turbine and 1000 noise realizations.

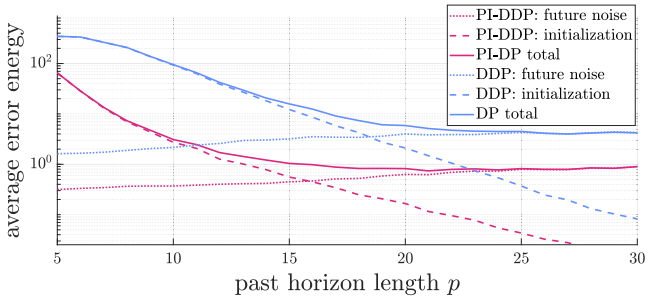

Fig. 2. Error energy for varying past horizon lengths.

The PI-DDP shows improved performance compared to the standard DDP. While the standard deviation is slightly larger, the mean value aligns with the true output.

Fig. 2 shows the average prediction error energy corresponding to the predictions in Fig. 1 for varying past horizon lengths.

The solid lines represent the total prediction error, the dotted lines the error from future noise realizations, and the dashed lines the error arising from the initialization. For short past horizon lengths, the error mostly arises from the initialization of the system. Here, the PI-DDP achieves better results. For increasing past horizon lengths, both initialization errors approach zero and the difference is mainly driven by the error arising from future noise. It is consistently smaller for the PI-DDP compared to the DDP, caused by a smaller ℓ_2-norm solution of g. Since the system is excited close to the eigenfrequency of the known poles, the ℓ_2-norm of the vector g is significantly smaller for the PI-DDP compared to the DDP. Moreover, the remaining unknown states are not excited and the initialization errors corresponding to these states are small. If the system is excited at eigenfrequencies far away from the known poles, the accuracy of the prediction increases less and the error arising from the initialization of the unknown dynamics might increase.

6. Conclusion

We presented a method to incorporate prior system knowledge into data-driven models by filtering input–output data using known pole or zero locations. Prediction errors with and without prior knowledge are explicitly derived. While prior knowledge of zero locations does not directly affect the output prediction, prior knowledge of pole locations mitigates the convergence error of the corresponding initial state and reduces the impact of future noise. Thus, prior knowledge of poles close to the unit circle allows to use shorter past horizon lengths.

Future work includes to further investigate the impact of prior knowledge on the initialization of the remaining unknown dynamics and on the reduced PE requirements on the input data. Additionally, the proposed method offers possibilities to include prior knowledge into various data-driven predictive control methods such as DeePC and SPC (Coulson et al., 2019; Favoreel et al., 1999).

CRediT authorship contribution statement

Fritz A. Engeln: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Jan-Willem van Wingerden:** Writing – review & editing, Supervision, Conceptualization. **Timm Faulwasser:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Berberich, J., & Allgöwer, F. (2020). A trajectory-based framework for data-driven system analysis and control. In *2020 European control conference* (pp. 1365–1370). http://dx.doi.org/10.23919/ECC51009.2020.9143608.

Coulson, J., Lygeros, J., & Dörfler, F. (2019). Data-enabled predictive control: In the shallows of the DeePC. In *Proceedings of the 18th European control conference* (pp. 307–312). http://dx.doi.org/10.23919/ECC.2019.8795639.

Dörfler, F. (2023). Data-driven control: Part one of two: A special issue sampling from a vast and dynamic landscape. *IEEE Control Systems Magazine*, *43*(5), 24–27. http://dx.doi.org/10.1109/MCS.2023.3291624.

Favoreel, W., De Moor, B., & Gevers, M. (1999). SPC: Subspace predictive control. In *Proceedings of the 14th world congress of the international federation of automatic control* (pp. 4004–4009). http://dx.doi.org/10.1016/S1474-6670(17)56683-5.

Hegazy, A., Naaijen, P., & van Wingerden, J. W. (2025). Control design for floating wind turbines. *Wind Energy Science Discussions*, 1–35. http://dx.doi.org/10.5194/wes-2025-68.

Klädtke, M., & Darup, M. S. (2023). Implicit predictors in regularized data-driven predictive control. *IEEE Control Systems Letters*, *7*, 2479–2484. http://dx.doi.org/10.1109/LCSYS.2023.3285104.

Li, D., Dong, H., & Song, Z. (2024). Physics-informed data-enabled predictive control for regulating mixed traffic flows. In *2024 IEEE transportation electrification conference and expo*. http://dx.doi.org/10.1109/ITEC60657.2024.10598889.

Ljung, L. (1999). *System identification: theory for the user* (2nd ed.). Upper Saddle River, NJ, USA: Prentice Hall.

Rawlings, J., Mayne, D., & Diehl, M. (2017). *Model predictive control: theory, computation, and design* (2nd ed.). Nob Hill Publishing.

Willems, J., Rapisarda, P., Markovsky, I., & De Moor, B. (2005). A note on persistency of excitation. *Systems & Control Letters*, *54*(4), 325–329. http://dx.doi.org/10.1016/j.sysconle.2004.09.003.

van Wingerden, J. W., Mulders, S., Dinkla, R., Oomen, T., & Verhaegen, M. (2022). Data-enabled predictive control with instrumental variables: the direct equivalence with subspace predictive control. In *2022 IEEE 61st conference on decision and control*. http://dx.doi.org/10.1109/CDC51059.2022.9992824.

Yang, H., & Li, S. (2015). A data-driven predictive controller design based on reduced Hankel matrix. In *2015 10th Asian control conference* (pp. 1–7). http://dx.doi.org/10.1109/ASCC.2015.7244723.

Zadeh, L. (1956). On the identification problem. *IRE Transactions on Circuit Theory*, *3*(4), 277–281. http://dx.doi.org/10.1109/TCT.1956.1086328.

Zieglmeier, S., de Badyn, M. H., Warakagoda, N., Krogstad, T., & Engelstad, P. (2025). Semi-data-driven model predictive control: A physics-informed data-driven control approach. arXiv:2504.00746.