

Received December 7, 2021, accepted December 27, 2021, date of publication January 6, 2022, date of current version January 13, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3140758

Task-Relevant Encoding of Domain Knowledge in Dynamics Modeling: Application to Furnace Forecasting From Video

BRENDAN MICHAEL¹, AKIFUMI ISE², KAORU KAWABATA²,
AND TAKAMITSU MATSUBARA¹, (Member, IEEE)

¹Robot Learning Laboratory, Division of Information Sciences, Nara Institute of Science and Technology, Nara 630-0192, Japan

²Hitachi Zosen Corporation, Osaka 559-8559, Japan

Corresponding author: Brendan Michael (brendan.michael@is.naist.jp)

This work was supported in part by the Hitachi Zosen Corporation, Japan.

ABSTRACT Waste incineration plants are complex dynamical systems that rely on expert human operators to maintain steady combustion, by observing real-time in-chamber video feeds. Real-time plant forecasting provides vital operational support in decision making, and applying machine learning to automatically learn dynamics forecast models from video feeds is an attractive means to realise this. However, learning complex dynamics in systems that requires cost-efficiency remains an open research problem. Specifically, modelling plant dynamics in real-time is challenging due to uncertainties caused by inhomogeneous waste inputs, requiring complex learning that impedes real-time modelling. To address this, this paper presents a real-time data-driven framework for generating video forecasts, by incorporating *task-relevant domain-knowledge*, during learning. Specifically, this method combines dynamics modelling and forecasting using *dynamic mode decomposition*, with *Fourier transformations* informed by expert operator heuristic knowledge for encoding task-relevant frequency information inside the learning process. Experiments in this paper demonstrate that the proposed framework captures intuitive physical aspects of the underlying physiochemical process, with a greatly reduced computational runtime in comparison to standard approaches, allowing for application in real-time domains. Forecasted video predictions are accurate over short time horizons, and capture important system characteristics over longer time periods.

INDEX TERMS Dynamic mode decomposition, forecasting, Fourier transforms, machine learning, video signal processing, waste handling.

I. INTRODUCTION

Municipal industrial waste-incineration plants are an increasing popular alternative to land-fills [1] for waste disposal and energy recuperation [2]. While prevention and minimization is key to waste-management, international policy objectives highlight the need for improving installation environmental performance [3], often via innovative technologies and automation [4]. Incineration automation is particularly desired to ensure consistent environmental and energy outcomes [5].

A key operational objective in waste-incineration automation, is real-time forecasting of long-term behavioural trends

The associate editor coordinating the review of this manuscript and approving it for publication was Yi Zhang¹.

within a combustion chamber (*e.g.*, combustion growth/decay over a period), to assess stability and ensure consistency of energy output. Currently, manual forecasting is performed by operators, using in-chamber video feeds to make speculative estimates of changes in the dynamical process, supported by simple automated solutions such as image classification [6]. Automated methods for modelling combustion, and generating *real-time videos forecasts* of future chamber behaviour, can help support operator decision making, and provide quantitative combustion assessments.

Traditional plant modelling and forecasting (*e.g.*, in manufacturing or energy production domains) often uses first-principle analytical models with detailed understanding of the known physiochemical process and prior identification of control parameters and processes laws [7]. However,

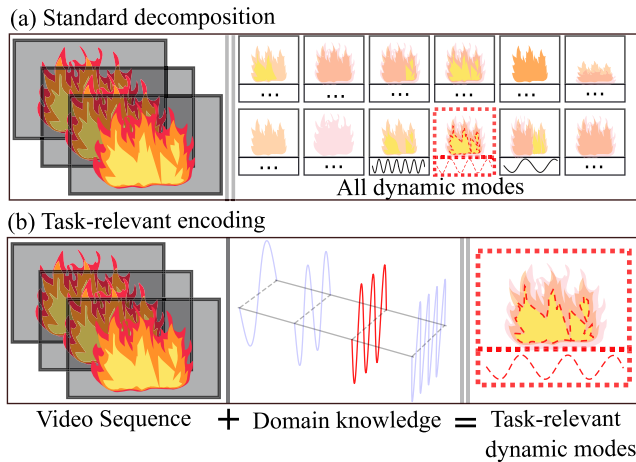


FIGURE 1. (a) Standard decomposition extracts all dynamic modes from a video sequence, even if only specific task dynamics (red dashed) are required, leading to high computational cost and subject to instabilities. (b) By encoding task-relevant domain knowledge (e.g., known frequencies of interest) into decomposition (red sinusoidal), cost is reduced and learning stabilised, resulting in only the extraction of relevant dynamic modes.

waste-incineration plants presents challenges not traditionally addressed, as combustion involves complex to model dynamical behaviour, with high uncertainty about the composition of the waste input. Specifically, inhomogeneous waste inputs [8] (e.g., uncertainty in qualitative characteristics such as calorific and moisture content of waste [9], [10]) cannot be determined *a priori*, and results in challenging to model combustion events, e.g., flame stagnation.

In the absence of detailed analytical models addressing uncertainties, an intuitive forecasting approach is to use *data-driven machine learning*, to automatically learn models of combustion behaviour, from the video feed. Specifically, in the context of forecasting long-term behavioural trends, *spatio-temporal decomposition* methods are well-suited to explicitly decompose high-dimensional time-series measurements into coherent spatial structures and corresponding time dynamics [11], and as such can be used to extract *task relevant* dynamics from video feeds, i.e., extract components of a video associated with known *dynamical frequencies*.

However, naive application of spatio-temporal methods (e.g., dynamic mode decomposition (DMD) [12]), is challenging to apply to real-time video systems, due to the high computational cost and complexity. Real-time implementations of DMD with improved computational benefits (e.g., Streaming DMD [13]), unfortunately, necessitate projecting measurements to a significantly low-rank subspace, which is challenging in situations involving non-autonomous (input-driven) uncertain dynamics, such as waste inhomogeneity. In addition, complexity and uncertainty in dynamics can result in unstable learnt dynamics, requiring complex modelling often via deep learning [14], [15].

As such, practical means for applying dynamical learning techniques in domains that require both cost-efficiency and intuitive modelling remains an open research question.

To address this, this paper proposes the Task-Relevant Encoding of Domain Knowledge (TREK) framework for the real-time modelling and forecasting of dynamical systems from high-dimensional, long-term measurements with uncertainties, by incorporating *task-relevant domain knowledge* into the learning process. Specifically, cost-efficiency is achieved by encoding expert heuristic knowledge about the dynamical system (e.g., combustion decay rates) as part of the decomposition process (FIGURE 1), thereby learning only dynamic modes relevant to the task. This approach exploits the relationship between Fourier transformations and dynamic mode decomposition, to preprocess measurements with uncertain dynamics into a decomposition amenable form. Specifically, a Fourier spectrum manipulation is used to extract *task relevant* frequency components from a sequence, which are then mapped back to the input space as a low-rank, task relevant approximation. As such, the Fourier transformation converts high-rank measurements into a low-rank coordinate system amenable to decomposition with DMD, while remaining agnostic about the complexity or uncertainty of the underlying system. By exploiting this relationship, dynamics modes can be quickly learnt that capture intuitive physical interpretations of task relevant combustion behaviour. Results in this paper demonstrate this framework learns interpretable models that capture important physiochemical characteristics of combustion, and can generate video forecasts in real-time, which remain accurate within a reasonable forecast horizon.

The key contributions of this paper are as follows:

- (i) *Identification of limitations*: This paper presents first evidence that current spatial-temporal methods are unable to model complex dynamics in the presence of uncertainties in real-time, severely limiting applications in real-world scenarios.
- (ii) *Novel real-time framework*: A computationally efficient framework, is presented to address these limitations, by exploiting spectral manipulation to reduce learning complexity. This framework is agnostic to the underlying dynamics complexity, and can provide robust and real-time dynamical learning for real-world systems.
- (iii) *Evaluation and Results*: This framework is evaluated in comparison to commonly used data-driven dynamics learning methods, in (a) a high-dimensional signal simulation example, and (b) in a dynamics learning task using real in-furnace video data. The framework extracts intuitive dynamical characteristics of the furnace in real-time, with potential for application to other fields that require time-efficient dynamical learning.

The remainder of the paper is organised as follows: §II presents a background of operational control for industrial waste incineration plants and data-driven dynamical modelling approaches, §III outlines prior work in extracting intuitive dynamic models via spatio-temporal decomposition, §IV presents the novel approach of exploiting spectral manipulation for performing decompositions in real-time. §V presents an experimental evaluation of the approach, both in

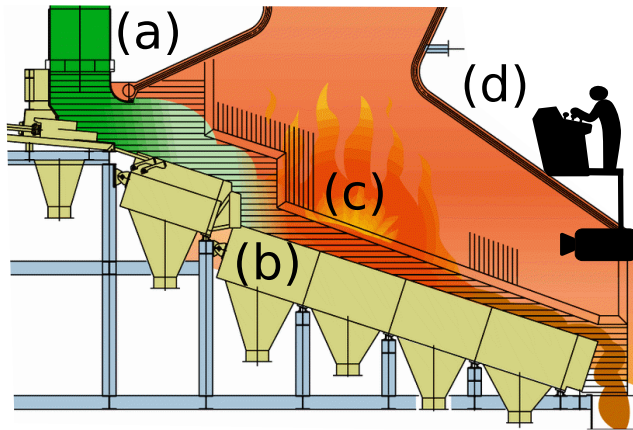


FIGURE 2. Combustion chamber processing: (a) Waste input, (b) Transportation via actuated grate agitation, (c) Combustion, (d) Operator control.

an intuitive simulation example, and applied to an long-term industrial furnace video feed. Finally, a discussion and future applications is presented in §VI and §VII.

II. RELATED WORK

In the following section, the role of human-expert knowledge in industrial furnace control is identified, as well as prior state-of-the-art machine learning methods for modelling dynamic behaviour.

A. INDUSTRIAL FURNACE CONTROL AND FORECASTING

To motivate the proposed approach, consider a general overview [16] of the standard waste-incineration process (FIGURE 2), whereby: (a) waste is inserted to a feeder via a separate crane [8], and released into the chamber, (b) actuated staircase grates slowly agitate and transport waste [17], (c) ignition and combustion is executed in a central region, (d) monitoring and control is typically performed by PID controllers using in-furnace sensors [18], with refined manual control performed by human operators monitoring in-furnace video cameras.

While this process seems relatively straightforward and autonomous, operators are vital in maintaining combustion stability by monitoring in-chamber video and periodically overriding process controls (e.g., actuation speed or airflow) to ensure consistent uniform combustion on the staircase. A key factor that affects control decisions, is the large lag time inherent in the system, often of the order of tens of minutes between observations and system state changes. Specifically, this delay is caused by numerous inherently slow processes, e.g., : (i) slow necessary agitation of the solid waste along the staircase, to ensure consistent and complete combustion, or (ii) variance in ignition and combustion dynamics caused by waste moisture and composition.

B. OPERATOR DOMAIN KNOWLEDGE

To maintain stable combustion given these delays, it is imperative that operators accurately anticipate changes in

combustion behaviour *ahead of time*, and provide control inputs accordingly. As such, operators develop forecasting heuristics to foresee long-term furnace state changes, thereby anticipating changes. This approach enables a generalised comprehensive outlook on the slow, delay induced chamber dynamics, while being unconcerned with insignificant short-term events such a local waste movement. Specifically, operators are attentive to factors that induce long-term behaviour seen in the video feed over a period of 5 – 20 minutes, e.g., : (i) whether waste fails to incinerate completely within the time period, indicating either an inadequate or oversaturated feed rate, (ii) changes in the fire strength (e.g., change in colour, peak height or distribution of flames).

Specifically, these heuristics can be described as *task-relevant domain knowledge*, and are directly related to specific aspects of furnace state behaviour, and as such will vary depending on the combustion scenario. This knowledge of long-term dynamical trends is quantified and collated as the set of *task-relevant domain knowledge*, $\tau = \{\tau_1, \tau_2, \dots, \tau_D\} \in \mathbb{R}^D$, that the operator employs when forecasting (e.g., the approximate time for waste to traverse the system is $\tau_d = 10$ minutes). Operators compare this knowledge against current observations to inform which control operations should be applied ahead of time to ensure stability. While this knowledge is only approximate (given variations in waste content), it provides key temporal information for forecasting, and is crucial for informing operational decisions in advance of furnace state changes.

C. DATA-DRIVEN MODELLING

To help support operator decisions, automating this speculative heuristic forecasting can provide numerous operational advantages, including: (i) using forecasting predictions of expected combustion changes ahead of time to ensure furnace efficiency, (ii) a quantitative means to assessing different control strategies, (iii) the ability to incorporating models into simulations for offline operator training.

In the absence of detailed understanding of the physio-chemical system and lack of explicit control laws due to waste-inhomogeneities, an appealing approach to automating forecasting is to use machine learning methods to model the dynamics from the available video data, and use learnt models to generate *videos of forecasted future behaviour* [19]. While traditional stochastic models are often used for forecasting (e.g., vector autoregressive moving average (VARIMA) [20]) these do not generalise well to high-dimensional complex inputs [21]. As such, prior work forecasting industrial furnace dynamics [22], [23] is generally limited to multivariate, but low-dimensional forecasting, and is unsuitable for generating high-dimensional video forecasts. To address this, state-of-the-art methods for generating video forecasts are often deep-learning based [24], [25], and specifically for learning combustion dynamics this includes recurrent and convolutional neural network approaches [19], [26]. These methods are subject to the standard deep-learning challenges,

e.g., lengthy model training time and large sample requirement, however, an additional key deficiency is the *lack of interpretability* in models (known as the *black box problem*). Learning interpretable models is of paramount importance [27], not only for furnace safety and integration with existing tools [28], but also to necessitate the desired aim of extracting frequency components corresponding to the long-term dynamics, while ignoring short-term anomalies. Standard dynamical modelling approaches that learn black box models of the dynamics from video [19], require just as complex post-hoc analysis to derive physical intuitive meanings of the model structure. As such, a key challenge is designing a modelling framework that provides both accurate generated video forecasts, as well as intuitive coherent outputs that explain the underlying physiochemical processes in terms of the dynamical behaviour.

III. PRELIMINARIES

To address the problems of dynamics modelling from video, a pertinent approach is *spatio-temporal decomposition*, which explicitly extracts intuitive spatial structures and corresponding time dynamics from high-dimensional measurement time-series. These structures provide clear physical interpretations of the underlying process, well-suited to the application goals of data-driven plant modelling. As opposed to traditional machine learning methods which require numerous samples for generalised modelling, decomposition requires only a single sample sequence in which to infer dynamics, making this vastly more computationally suitable to data-limited environments. Decomposition has previously been applied to many tasks involving complex dynamical systems including jet [29] and nuclear reactor analysis [30], soft robot identification [31], epidemiology [32], and even financial trading [33].

Specifically, given K time-sequential measurements $\mathbf{X} \in \mathbb{C}^{P \times K}$ from a dynamic system of the form $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, t)$, *spatio-temporal decomposition* describes this system as a superposition of empirically estimated basis vectors, known as the dynamic modes [34]. While outside the scope of this paper, the interested reader is referred to [35] for a formal rigorous definition, and [11] for illustrative concrete examples.

In the following, a brief outline of *Dynamic mode decomposition* (DMD) [12] is presented for extracting these modes. For detailed derivations and additional implementations, the reader is referred to the following well-established texts [11], [34], [35].

A. DYNAMIC MODE DECOMPOSITION

Dynamic mode decomposition is a common approach to estimating dynamic modes, by assuming that dynamics are driven by a linear operator $\mathbf{A} \in \mathbb{C}^{P \times P}$, where:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k, \text{ for } k = 1, \dots, K - 1. \quad (1)$$

This is commonly described as finding the operator \mathbf{A} that minimises the error between two snapshot matrices [11]:

$$\|\mathbf{X}_2 - \mathbf{A}\mathbf{X}_1\|_F, \quad (2)$$

which can be solved in a least-squares sense by:

$$\mathbf{A} = \mathbf{X}_2\mathbf{X}_1^\dagger, \quad (3)$$

where $\mathbf{X}_1 = [\mathbf{x}_1, \dots, \mathbf{x}_{k-1}]$, $\mathbf{X}_2 = [\mathbf{x}_2, \dots, \mathbf{x}_k]$, \dagger denotes the Moore-Penrose pseudoinverse, and F is the Frobenius norm. In the context of video processing, a measurement \mathbf{x}_k can be a tall vector consisting of RGB values for each pixel in a frame.

Given this linear dynamics form, it is assumed that measurements at time t_k are represented as a Fourier-like expansion of \mathcal{M} spatio-temporal structures $\phi_m \in \mathbb{C}^P$ (also known as dynamic modes), characterised by growth rates δ and frequencies ω [36]:

$$\mathbf{x}(t_k) = \sum_{m=1}^{\mathcal{M}} \phi_m e^{(\delta_m + i\omega_m)(k-1)(\Delta t)} b_m, \quad (4)$$

where b_m is the amplitude of the corresponding mode. This is more commonly expressed in discrete form as [11]:

$$\mathbf{x}_t = \Phi \Lambda^t \mathbf{b} = \sum_{m=1}^{\mathcal{M}} \phi_m \lambda_m^t b_m \quad (5)$$

where $\Lambda = \lambda \mathbf{I}_{\mathcal{M}} \in \mathbb{C}^{\mathcal{M} \times \mathcal{M}}$ is a matrix of eigenvalues that define the dynamical time evolution of the dynamic modes respectively as growth/decay and oscillation frequencies [34]. As such, the goal of DMD is to find these \mathcal{M} dynamic modes via an eigendecomposition of \mathbf{A} .

Practically, finding this eigendecomposition is computationally challenging, and commonly a low-rank approximation $\tilde{\mathbf{A}} \in \mathbb{C}^{\mathcal{N} \times \mathcal{N}}$ is sort after instead. Specifically, proper orthogonal decomposition (POD), also known as principal component analysis (PCA), is first performed to find a low-rank subspace spanned by the first \mathcal{N} highest energy *spatial modes*. POD *spatial modes* are not the same as the spatial structures of DMD *dynamic modes*, as POD modes are extracted solely based on sequence energy ignoring dynamics. Given these highest energy spatial modes $\tilde{\mathbf{U}} \in \mathbb{C}^{P \times \mathcal{N}}$, a low-rank approximation of the data $\tilde{\mathbf{X}}$ is given by the singular value decomposition [37]:

$$\tilde{\mathbf{X}} = \tilde{\mathbf{U}} \tilde{\Sigma} \tilde{\mathbf{V}}, \quad (6)$$

where Σ is a diagonal matrix of the singular values of \mathbf{X} , and \mathbf{U}, \mathbf{V} are matrices of the left and right singular vectors respectively.

Selecting the number \mathcal{N} of columns of \mathbf{U} to use as the basis $\tilde{\mathbf{U}}$ is often performed by evaluating the singular values in Σ against some threshold, and the approach taken in this paper follows the one in [36]:

$$\frac{\sigma_p^2 + \dots + \sigma_{\mathcal{N}}^2}{\sigma_1^2 + \dots + \sigma_{\mathcal{N}}^2} \leq \varepsilon_1, \quad (7)$$

where σ^2 are the singular values of \mathbf{X} sorted in decreasing order, and ε_1 is a tunable threshold.

Given low-rank POD modes $\tilde{\mathbf{U}}$, this defines the subspace on which the low-rank operator $\tilde{\mathbf{A}}$ evolves the dynamics:

$$\tilde{\mathbf{x}}_{k+1} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}_k, \quad (8)$$

and the goal becomes finding an eigendecomposition of $\tilde{\mathbf{A}}$. Note, commonly it is assumed there is a one-to-one correspondence between the number of spatial and spectral modes, *i.e.*, that the number of POD modes (the spatial complexity \mathcal{N}) is equal to the number of dynamic modes (the spectral complexity \mathcal{M}). For further details on dealing with situations not of this form, see [36].

B. STREAMING DMD

In the context of real-time systems, DMD can be prohibitively computationally expensive. Specifically, for a measurement sequence $\mathbf{X} \in \mathbb{C}^{\mathcal{P} \times \mathcal{K}}$, computing the low-rank basis (6) involves a singular value decomposition (see TABLE 1), which is computationally infeasible for real-time systems involving high-resolution, long-term video sequences where both terms are very large.

Streaming DMD (SDMD) [13] is well suited for large values of \mathcal{P} or \mathcal{K} , by bypassing the need for computationally expensive SVD operations to find the low-rank subspace. Specifically, an iterative Gram-Schmidt approach to low-rank approximation of (6) is used, which is shown to be equivalent to standard DMD and the interested reader is referred to derivations in [13].

C. EXTRACTING TASK-RELEVANT DYNAMIC MODES

Given a learnt decomposition approximating the *global dynamics* (§III-A), further analysis can be performed to extract only modes associated with known frequencies. Prior domain knowledge τ of the system §II-B can be used to select only task-relevant dynamics, known as *task-relevant dynamic modes*, from the global model, in a *post-hoc* manner *i.e.*, $\{\Phi_\tau, \Lambda_\tau\} \in \{\Phi, \Lambda\}$. Reconstructing a dynamical system using only these task-relevant modes, as outlined in FIGURE 3, is formally given as the linear reconstruction of these task-relevant terms:

$$\mathbf{x}_t = \Phi_\tau \Lambda_\tau^t \mathbf{b}_\tau = \sum_{d \in \tau} \phi_d \lambda_d^t b_d. \quad (9)$$

In the context of long-term furnace forecasting, these task-relevant modes are dynamical modes associated with task-relevant changes in the furnace dynamics, *e.g.*, long-term (low frequency) changes in combustion. In summary, this approach first decomposes the entirety of the *global dynamics* with DMD eigendecomposition, and then performs *post-hoc* extraction of task-relevant frequencies by using domain knowledge τ to identify *locally* relevant components.

D. LIMITATIONS

However, while *post-hoc* extraction (§III-C) of dynamic modes via either standard or streaming-based DMD is

appealing, these methods suffer from a number of key problems that limit application in complex real-time systems such as furnace modelling, specifically:

- (i) Inefficiency of decomposition: As outlined in §III-C and FIGURE 3, the standard DMD approach first decomposes the global dynamics, and then selects the task-relevant dynamic modes based on consulting τ . This *all-inclusive* approach introduces inefficiencies, and does not allow for selective decomposition of only relevant dynamic modes based on τ . Alternative formulations such as Bayesian priors [39], [40] can be used to bias learning in favour of known task-knowledge, however these often involve computationally expensive Monte Carlo sampling making them ill-suited to this domain.
- (ii) Uncertain dynamics: In the context of furnace video decomposition, the underlying physiochemical combustion process are non-autonomous, with inhomogeneous input waste driving the dynamics. While the DMD formulation assumes a linear dynamical system (4), the relationship to Koopman spectral analysis [34], [41], [42] allows for learning non-linear dynamical systems, either via approximation to a linear system with the standard DMD approach or the use of embedding functions to explicitly linearise the dynamics in a higher-dimensional space [14], [43]. However, linear approximation can be unstable and provide poor forecasting, and finding appropriate embedding functions for real-world systems is challenging, often requiring complex deep learning [14], [15] for relatively simple non-linear systems.
- (iii) Computation speed of Streaming DMD: While iterative approaches to DMD greatly reduces the computational complexity (as seen in TABLE 1), this speed benefit is conditional on $\mathcal{N} \ll \min(\mathcal{K}, \mathcal{P})$, *i.e.*, there is a low-rank subspace that can be easily found via iteration. In the context of real-world systems, it is often not easy to find this low-rank subspace [44], with factors such as noise [36] or rank-deficiency [34] requiring either a large \mathcal{N} , or spurious results requiring alternative (computationally expensive) DMD formulations such as delay-embeddings [36].

As such, real-time modelling of complex systems such as furnace dynamics is challenging even with iterative streaming approaches. The uncertainty of the system introduces an inherent contradiction, requiring both speed, and computationally expensive methods to overcome.

IV. METHOD

To address the problem of real-time learning of dynamical systems, this paper presents a method for approximating uncertain dynamical systems within an efficient low-rank framework, by exploiting the relationship between spectrum manipulation with Fourier transformations and DMD, as outlined in FIGURE 3.

TABLE 1. Computational complexity: (P)reprocessing, (I)teration, (M)ode computation, (B)ound.

	DMD	SDMD [13]	FFT	TREK
(P)	$\mathcal{O}(\min(\mathcal{P}K^2, \mathcal{P}^2K))$	–	$\mathcal{O}(K \log K)$	$\mathcal{O}(K \log K)$
(I)	–	$\mathcal{O}(\mathcal{P}\mathcal{N})$	–	$\mathcal{O}(\mathcal{P}\mathcal{N})$
(M)	$\mathcal{O}(\mathcal{N}^3)$ [38]	$\mathcal{O}(\mathcal{P}\mathcal{N}^2)$	–	$\mathcal{O}(\mathcal{P}\mathcal{N}^2)$
B	$\mathcal{O}(\min(\mathcal{P}K^2, \mathcal{P}^2K) + \mathcal{N}^3)$	$\mathcal{O}(K\mathcal{P}\mathcal{N} + \mathcal{P}\mathcal{N}^2)$	$\mathcal{O}(K \log K)$	$\mathcal{O}(K\mathcal{P}\mathcal{N} + \mathcal{P}\mathcal{N}^2 + K \log K)$

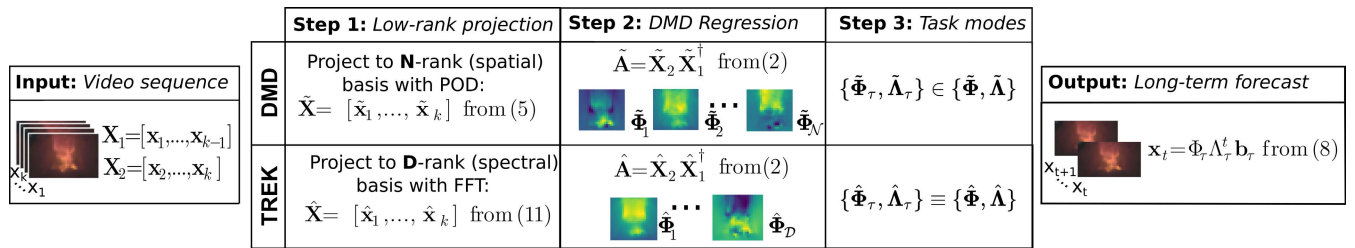


FIGURE 3. Framework for decomposing video sequence and forecasting long-term trends. **Input:** Video sequence of K frames. **(Top) Standard DMD:** Step 1) Low-rank projection computes \mathcal{N} POD bases, Step 2) DMD regression characterises the entire global dynamics, Step 3) Post-processing extracts \mathcal{D} task-relevant dynamic modes based on prior knowledge τ . In contrast, **(Bottom) TREK:** Step 1) Preprocesses via a low-cost FFT to find \mathcal{D} bases, with one-to-one correspondence to terms in τ , and Step 2) DMD characterises only \mathcal{D} terms related to the task-relevant dynamics, with no need for additional post-processing. **Output:** Given task-relevant dynamic modes, forecasts are computed for any t -timesteps into the future.

A. OVERVIEW

As a high-level overview, this framework (i) uses a discrete Fourier transformation as an alternative to POD preprocessing, to instead decompose the sequence into a set of spectral Fourier components, (ii) uses spectrum manipulation to extract only task-relevant frequencies defined by τ , thereby approximating a low-rank spectral subspace of rank $\mathcal{D} < \mathcal{N}$, (iii) performs streaming DMD on data in this low-rank subspace to learn real-time decomposition models.

In the context of DMD, this approach could be seen analogous to the standard preprocessing approach of finding \mathcal{N} maximum energy spatial (POD) modes (6), except in this case the aim is to find \mathcal{D} frequency spectral (FFT) modes defined by the task-knowledge τ . As such, both the computational advantages of FFT over POD, and the ability to selectively extract relevant Fourier terms, enables this framework to be both computationally efficient and embeddable with task-relevant knowledge.

B. METHOD

To illustrate this approach, without loss of generality consider the standard application of a discrete Fourier transform (DFT) to a one-dimensional time-series $\mathbf{x} \in \mathbb{C}^K$ sampled at rate f_s . A discrete Fourier transform decomposes this sequence into $J = K$ frequency components, with sample frequencies $\mathbf{s} = [0, 1, 2, \dots, (K - 1)]/f_s$, giving Fourier terms $\mathbf{y} \in \mathbb{C}^J$. Each term y_j is expressed as a weighted summation of all elements in \mathbf{x} , with a corresponding trigonometric dynamics [45]:

$$\begin{aligned}
 \mathbf{y} := \{y_k\}_{k=1}^K &= \sum_{j=0}^{J-1} x_j [\cos(2\pi kj/J) - i \sin(2\pi kj/J)], \\
 &= \sum_{j=0}^{J-1} x_j e^{-i2\pi kj/J} \quad (10)
 \end{aligned}$$

with inverse transformation:

$$\mathbf{x} := \{x_k\}_{k=1}^K = \frac{1}{\sqrt{J}} \sum_{j=0}^{J-1} y_j e^{i2\pi kj/J} \quad (11)$$

As such, (10) shows any sequence can be approximated into J discrete terms y_j , with each term driven by a complex cosine/sine pair. In the context of DMD, this has a clear relationship to the linear dynamical system in (4), and as such this sequence is perfectly expressible with a linear operator with J complex eigenvalues. This relationship has previously been analysed in [46], and has been shown that under certain conditions applying DMD to mean-subtracted linearly independent datasets is equivalent to DFT. In addition, this is a trivial solution to Koopman’s theory in J -dimensional space [47], as such guaranteeing the sequence is decomposable.

However, while (10) states that the sequence is driven by a linear dynamical system in an J -dimensional-space (equivalent to a K -dimensional-space as all terms y_j are used), this is not practically useful, as DMD requires that there is a low-rank (linear) subspace $\mathcal{N} < K$ in which to drive the $\tilde{\mathbf{A}}$ linear operator (8) along.

To address this, note that these Fourier basis are orthogonal, and therefore the inverse transformation (11) using only a subset of $\mathcal{D} < J$ terms will also remain linear. Performing the inverse transformation using \mathcal{D} terms results in a new measurement sequence perfectly expressible in a \mathcal{D} -dimensional space. As such, this approach is amenable to using the task-relevant domain knowledge $\tau \in \mathbb{R}^{\mathcal{D}}$ to select only a subset of those coefficients relevant to the task at hand to form a spectrally-rank-reduced sequence $\hat{\mathbf{x}}_\tau \in \mathbb{C}^K$.

Specifically, given task-relevant frequency information $\tau = \{\tau_1, \tau_2, \dots, \tau_{\mathcal{D}}\}$ the inverse FFT computation can be

performed as:

$$\hat{\mathbf{x}} := \{\hat{x}_k\}_{k=1}^K = \frac{1}{\sqrt{J}} \sum_{j=0}^{J-1} \zeta y_j e^{i2\pi kj/J} \quad (12)$$

where ζ is an indicator function determining if a specific term y_j should be included in the reconstruction. For example, a low-pass filter could be given as:

$$\zeta = \begin{cases} 1, & \text{if } s[j] < \tau \\ 0, & \text{otherwise,} \end{cases} \quad (13)$$

indicating that if the corresponding sample frequency of this y_j is less than a desired task-relevant term τ , it should be included in the reconstruction. As such, the number of dynamics modes \mathcal{N} used for the DMD computation is given by the number of terms y_j where $\zeta = 1$

By incorporating task-relevant frequency information τ the sequence \mathbf{X} is transformed to a new sequence $\hat{\mathbf{X}}$ that is guaranteed to be linear in terms of $\mathcal{D} < K$ frequency components, and as such is a linear solution in \mathcal{D} . In the context of furnace modelling from video, this allows for the extraction of key long-term characteristics which are indicative of incineration state changes, and discarding of unnecessary frequencies such as flame flickering.

V. EXPERIMENTS

To evaluate the proposed TREK framework, this section outlines experiments examining the suitability for both modelling and forecasting. Without loss of generality, experiments use the *Total-Least Squares* implementation of standard DMD [48] and Streaming DMD [49], which minimises measurement error bias and is more suitable for real world systems such as video.

A. SIMULATION

To demonstrate this approach, consider the following illustrative example of decomposing a high-dimensional sequence with a large number of measurements, into its constituent dynamic modes. As outlined in §III, decomposition and forecasting requires only a single measurement sequence, as opposed to traditional machine learning.

In this, a sequence is generated where $\mathcal{N} = 101$ random spatial modes, each of size $\mathcal{P} = 12000$ (equivalent to the dimensionality of an RGB image) are generated from a random \mathcal{P} dimensional orthonormal basis. These modes are evolved according to (4) using \mathcal{N} corresponding linearly spaced frequencies $\delta = [-\Delta t^{-1}, \dots, \Delta t^{-1}]$, where $\Delta t = 1e - 2$, for $t = 150s$. The resultant of this is a data sequence $\mathbf{X} \in \mathbb{C}^{12000 \times 15000}$ comprised of exactly 101 dynamic modes. The sequence energy is shown in FIGURE 4(a) (light grey), where it is seen that this is a relatively complex mixture of trigonometric functions.

The goal is to characterise only long-term trends in the sequence, corresponding to the task-relevant information $\tau = [-1\text{Hz}, 1\text{Hz}]$, i.e., $\mathcal{D} = 2$, shown in FIGURE 4(b) (light grey).

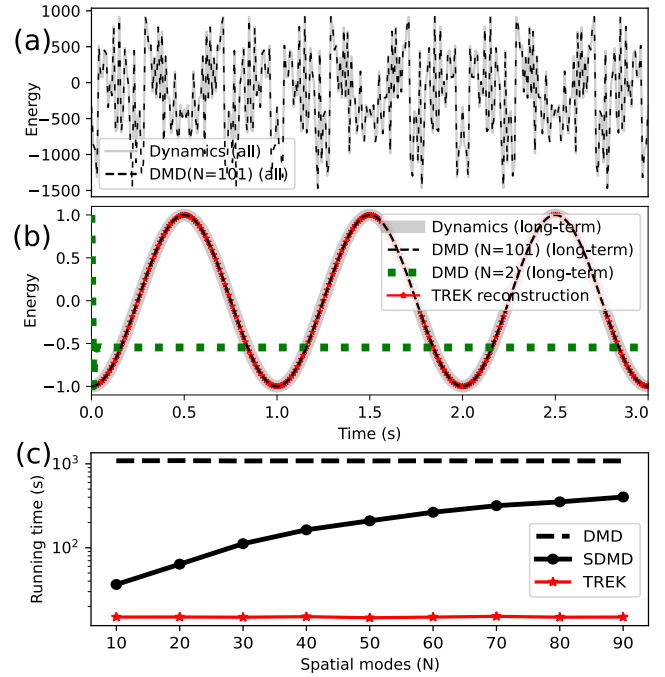


FIGURE 4. (a) Full dynamics (light grey) are complex, but can be learnt by standard DMD. (b) Long-term dynamics (light grey) can be extracted post-hoc from DMD's model of the global dynamics. TREK (red stars) extracts long-term dynamics using frequency knowledge from τ . (c) Running time (log-scale) for each method demonstrates that both DMD and SDMD are poorly suited to decomposing sequences with a large number of spatial modes.

Experiments are performed using an Intel i9-9900K 3.60GHz, with 128GB RAM, conducted in Python 3 with experimental notebook located at ¹. In this paper, the standard DMD approach uses the well-known PyDMD implementation [50] and the streaming DMD approach is based on the *dmtools* codebase [51] supplementary to [13]. Without loss of generality, a forked version of *dmtools* is used in this paper [52], which includes additional Python implementations and minor bug-fixes.

Initially, standard DMD (§III-A) is used to decompose this sequence into $\mathcal{N} = 101$ dynamic modes. Given this decomposition, a reconstruction of the sequence is performed, as shown in FIGURE 4(a). In this, it is seen that the DMD reconstruction utilising all learnt dynamics (both short and long term) perfectly matches the sequence energy, and therefore characterises the dynamics of the sequence. However, as the goal is to extract long-term dynamics from this sequence, a second reconstruction is generated utilising only the *long-term dynamics* learnt from the decomposition, following the post-hoc extraction methodology outlined in §III-C. From this, FIGURE 4(b) demonstrates that DMD has also learnt the long-term components, as the reconstruction matches the long-term dynamical component of the sequence. Additionally, as shown in TABLE 2, standard DMD correctly extracts all 101 dynamic components from the sequence (including the desired task relevant ones in τ), resulting in a

¹https://gitlab.com/Brendan_Michael/TREK

low mean squared error (MSE) between the original sequence and the DMD reconstruction. However, while accurate, this approach is computationally expensive, due to computing the low-rank basis (6) of \mathbf{X} as seen in TABLE 1. In fact, the time taken to compute a DMD model vastly exceeds the sequence runtime (as seen in TABLE 2), and is therefore unsuitable for real-time systems.

To address this slow learning speed, streaming DMD (SDMD) can be used as an iterative solution, and as such this method is evaluated on the same sequence. As shown in TABLE 2, SDMD also correctly identifies the dynamic components, and obtains a similarly small MSE. However, the runtime remains high, due to the high rank requirement of the data ($\mathcal{N} = 101$) dramatically scaling the computation time (as seen previously in TABLE 1).

Given the poor runtime of both standard DMD and the iterative SDMD, a naive approach to address this problem is to choose fewer modes ($\mathcal{N} = 2$) to decrease runtime. However, this results in poor dynamics learning, seen in the failure to characterise any relevant frequencies in FIGURE 4(b), and high MSE in TABLE 2. This is due to a key failure, that being the dynamics must evolve along a low-rank operators (8). Simply choosing $\mathcal{N} = 2$ from a system explicitly containing 101 dynamic modes, provides no guarantees that this is the correct dimensionality for the low-rank operator, as such no correct dynamic modes are extracted.

As an alternative to all these above approaches, the TREK framework is applied with the aim of extracting the task-relevant dynamic components of the sequence, within a reasonable running time. Following the methodology in §IV, task-relevant information τ is used to inform extracting $\mathcal{N} = 2$ task-relevant spectral components from the data with (10) and (12). Streaming DMD is applied on this spectral low-rank data to decompose only the task-relevant (long-term) dynamics. The results for the corresponding reconstruction are shown in FIGURE 4(b), where it is seen that like standard DMD, the TREK reconstruction has correctly characterised the long-term dynamics. Results in TABLE 2 show that similarly the MSE as compared to the long-term dynamics remains small, and it is seen that the computation time is vastly more suitable for real-time, resulting in a 98.8% decrease in runtime. Note, the *non-task relevant* dynamics are not learnt with TREK, and as such there is a high MSE when comparing against all dynamic characteristics in TABLE 2. However that is both unsurprising and unessential for the goals of this approach.

To evaluate the robustness of the approach, this experiment is repeated while varying the number of constituent spatial modes in the range $N = [10, 90]$, thereby exploring application of each method to dynamical systems of varying complexity. These results are shown in FIGURE 4(c), where initially it is seen that the runtime for DMD remains high regardless of model complexity, due to the expensive SVD based preprocessing stage being independent of the number of modes (seen in TABLE 1). In comparison, the

TABLE 2. Simulation results: Number of modes used (\mathcal{N}), running time, MSE between original sequence and learnt reconstruction, MSE between only long-term components of sequence and learnt reconstruction, and number of correctly extracted dynamic modes.

Method	\mathcal{N}	Time(s)	MSE (<i>all</i>)	MSE (τ)	Correct
DMD	101	1217.76	$2.998 \times 10e-16$	$1.017 \times 10e-19$	101/101
SDMD	101	482.85	$1.481 \times 10e-15$	$9.123 \times 10e-20$	101/101
DMD	2	1003.50	$4.358 \times 10e5$	$7.98 \times 10e1$	0/101
SDMD	2	28.46	$4.357 \times 10e5$	$1.490 \times 10e0$	0/101
TREK	2	14.58	$4.358 \times 10e5$	$1.186 \times 10e-20$	2/101

TABLE 3. Furnace video dataset characteristics.

Experiment	\mathcal{P}	Time (min.)	K
Burnout point detection	9000	20	36,000
Long-term forecasting	9000	80	143,999

runtime of SDMD increases with model complexity, due to the larger iterative cost which includes the number of modes (also seen in TABLE 1). As such, both standard methods fail to complete model learning within a reasonable timeframe. In comparison to these, the runtime for TREK remains low regardless of the underlying model complexity, due to its functionality of extracting only task-relevant dynamic modes as part of the pre-processing stage, thereby ensuring the amount of *useful* modes remains small during computation.

B. FURNACE EXPERIMENT

The proposed TREK framework is applied to a real-plant industrial furnace video feed, with the aim of modelling and forecasting dynamic furnace state changes. Input data consists of a three-color RGB sequence of video frames, sampled at 30Hz with original size (1920, 1080, 3), resized to (40, 75, 3) prior to learning (*i.e.*, $\mathcal{P} = 9000$).

1) MODELLING - BURNOUT POINT DETECTION

Initially, the proposed approach is evaluated for modelling a dynamical event, the shifting of a *burnout-point*, occurring over 20 minutes ($K = 36000$ frames). A burnout-point is the primary combustion position in the chamber, and the general control aim is to maintain a steady consistent burn at this fixed position. An example of burnout-point shifting is shown in FIGURE 5, where it is seen that initially the fire burns uniformly on the platform. Over the next 10 minutes, inhomogeneity in waste combustion causes waste at the centre to fail to ignite resulting in waste splitting into two separate upper and lower combustion regions. Following this, at 20 minutes the lower of the two regions extinguishes due to lack of combustible material. This is an unstable state, and the operator will need to actuate the platform to drive reignition. The ability to model this event would be a useful operator tool to help in the control decision making process. Specifically, the operator is interested in characterising long-term dynamic activity greater than five-minutes, while being disinterested

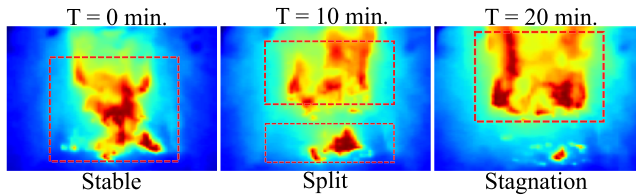


FIGURE 5. Shifting of in-furnace burnout position (pseudo-colour) resulting in stagnation.

in short-term events. As such, the prior task-relevant information involved in this task is the frequency $\tau = [300^{-1}\text{Hz}]$.

To investigate modelling this event, an experiment is performed to extract the long-term dynamics from the sequence. Specifically, the aim is to learn a dynamic model that has: (i) dynamic modes that are interpretable in the context of the plant, (ii) a fast computation time, greater than 30Hz.

Model performance is evaluated by examining three key criteria [11]: (i) the stability of the learnt eigenvalues, indicating the growth and oscillation of the learnt dynamics model, (ii) the coherency of the extracted spatial modes, i.e., the ability to qualitatively explain mechanics of the underlying furnace system, (iii) the mean squared error (MSE) between the reconstructed sequence and the original video, quantifying the prediction accuracy of the modelling.

Initially, the standard DMD approach is applied, by selecting a low-rank basis of spatial modes by (6) using the singular value thresholding approach by (7), resulting in $\mathcal{N} = 9000$, and then computing an eigendecomposition of $\tilde{\mathbf{A}}$. The results for this are seen in FIGURE 6 (a-b), where it is shown that modelling is very unstable and of poor quality. Specifically, learnt eigenvalues in FIGURE 6 (a) lie far from the unit circle (along which stable eigenvalues should reside), and undergo extreme growth, decay, and oscillation, which is obviously not present in the original video sequence. This instability results in uninformative unstable dynamic modes (FIGURE 6 (b)) and the corresponding spatial models do not show any intuitive structures with which to interpret the sequence, instead highlighting hyper-localised regions of combustion. This is expected, as due to the complexity and uncertainty in the video, as discussed in §III-D, fitting linear approximating models to (potentially) non-linear data can result in uninformative models. Given this instability, no useful dynamics are extracted, resulting in an unbounded reconstruction error (seen in TABLE 4).

Additionally, given the data is full-rank ($\mathcal{N} = \mathcal{P}$), SDMD cannot be used to provide real-time updates, due to the reliance on a low-rank basis (§III-B). Simply selecting fewer modes to decrease Streaming DMD runtime is meaningless, resulting again in unstable eigenvalues (FIGURE 6 (a)).

As such, standard DMD approaches fails to learn this complex video data, and reconstruction or modelling of this sequence will be inaccurate and unstable. Additionally, due to the size of \mathcal{P} and K , model training takes approximately 74 minutes to compute, and as such is infeasible for practical application on this 20 minute sequence. It is clear that

TABLE 4. Burnout Point Detection Results: Running time, MSE between original sequence and learnt reconstruction, and number of stable extracted dynamic modes.

Method	Time (min.)	MSE	Stable Eigenvalues
DMD	74.31	<i>NaN</i>	106/9000
TREK	1.12	5.07	8/8

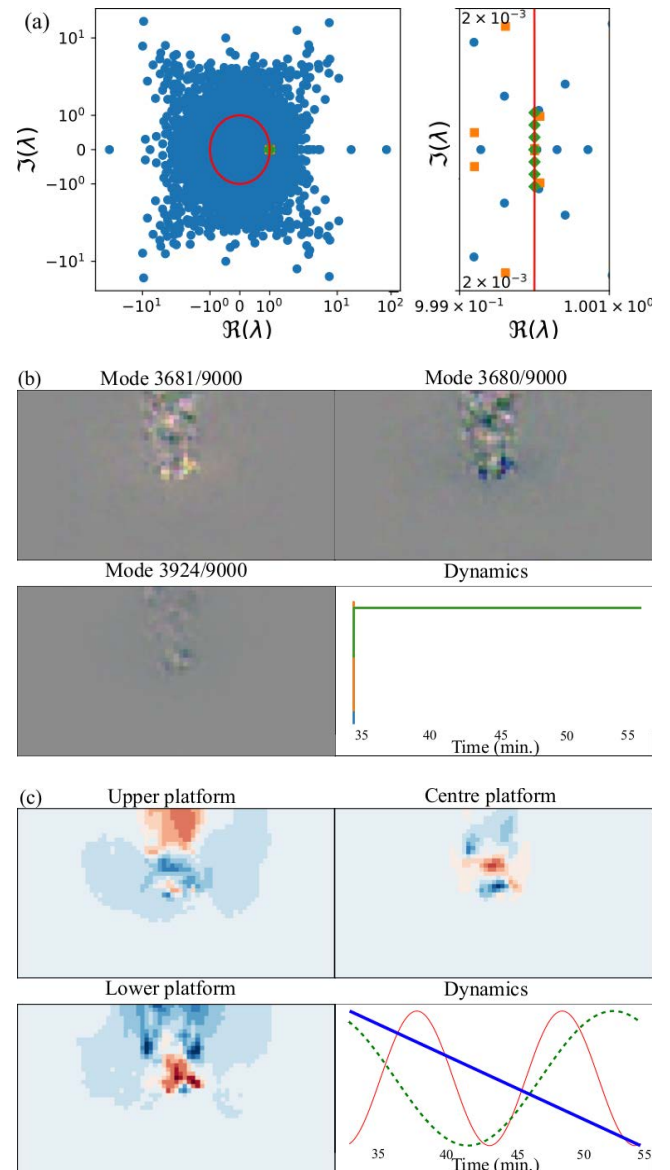


FIGURE 6. Burnout-point modelling: (a) Eigenvalues for DMD (blue circles), Streaming DMD with minimal but incorrect rank (orange squares), and TREK (green diamonds). Unit circle (red) indicates stability, and note symmetric log scale. (b) Standard DMD's dynamic modes (normalised colour) are unstable and captures no intuitive dynamics. (c) In comparison, TREK framework explains characteristics of the furnace, with dynamic modes (normalised colours) corresponding to spatial platform regions: upper (thin red line), centre (dashed green) and lower (thick blue).

standard DMD is not applicable to this scenario, and this approach suffers from limitations as discussed in §III-D.

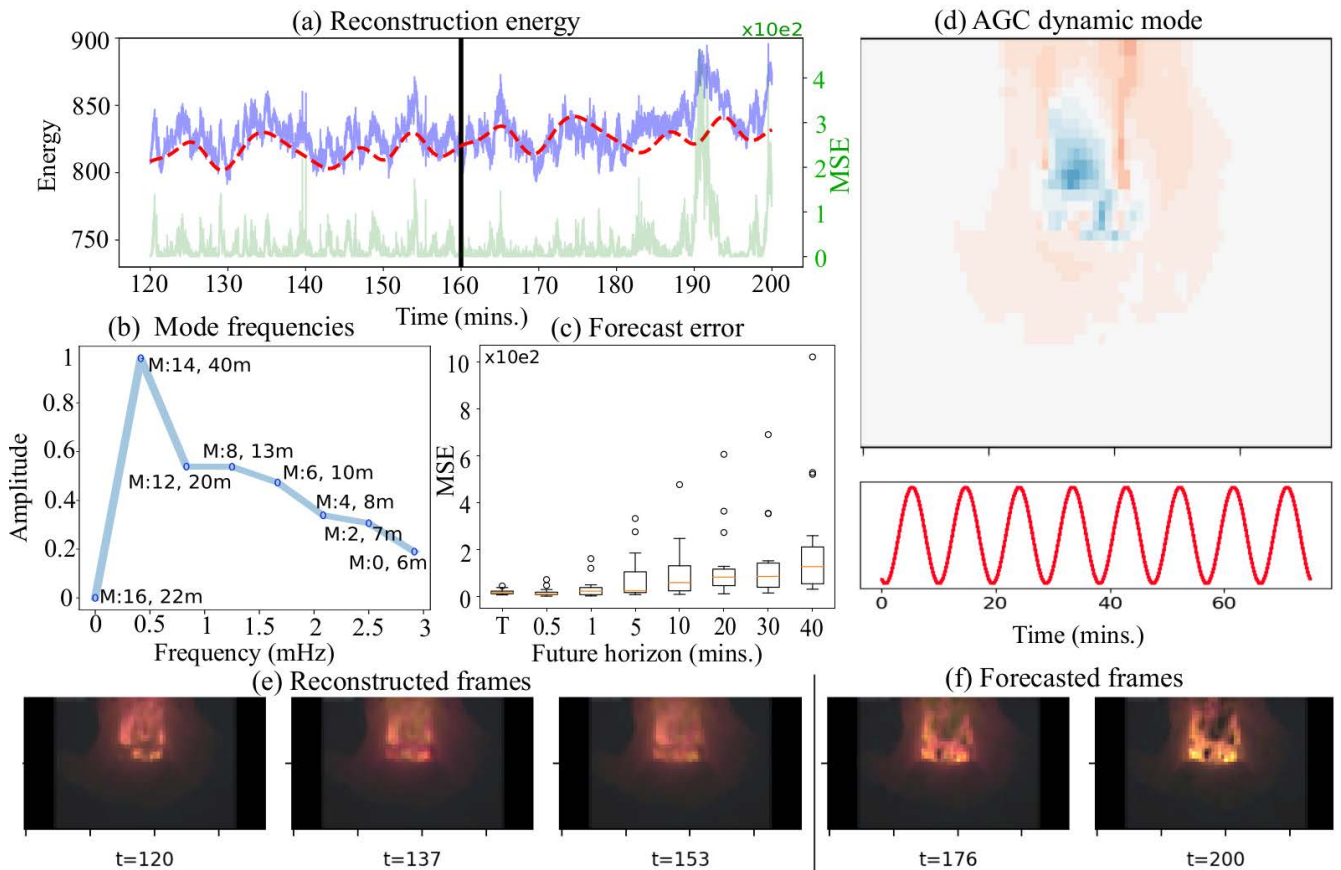


FIGURE 7. (a) Video sequence energy (light blue) displays oscillation behaviour, consistent with the AGC period. Predictions from the TREK framework (red dashed) align closely to sequence energy for both the training period (left of black vertical line), and the (unobserved) forecast period (right of line). Frame-by-frame MSE also shown (light green). (b) Frequency spectrum (millihertz) of dynamic modes show long-term trends are identified. (c) Reconstruction error in training data (T) and forecast error for subsequent 40 minutes. (d) The dynamic mode associated with the 10 minute AGC behaviour (M6), with spatial mode (pseudo-colour) corresponding to a smoke-filled area, and a 10 minute dynamic oscillation period. (e) Reconstructed frames and forecasts show physical characteristics of combustion.

To address these problems, the TREK approach as outlined in §IV is used to first extract spectral components based on task-relevant information τ , and then learn a low-rank Streaming DMD model. Initially, the task-relevant Fourier transform (10) is applied to this data sequence to extract all spectral components. Subsequently the inverse FFT (12) is performed, using the low-pass indicator function (13). As such, extracting only non-zero task-relevant spectral components results in $\mathcal{N} = 8$ Fourier modes with frequencies less than τ .

Eigenvalues from the learnt dynamic model are shown in FIGURE 6 (a), where it is seen that these are stable (as they sit on the unit circle). In addition, the learnt dynamic modes are shown in FIGURE 6 (c) where it is seen that three key spatial regions that characterise this burnout event are identified: an upper, middle, and lower region of the fire. This corresponds to the spatial components that describe the burnout event as seen in FIGURE 5, *i.e.*, a separation of a single region into three separate components. The corresponding learnt dynamics show that these three spatial regions are modelled together to characterise the split and decaying behaviour. As such,

through examination of the outputs, we can characterise the event that occurred and analyse it in terms of spatial-temporal dynamics.

2) LONG-TERM FORECASTING - PATTERN DETECTION

Experiments in §V-B1 demonstrate the proposed method is suitable for real-time analysis of furnace events. However, a desirable property for dynamical modelling is the ability to forecast. To evaluate the suitability of the proposed approach, an experiment is outlined to detect patterns in long-term forecasts. Specifically, in the context of an industrial furnace, the aim is to detect a 10 minute periodic event, the *automated grate control* (AGC), which actuates platform rollers to sift waste, which as a side effect spews smoke and debris.

In this experiment, a video sequence of 80 minutes ($K = 143,999$ frames) is split into two independent sequences, a training sequence covering the first 40 minutes, and a testing sequence covering the later 40 minutes. As in §V-B1, TREK is applied to identify frequency components less than five minutes, resulting in $\mathcal{N} = 20$ dynamic

modes, and decomposition is applied on the low-rank sequence.

The results are shown in FIGURE 7. Initially, it is seen in FIGURE 7 (a) that the energy of the video sequence is stable and oscillates with a fixed period. The energy of the TREK reconstruction also follows this trend, for both the training and forecasting period. As such, the framework has learnt a model that captures the general trend of the measurement sequence and forecasts oscillatory behaviour.

To understand what drives this oscillation behaviour, and determine if the learnt model has learnt a characteristic decomposition, the frequency-power graph of the learnt dynamics is plotted in FIGURE 7(b). In this, it is seen that eigenvalues characterises the sequences in terms of long-term oscillatory components (between 6 and 40 minutes oscillation periods). FIGURE 7(d) shows a visual output of one of these components, the learnt component with period ten minutes (denoted M_6). It is seen that spatially, this mode is characterised by a hazy smoke field region surrounding the main flame body (light red), and combined with the corresponding dynamics showing a 10 minute period, it is clear that this components corresponds to the underlying AGC process. As such, the TREK framework has extracted key spatio-temporal components of a long-term video sequence, and is able to forecast these dynamics for a long horizon. Importantly, the components remain intuitive in the context of the furnace system, and can be used to infer underlying physiochemical characteristics of the plant.

To examine the robustness of the approach, the above experiment is repeated on 18 independent train/test video instances, from a 13 hour video feed. The mean average error of the reconstruction for multiple forecast horizons is shown in FIGURE 7(c). In this, it is seen that TREK is accurate for both reconstructing the training data (while providing intuitive decompositions of the data), and short-term horizon tasks (up to one minute). However the long term forecasting accuracy remains challenging, due to the problems of inhomogeneity of inputs and non-stationary driven dynamics. This is a common limitation of spatio-temporal modelling, and as such, while the proposed approach is suitable for the real-time learning and forecasting, incorporating control and inputs into forecasting remains an open problem.

VI. DISCUSSION

In this paper, a framework is derived and evaluated for the learning of combustion dynamics of an industrial furnace from video-feed data, via a combination of dynamic mode decomposition for extracting intuitive dynamic modes, and Fourier analysis for incorporating task-relevant information. The findings outlined here show that by including this information, complexities and uncertainties in the data can be mitigated during dynamics modelling, resulting in computationally inexpensive, stable predictions. Even when observed combustion is comprised of multiple complex interacting dynamical processes, the use of task-relevant information

eliminates undesired dynamical information, allowing for learning relevant long-term dynamical patterns. In forecasting tasks, the use of task-relevant information can result in longer, more accurate long-term predictions, which can be made quickly and at a reduced computational cost compared to traditional modelling methods. This long-term forecasting coupled with low-cost predictions, demonstrates that this approach for learning dynamics is suitable for real-world scenarios involving uncertain, high-dimensional dynamics.

VII. CONCLUSION AND FUTURE WORK

Generating video forecasts of in-furnace dynamics behaviour is key to supporting plant operator decision making. The approach presented in this paper allows for learning real-time complex dynamics, in a framework that captures intuitive spatio-temporal physical aspects. In the context of real-world dynamics modelling, this method in its current form would find utility in forecasting long-term, steady dynamical behaviour from video. Given the low-computational cost, there is potential for integrating these forecasts with existing real-time control tools, such as state identification or classification. Additionally, the ability of this method to remain agnostic to the complexity of the underlying dynamics has wider implications outside of the field of combustion monitoring from video. This framework may also find utility in other scenarios requiring real-time forecasting with dynamics uncertainty, e.g., real-time adaptation for robot control, specifically, highly non-linear system such as soft robotics or environmental stimuli and interaction. Not only would the ability to encode task-relevant information into dynamics modelling enhance applications, but this approach enables the analysis of non-linear dynamics systems previously thought to be too complex for standard spatio-temporal approaches.

REFERENCES

- [1] L. Makarichi, W. Jutidamrongphan, and K.-A. Techato, "The evolution of waste-to-energy incineration: A review," *Renew. Sustain. Energy Rev.*, vol. 91, pp. 812–821, Aug. 2018.
- [2] F. Cucchiella, I. D'Adamo, and M. Gastaldi, "Sustainable waste management: Waste to energy plant as an alternative to landfill," *Energy Convers. Manage.*, vol. 131, pp. 18–31, Jan. 2017.
- [3] F. Neuwahl, G. Cusano, J. G. Benavides, S. Holbrook, and S. Roudier, "Best available techniques (BAT) reference document for waste incineration," Publications Office Eur. Union, Luxembourg, Tech. Rep. EUR 29971 EN, 2019.
- [4] S.-Y. Pan, M. A. Du, I.-T. Huang, I.-H. Liu, E.-E. Chang, and P.-C. Chiang, "Strategies on implementation of waste-to-energy (WTE) supply chain for circular economy system: A review," *J. Cleaner Prod.*, vol. 108, pp. 409–421, Dec. 2015.
- [5] T. Astrup, C. Riber, and A. J. Pedersen, "Incinerator performance: Effects of changes in waste input and furnace operation on air emissions and residues," *Waste Manage. Res.*, vol. 29, pp. 57–68, Oct. 2011.
- [6] H. Ichihashi, K. Honda, T. Katada, and M. Fujiyoshi, "Post-supervised fuzzy classifier and application to surveillance of combustion in incinerator," in *Proc. Int. Conf. Soft Comput. Intell. Syst.*, 2006, pp. 1643–1648.
- [7] J. Mikleš and M. Fikar, *Process Modelling, Identification, and Control*. Berlin, Germany: Springer, 2007.
- [8] H. Sasaki, T. Hirabayashi, K. Kawabata, Y. Onuki, and T. Matsubara, "Bayesian policy optimization for waste crane with garbage inhomogeneity," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 4533–4540, Jul. 2020.

- [9] M. Rovaglio, D. Manca, and G. Biardi, "Dynamic modeling of waste incineration plants with rotary kilns: Comparisons between experimental and simulation data," *Chem. Eng. Sci.*, vol. 53, pp. 2727–2742, Aug. 1998.
- [10] X. Li, Y. Ma, M. Zhang, M. Zhan, P. Wang, X. Lin, T. Chen, S. Lu, and J. Yan, "Study on the relationship between waste classification, combustion condition and dioxin emission from waste incineration," *Waste Disposal Sustain. Energy*, vol. 1, no. 2, pp. 91–98, Aug. 2019.
- [11] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. Philadelphia, PA, USA: SIAM, 2016.
- [12] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *J. Fluid Mech.*, vol. 656, pp. 5–28, Jan. 2010.
- [13] M. S. Hemati, M. O. Williams, and C. W. Rowley, "Dynamic mode decomposition for large and streaming datasets," *Phys. Fluids*, vol. 26, no. 11, Nov. 2014, Art. no. 111701.
- [14] B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep learning for universal linear embeddings of nonlinear dynamics," *Nature Commun.*, vol. 9, no. 1, pp. 1–10, Dec. 2018.
- [15] N. Takeishi, Y. Kawahara, and T. Yairi, "Learning Koopman invariant subspaces for dynamic mode decomposition," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1130–1140.
- [16] S.-I. Sakai and M. Hiraoka, "Municipal solid waste incinerator residue recycling by thermal processes," *Waste Manage.*, vol. 20, nos. 2–3, pp. 249–258, Apr. 2000.
- [17] M. Waldner, R. Halter, A. Sigg, B. Brosch, H. Gehrman, and M. Keunecke, "Energy from waste—clean, efficient, renewable: Transitions in combustion efficiency and NO_x control," *Waste Manage.*, vol. 33, no. 2, pp. 317–326, 2013.
- [18] S. Martineau, K. J. Burnham, O. C. L. Haas, G. Andrews, and A. Heeley, "Four-term bilinear PID controller applied to an industrial furnace," *Control Eng. Pract.*, vol. 12, no. 4, pp. 457–464, Apr. 2004.
- [19] T. Kanekol, Y. Tsuruminel, J. Poon, Y. Onuki, Y. Dai, K. Kawabata, and T. Matsubara, "Learning deep dynamical models of a waste incineration plant from in-furnace images and process data," in *Proc. IEEE 15th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2019, pp. 873–878.
- [20] G. C. Tiao and G. E. Box, "Modeling multiple time series with applications," *J. Amer. Statist. Assoc.*, vol. 76, no. 376, pp. 802–816, 1981.
- [21] Z. Gao and R. S. Tsay, "Modeling high-dimensional unit-root time series," *Int. J. Forecasting*, vol. 37, no. 4, pp. 1535–1555, Oct. 2021.
- [22] D. Quesada, G. Valverde, P. Larrañaga, and C. Bielza, "Long-term forecasting of multivariate time series in industrial furnaces with dynamic Gaussian Bayesian networks," *Eng. Appl. Artif. Intell.*, vol. 103, Aug. 2021, Art. no. 104301.
- [23] G. Tavares, Z. Zsigraiová, V. Semiao, and M. D. G. Carvalho, "Monitoring, fault detection and operation prediction of MSW incinerators using multivariate statistical methods," *Waste Manage.*, vol. 31, no. 7, pp. 1635–1644, Jul. 2011.
- [24] V. Le Guen and N. Thome, "Disentangling physical dynamics from unknown factors for unsupervised video prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11474–11484.
- [25] Y. Wang, P. Bilinski, F. Bremond, and A. Dantcheva, "ImaGINator: Conditional spatio-temporal GAN for video generation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1160–1169.
- [26] Z. Han, M. M. Hossain, Y. Wang, J. Li, and C. Xu, "Combustion stability monitoring through flame imaging and stacked sparse autoencoder based deep neural network," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114159.
- [27] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019.
- [28] S. R. Hong, J. Hullman, and E. Bertini, "Human factors in model interpretability: Industry practices, challenges, and needs," *Proc. ACM Human-Comput. Interact.*, vol. 4, pp. 1–26, May 2020.
- [29] P. J. Schmid, L. Li, M. P. Juniper, and O. Pust, "Applications of the dynamic mode decomposition," *Theor. Comput. Fluid Dyn.*, vol. 25, no. 1, pp. 249–259, 2011.
- [30] A. Di Ronco, C. Introini, E. Cervi, S. Lorenzi, Y. S. Jeong, S. B. Seo, I. C. Bang, F. Giacobbo, and A. Cammi, "Dynamic mode decomposition for the stability analysis of the molten salt fast reactor core," *Nucl. Eng. Design*, vol. 362, Jun. 2020, Art. no. 110529.
- [31] D. Bruder, C. D. Remy, and R. Vasudevan, "Nonlinear system identification of soft robot dynamics using Koopman operator theory," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 6244–6250.
- [32] J. L. Proctor and P. A. Eckhoff, "Discovering dynamic patterns from infectious disease data using dynamic mode decomposition," *Int. Health*, vol. 7, no. 2, pp. 139–145, Mar. 2015.
- [33] J. Mann and J. N. Kutz, "Dynamic mode decomposition for financial trading strategies," *Quant. Finance*, vol. 16, no. 11, pp. 1643–1655, Nov. 2016.
- [34] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, "On dynamic mode decomposition: Theory and applications," *J. Comput. Dyn.*, vol. 1, no. 2, pp. 391–421, 2014.
- [35] A. Mauroy, I. Mezić, and Y. Susuki, *Koopman Operator in Systems and Control*. Cham, Switzerland: Springer, 2020.
- [36] S. L. Clainche and J. M. Vega, "Higher order dynamic mode decomposition," *SIAM J. Appl. Dyn. Syst.*, vol. 16, no. 2, pp. 882–925, 2017.
- [37] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," in *A Practical Approach to Microarray Data Analysis*. Boston, MA, USA: Springer, 2003, pp. 91–109.
- [38] J. Page and R. R. Kerswell, "Searching turbulence for periodic orbits with dynamic mode decomposition," *J. Fluid Mech.*, vol. 886, Mar. 2020.
- [39] N. Takeishi, Y. Kawahara, Y. Tabei, and T. Yairi, "Bayesian dynamic mode decomposition," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2814–2821.
- [40] T. Kawashima, H. Shouno, and H. Hino, "Bayesian dynamic mode decomposition with variational matrix factorization," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 9, pp. 8083–8091.
- [41] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, "Spectral analysis of nonlinear flows," *J. Fluid Mech.*, vol. 641, pp. 115–127, Dec. 2009.
- [42] I. Mezić, "Analysis of fluid flows via spectral properties of the Koopman operator," *Annu. Rev. Fluid Mech.*, vol. 45, no. 1, pp. 357–378, Jan. 2013.
- [43] C. Folkestad, D. Pastor, I. Mezić, R. Mohr, M. Fonoberova, and J. Burdick, "Extended dynamic mode decomposition with learned Koopman eigenfunctions for prediction and control," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2020, pp. 3906–3913.
- [44] M. R. Jovanović, P. J. Schmid, and J. W. Nichols, "Sparsity-promoting dynamic mode decomposition," *Phys. Fluids*, vol. 26, no. 2, Feb. 2014, Art. no. 024103.
- [45] R. N. Bracewell and R. N. Bracewell, *The Fourier Transform and Its Applications*. New York, NY, USA: McGraw-Hill, 1986.
- [46] K. K. Chen, J. H. Tu, and C. W. Rowley, "Variants of dynamic mode decomposition: Boundary condition, Koopman, and Fourier analyses," *J. Nonlinear Sci.*, vol. 22, no. 6, pp. 887–915, Mar. 2012.
- [47] H. Lange, S. L. Brunton, and J. N. Kutz, "From Fourier to Koopman: Spectral methods for long-term time series prediction," *J. Mach. Learn. Res.*, vol. 22, no. 41, pp. 1–38, 2021.
- [48] S. T. M. Dawson, M. S. Hemati, M. O. Williams, and C. W. Rowley, "Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition," *Exp. Fluids*, vol. 57, no. 3, p. 42, Mar. 2016.
- [49] M. S. Hemati, C. W. Rowley, E. A. Deem, and L. N. Cattafesta, "De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets," *Theor. Comput. Fluid Dyn.*, vol. 31, no. 4, pp. 349–368, 2017.
- [50] N. Demo, M. Tezzele, and G. Rozza, "PyDMD: Python dynamic mode decomposition," *J. Open Source Softw.*, vol. 3, no. 22, p. 530, Feb. 2018.
- [51] M. H. Clarence Rowley and M. O. Williams. (2015). *dmdtools*. [Online]. Available: <https://github.com/cwrowley/dmdtools>
- [52] B. Michael. (2021). *dmdtools Fork*. [Online]. Available: https://gitlab.com/Brendan_Michael/dmdtools



BRENDAN MICHAEL received the Ph.D. degree in computer science from King's College London (KCL), U.K., in 2017. From 2017 to 2020, he was a Teaching Fellow at KCL. He is currently an Assistant Professor with the Robot Learning Laboratory, Nara Institute of Science and Technology, Japan. His research interests include machine learning applications to systems with uncertainty and noise.



AKIFUMI ISE received the B.S. degree in computer science and intelligent systems and the M.S. degree in computer science and intelligent systems from Osaka Prefecture University, Japan, in 2010 and 2012, respectively. He previously worked as a Game Programmer and a Visiting Researcher at Osaka Prefecture University, and currently performs development work at the Environment Business Headquarters and Development Center, Hitachi Zosen Corporation, Japan.



KAORU KAWABATA joined Hitachi Zosen Corporation, Japan, in 1990 and worked in electrical instrumentation design. He worked on many projects in environmental business: long-term management of environmental facilities in 2003, setting up and management of a Remote Operations Center (ROC) for waste incineration plants in 2011, development of 24-hour remote monitoring system from the ROC in 2012, development of the image recognition system CoSMoS in 2013, devel-

opment of a large-capacity distributed system for collection and monitoring plant data with JASRI in 2015 and development of an automatic combustion advanced system for waste incineration called Comp. ACC in 2018. Since 2020, he has been an Engineering Leader of the Environmental Business Headquarters and Development Center at Hitachi Zosen Corporation, where he manages developments of new products and technologies.



TAKAMITSU MATSUBARA (Member, IEEE) received the Ph.D. degree in information science from the Nara Institute of Science and Technology, Nara, Japan, in 2007. From 2005 to 2007, he was a Research Fellow (DC1) of the Japan Society for the Promotion of Science. From 2013 to 2014, he was a Visiting Researcher with the Donders Institute for Brain Cognition and Behavior, Radboud University Nijmegen, Nijmegen, The Netherlands. He is currently an Associate Professor with the Nara Institute of Science and Technology and a Visiting Researcher with ATR, Kyoto, Japan, and AIST AI Center, Tokyo, Japan. His research interests include machine learning and control theory for robotics.

... ..