

Markov Modeling of Time Series via Spectral Analysis for Detection of Combustion Instabilities ^{*}

Devesh K. Jha, Nurali Virani, and Asok Ray

Abstract Modeling of temporal patterns to infer generative models from measurement data is critical for dynamic data-driven application systems (DDDAS). Markov models are often used to capture temporal patterns in sequential data for statistical learning applications. This chapter presents a methodology for reduced-order Markov modeling of time-series data based on spectral properties of stochastic matrix and clustering of directed graphs. Instead of the common Hidden Markov model (HMM)-inspired techniques, a symbolic dynamics-based approach to infer an approximate generative Markov model for the data. The time-series data is first symbolized by partitioning of the discrete-valued signal in continuous domain. The size of temporal memory of the discretized symbol sequence is then estimated using spectral properties of the stochastic matrix created from the symbol sequence for a first-order Markov model of the symbol sequence. Then, a graphical method is used to cluster the states of the corresponding high-order Markov model to infer a reduced-size Markov model with a non-deterministic algebraic structure. A Bayesian inference rule captures the parameters of the reduced-size Markov model from the original model. The proposed idea is illustrated by creating Markov models for pressure time-series data from a swirl stabilized combustor where some controlled protocols are used to induce instability. Results demonstrate complexity modeling of the underlying Markov model as the system operating condition changes from stable to unstable which is useful in combustion applications such as detection and control of thermo-acoustic instabilities.

Department of Mechanical and Nuclear Engineering, The Pennsylvania State University, University Park, PA, 16802, USA
e-mail: (dkj5042, nnv105, axr2)@psu.edu

^{*} This work has been supported in part by the U.S. Air Force Office of Scientific Research under Grant No. FA9550-15-1-0400.

1 Motivation and Introduction

The underlying theory of symbolic time-series analysis (STSA) [1, 2] has led to the development of signal representation tools in the paradigm of dynamic data-driven application systems (DDDAS) [3, 4], where time series of sensor signals are partitioned to obtain respective symbol strings. In general, STSA is a nonlinear technique for representation of temporal patterns in sequential data, where the underlying continuous attributes are projected onto a symbolic space. This step is followed by identification of concise probabilistic patterns for compression of the discretized information. Within this framework, *finite-memory* Markov models have been shown to be a reasonable *finite-memory* approximation (or representation) of systems with fading memory (e.g., engineering systems that exhibit stable orbits or mixing) [5, 6].

Once the continuous data are discretized, the memory estimate for the discretized sequence is used for compression as a finite-memory Markov process, which is represented by a state transition matrix. The transition matrix is estimated by a maximum likelihood estimator (MLE) under the assumption of infinite data and uniform priors for all elements of the transition matrix. In contrast to the probabilistic finite state automaton (PFSA)-based approach to infer a Markov model for time-series data presented in [5–7], an alternative method has been proposed in this chapter, where the constraints of the deterministic algebraic structure of finite-state automata are relaxed to allow non-deterministic transitions for the PFSA inferred from the time-series data. This task has been performed by making a trade-off for lower complexity of the generated model (possibly) at the expense of resolution loss. The proposed concept is validated for model inferencing using time-series data from a swirl-stabilized combustor and identify the different stages of the complex instability phenomenon from a completely data-driven perspective. We also point-out to the changes in the model structure and their physical interpretations based on the data from the process.

Hidden Markov Modeling (HMM) is the most-widely used statistical learning tool for modeling time-series data [8] where the data is modeled as a Markov process with unobserved states. The learning task is to infer the states and the corresponding parameters of the Markov chain. In contrast to HMM, some other non-linear techniques have also been proposed for Markov modeling of time-series data where the states of the Markov chain are some collection of words of different lengths which can be obtained from the time-series data up on projecting the data to a discrete space with finite cardinality [5–7, 9, 10]. The common concept in all these techniques, based on Markov modeling of discrete sequences, is that the Markov chain is induced by probabilistic version of a deterministic FSA [5]. While the PFSA-based inference provides a consistent, deterministic graph structure for learning, the deterministic algebraic structure is generally redundant and can often lead to large number of states in the induced Markov model. Merging the states of the PFSA for dimensionality reduction is often inconsistent due to the algebraic constraints [6]. Some other approaches for state aggregation in Markov chains could be found in [11–14]. However, these papers present aggregation of

states in a Markov chains; construction of the Markov chain from data is never considered. It is important that these two problems (i.e., state merging and Markov chain construction) be studied together for analysis of dynamic data-driven application systems (DDAS) [3, 4]. Moreover, the optimal model selection is inspired by *wrapper*-based techniques where the system searches for the best one in all the above techniques as the similarity is measured between the Markov chains of different dimensions obtained by merging certain parts of the state-space.

This chapter presents a Markov modeling technique for time-series data where the size of temporal memory of the symbolic data is estimated by using the spectral properties of a PFSA whose states are words of length one [15, 16]. Next the states are merged and the deterministic algebraic properties associated with PFSA are removed, where the states of the Markov chain is now some collection of words from its alphabet of length estimated in the last step. The parameters of the reduced-order Markov model are estimated using a Bayesian inference technique from the parameters associated with the higher-order Markov model. The final model obtained is a generative model for the data; however, some information is lost as parts of the deterministic structure of a finite state automaton (FSA) are removed. This approach is used to construct Markov models for pressure data obtained from a combustion instability [17–19] phenomenon, which is a highly non-linear thermo-acoustic process and very hard to model completely using first principles of physics. This chapter addresses data-driven modeling for real-time detection of changes in the underlying process. Specifically these data-driven models can assist prediction and control of combustion instabilities and thus, allow for more reliable and efficient operation of modern-day combustors such as those used in aircraft gas turbine engines.

2 Background and Mathematical Preliminaries

Symbolic analysis of time-series data is a recent approach where continuous sensor data are converted to symbol sequences via partitioning of the continuous domain [5, 20]. The dynamics of the symbols sequences are then modeled as a Probabilistic Finite State Automata (PFSA), which is defined as follows:

Definition 1 (PFSA). A Probabilistic Finite State Automata (PFSA) is a tuple $G = (\mathcal{Q}, \mathcal{A}, \delta, \mathbf{M})$ where

- \mathcal{Q} is a finite set of states of the automata;
- \mathcal{A} is a finite alphabet set of symbols $a \in \mathcal{A}$;
- $\delta : \mathcal{Q} \times \mathcal{A} \rightarrow \mathcal{Q}$ is the state transition function;
- $\mathbf{M} : \mathcal{Q} \times \mathcal{A} \rightarrow [0, 1]$ is the $|\mathcal{Q}| \times |\mathcal{A}|$ emission matrix. The matrix $\mathbf{M} = [m_{ij}]$ is row stochastic such that m_{ij} is the probability of generating symbol a_j from state q_i .

For symbolic analysis of time-series data, a class of PFSAs called the D -Markov machine have been proposed [5] as a sub-optimal but computationally efficient approach to encode the dynamics of symbol sequences as a finite state machine. For

most stable and controlled engineering systems that tend to forget their initial conditions, a finite length memory assumption is reasonable. The states of this PFSA are words over \mathcal{A} of length D (or less); and state transitions are described by a sliding block code of memory D and anticipation length of one [21]. The dynamics of this PFSA can both be described by the $|\mathcal{Q}| \times |\mathcal{Q}|$ state transition matrix $\mathbf{\Pi}$ or the $|\mathcal{Q}| \times 1$ state visit probability vector \mathbf{p} . The alphabet size or the level of coarse-graining of the continuous domain is driven by the resolution level required to capture the dynamics of the system – domain knowledge or data-driven partitioning techniques [22] can be used for this purpose. Estimating the depth of historical influences, on the other hand, requires estimation of the decay-rate of the memory of a dynamical system.

For systems with fading memory it is expected that the predictive influence of a symbol progressively diminishes further into the future. Formally depth is defined as follows:

Definition 2 (Depth). Let $\mathbf{s} = s_1 \dots s_k s_{k+1} s_{k+2} \dots$ be the observed symbol sequence where each $s_j \in \mathcal{A} \forall j \in \mathbb{N}$. Then, the depth of the process generating \mathbf{s} is defined as the length D such that:

$$\Pr(s_k | s_{k-1}, \dots, s_1) = \Pr(s_k | s_{k-1}, \dots, s_{k-D}) \quad (1)$$

An accurate estimation of depth for the symbolic dynamical process is required for the precise modeling of the underlying dynamics of the discrete sequence. Next an information-theoretic metric is introduced, which is used for merging the states of the Markov model later in next section.

Definition 3 (Kullback-Leibler Divergence). The Kullback-Leibler (K-L) divergence of a discrete probability distribution P from another distribution \tilde{P} is defined as follows.

$$D_{\text{KL}}(P \parallel \tilde{P}) = \sum_{x \in X} p_X(x) \log \left(\frac{p_X(x)}{\tilde{p}_X(x)} \right)$$

It is noted that K-L divergence is not symmetric; however, it can be converted to a symmetric distance as follows: $d(P, \tilde{P}) = D_{\text{KL}}(P \parallel \tilde{P}) + D_{\text{KL}}(\tilde{P} \parallel P)$. This is defined as the K-L distance between the distributions P and \tilde{P} .

This distance is used to determine the structure in the set of the states of the PFSA-based Markov model whose states are words, over the alphabet of the PFSA, of length equal to the depth estimated for the discretized sequence.

3 Proposed Approach

This section presents the details of the proposed approach for inferring a Markov model from the time-series data. As discussed earlier, the first step is the discretization of the time-series data to generate a discrete symbol sequence. It is possible to optimize the symbolization of time-series using an optimization criterion and the

details are available in literature (e.g., see [17]). The data are discretized using the unbiased principle of entropy maximization of the discrete sequence using Maximum Entropy Partitioning (MEP) [23]. The proposed approach consists of three critical steps and is also shown in Figure 1 for pedagogical purposes.

- Estimate the approximate size of temporal memory (or order) of the symbol sequence.
- Cluster the states of the high-order Markov model.
- Estimate the parameters of the reduced-order Markov model (i.e., the Transition matrix).

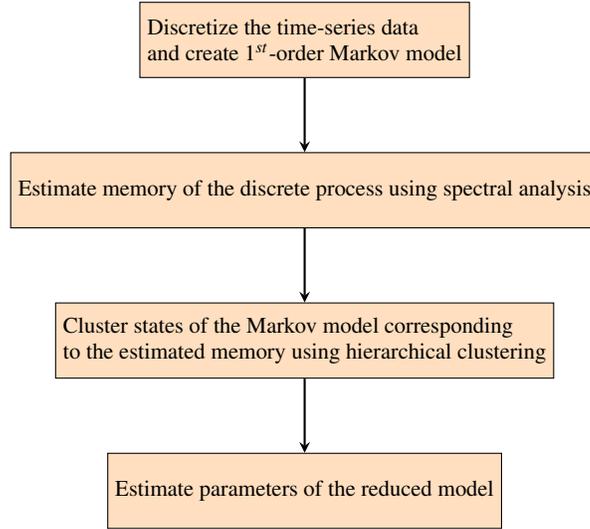


Fig. 1: Flowchart for the proposed reduced-order modeling

Memory of the discrete sequence is estimated using a recently introduced method based on the spectral analysis of the 1st order Markov model induced by a PFSA [15, 16]. The key ideas behind the three steps are explained next.

3.1 Estimation of Reduced-Order Markov model

Depth D of a symbol sequence has been redefined in [15] as the number of time steps after which probability of current symbol is independent of any past symbol i.e.:

$$\Pr(s_k | s_{k-n}) = \Pr(s_k) \quad \forall n > D \quad (2)$$

Note that dependence in the proposed definition (eq. 2) is evaluated on individual past symbols using $\Pr(s_k | s_{k-n})$ as opposed to the assessing dependence on words of length D using $\Pr(s_k | s_{k-1}, \dots, s_{k-D})$. It is shown that if the observed process is

forward causal then observing any additional intermediate symbols $s_{k-1}, \dots, s_{k-n+1}$ cannot induce a dependence between s_k and s_{k-n} if it did not exist on individual level.

Let $\mathbf{\Pi} = [\pi_{ij}^{(1)}]$ be the one-step transition probability matrix of the PFSA G constructed from this symbol sequence i.e.

$$\mathbf{\Pi} = \Pr(s_k | s_{k-1}) \quad (3)$$

Then using the distance of the transition matrix after steps from the stationary point, depth can be defined as a length D such that

$$|\text{trace}(\mathbf{\Pi}^n) - \text{trace}(\mathbf{\Pi}^\infty)| \leq \sum_{j=2}^J |\lambda_j|^n < \varepsilon \quad \forall n > D \quad (4)$$

J is number of non-zero eigenvalues of $\mathbf{\Pi}$. Thus, the depth D of the symbol sequence is estimated for a choice of ε by estimating the stochastic matrix for the one-step PFSA. Next, another pass of data is done through the module to estimate the PFSA parameters whose states are words over \mathcal{A} of length D , i.e., $\mathbf{\Pi} = \Pr(s_k | s_{k-1}, \dots, s_{k-D})$.

The states of the reduced-order Markov model are then estimated by partitioning the set of words over \mathcal{A} of length D estimated in the last step. This is done by using an agglomerative hierarchical clustering approach. The advantage of using the hierarchical clustering approach is that it helps visualize the structure of the set of the original states using an appropriate metric. Agglomerative hierarchical clustering is a bottom-up clustering approach [24] that generates a sparse network (e.g., a binary tree) of the state set \mathcal{Q} (where $|\mathcal{Q}| = |\mathcal{A}|^D$) by successive addition of edges between the elements of \mathcal{Q} . Initially, each of the states q_1, q_2, \dots, q_n is in its own cluster C_1, C_2, \dots, C_n where $C_i \in \mathcal{C}$, which is the set of all clusters for the hierarchical cluster tree. The distance between any two states, q_i and q_j , in \mathcal{Q} is measured by using the K-L distance between the symbol emission probabilities conditioned on these states, i.e.,

$$d(q_i, q_j) = D_{\text{KL}}(\Pr(\mathcal{A} | q_i) \| \Pr(\mathcal{A} | q_j)) + D_{\text{KL}}(\Pr(\mathcal{A} | q_j) \| \Pr(\mathcal{A} | q_i)) \quad (5)$$

In terms of the distance measured by Eq. (5), the pair of clusters that are nearest to each other are merged and this step is repeated till only one cluster is left. The tree structure displays the order of splits in the state set of the higher-order Markov model and is used to aggregate the states close to each other. The set of states clustered together could be obtained based on the number of final states required in the final Markov model.

Remark 1 (Stopping Criterion for Merging). The stopping criterion for the algorithm could be reached based on the modeling objective. In the absence of any defined end objective, the criterion for stopping the algorithm could be found using approaches like Minimum description length (MDL) for signal representation. If stopping criterion corresponds to another end objective (e.g., class separability),

then a Bayesian inference rule could be used to arrive at a consistent cardinality of clusters in \mathcal{Q} . However, in this chapter, the algorithm is terminated by fixing the desired number of states a priori. For a detailed discussion, interested readers are referred to [17].

Remark 2. The final Markov model is a finite depth approximation of the original time-series data. However, compared to the PFSA-based D-Markov machines in [5, 6], the current aggregated model has a non-deterministic algebraic structure, i.e., the same symbol emissions from a state can lead to different states. While this leads to some information loss as compared to the models in [5, 6], this facilitates compression of the size of the model as per the application requirements. For example, even though the optimal model might require a higher finite memory adding all words corresponding to that length might not be necessary to preserve the statistical behavior or class separability. Furthermore, the aggregated model would allow faster convergence rates for the symbol emission probabilities which can be calculated using Glivenko-Cantelli theorem [25]. For a detailed discussion interested readers are referred to [17].

3.2 Estimation of Parameters for the Reduced-Order Markov model

The parameters of the Markov model obtained after clustering the states of the original PFSA with $|\mathcal{A}|^D$ states is obtained using a Bayesian inference technique using the parameters estimated for the PFSA. In this proposed approach, the state transition matrix $\mathbf{\Pi}$, the emission matrix \mathbf{M} , and the state probability vector \mathbf{p} of the original PFSA model G are available, along with the deterministic assignment map $f: \mathcal{Q} \rightarrow \tilde{\mathcal{Q}}$ of the state in \mathcal{Q} (i.e., state set of original model) to one of the state in $\tilde{\mathcal{Q}}$ (i.e., state set of the reduced order model). Since the reduced order model can be represented by the tuple $\tilde{G} = (\tilde{\mathcal{Q}}, \tilde{\mathbf{\Pi}})$, where $\tilde{\mathbf{\Pi}} = [\tilde{\pi}_{ij}]$ is the state transition matrix, a Bayesian inference technique is employed to infer the individual values of transition probabilities $\tilde{\pi}_{ij} = \Pr(\tilde{q}_{k+1} = j \mid \tilde{q}_k = i)$ for all $i, j \in \tilde{\mathcal{Q}}$.

Let Q_k be the random variable denoting the state of PFSA model at some time step $k \in \mathbb{N}$ and S_k denotes the symbol emitted from that state, this probabilistic emission process is governed by the emission matrix \mathbf{M} . The state of the reduced order model is obtained from a deterministic mapping of the state of the PFSA model, thus the state of this model is also a random variable, which is denoted by $\tilde{Q}_k = f(Q_k)$. The Bayesian network representing the dependencies between these variables is shown in the recursive as well as unrolled form in the Figure 2. The conditional density $\Pr(\tilde{Q}_k = \tilde{q} \mid Q_k = q)$ can be evaluated by checking if state q belongs to the state cluster \tilde{q} and assigning the value of 1 if true, else assign it the value of 0. Since it is known that $\tilde{\mathcal{Q}}$ partitions the set \mathcal{Q} , the conditional density is well-defined. Thus, it can be written as

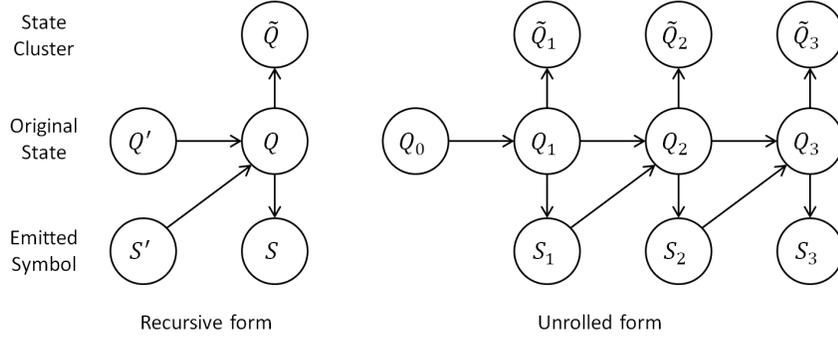


Fig. 2: Graphical models representing dependencies between the random variables

$$\Pr(\tilde{Q}_k = \tilde{q} \mid Q_k = q) = I_{\tilde{q}}(q), \quad (6)$$

where I is the indicator function with $I_{\tilde{q}}(q) = 1$, if element q belongs to the set \tilde{q} , else it is 0. The derivation of the Markov model $\Pr(\tilde{Q}_{k+1} \mid \tilde{Q}_k)$ using $\Pr(Q_{k+1} \mid Q_k)$, stationary probability vector \mathbf{p} , and assignment map f is shown ahead.

$$\Pr(\tilde{Q}_{k+1} \mid \tilde{Q}_k) = \sum_{q \in \mathcal{Q}} \Pr(\tilde{Q}_{k+1}, Q_{k+1} = q \mid \tilde{Q}_k) \quad (7)$$

(Marginalization)

$$= \sum_{q \in \mathcal{Q}} \Pr(Q_{k+1} = q \mid \tilde{Q}_k) \Pr(\tilde{Q}_{k+1} \mid Q_{k+1} = q) \quad (8)$$

(Factorization using Figure 2)

$$= \sum_{q \in \mathcal{Q}} \Pr(Q_{k+1} = q \mid \tilde{Q}_k) I_{\tilde{Q}_{k+1}}(q) \quad (9)$$

(using (6))

$$= \sum_{q \in \mathcal{Q}_{k+1}} \Pr(Q_{k+1} = q \mid \tilde{Q}_k). \quad (10)$$

where $\Pr(Q_{k+1} \mid \tilde{Q}_k)$ is obtained from Bayes' rule as

$$\Pr(Q_{k+1} \mid \tilde{Q}_k) = \frac{\Pr(\tilde{Q}_k \mid Q_{k+1}) \Pr(Q_{k+1})}{\sum_{q \in \mathcal{Q}} \Pr(\tilde{Q}_k \mid Q_{k+1} = q) \Pr(Q_{k+1} = q)}. \quad (11)$$

By following the steps to obtain (10),

$$\Pr(\tilde{Q}_k \mid Q_{k+1}) = \sum_{q \in \tilde{Q}_k} \Pr(Q_k = q \mid Q_{k+1}). \quad (12)$$

where $\Pr(Q_k \mid Q_{k+1})$ results from Bayes' rule as

$$\Pr(Q_k | Q_{k+1}) = \frac{\Pr(Q_{k+1} | Q_k) \Pr(Q_k)}{\sum_{q \in \mathcal{Q}} \Pr(Q_{k+1} | Q_k = q) \Pr(Q_k = q)}. \quad (13)$$

It is noted that, for the distribution $\Pr(Q_k)$ and $\Pr(Q_{k+1})$, a stationary probability \mathbf{p} is available. Using equations (10), (11), (12), and (13) together, one can easily obtain the desired state transition matrix $\tilde{\mathbf{\Pi}}$ of the reduced order model. Once the state cluster set $\tilde{\mathcal{Q}}$ and state transition matrix $\tilde{\mathbf{\Pi}}$ are available, the reduced order model is completely defined. The rest of the chapter will demonstrate the utility of these models in a practical problem of modeling combustion instabilities from time-series data.

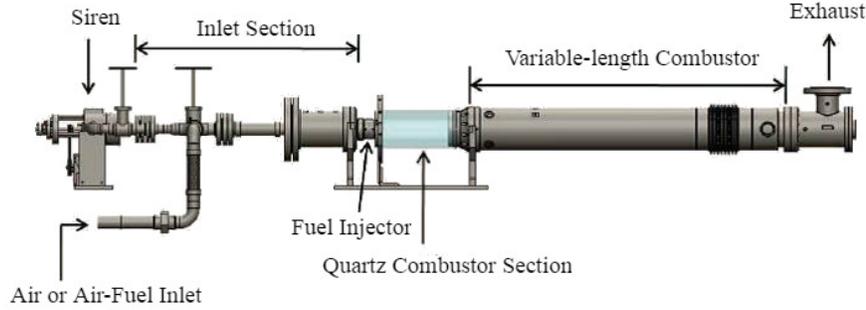


Fig. 3: Schematic drawing of the test apparatus

4 Combustion Experiment Details

This section presents the experimental details for collecting data to analyze the complex non-linear phenomena that occurs during the instability phenomena, in a laboratory-scale combustor. A swirl-stabilized, lean-premixed, laboratory-scale combustor was used to perform the experimental study. Figure 3 shows a schematic drawing of the variable-length combustor. The combustor consists of an inlet section, an injector, a combustion chamber, and an exhaust section. The combustion chamber consists of an optically-accessible quartz section followed by a variable length steel section.

Table 1: Operating conditions

Parameters	Value
Equivalence Ratio	0.525, 0.55, 0.60, 0.65
Inlet Velocity	25-50 m/s in 5 m/s increments
Combustor Length	25-59 inch in 1 inch increments

High pressure air is delivered to the experiment from a compressor system after passing through filters to remove any liquid or particles that might be present. The air supply pressure is set to 180 psig using a dome pressure regulator. The air is pre-heated to a maximum temperature of 250 °C by an 88kW electric heater. The fuel for

this study is natural gas (approximately 95% methane). It is supplied to the system at a pressure of 200 psig. The flow rates of the air and natural gas are measured by thermal mass flow meters. The desired equivalence ratio and mean inlet velocity is set by adjusting these flow rates with needle valves. For fully pre-mixed experiments (FPM), the fuel is injected far upstream of a choke plate to prevent equivalence ratio fluctuations. For technically pre-mixed experiments (TPM), fuel is injected in the injector section near the swirler. It mixes with air over a short distance between the swirler and the injector exit. Tests were conducted at a nominal combustor pressure of 1 atm over a range of operating conditions, as listed in Table 1. Other details, which are reported in [4], are skipped for brevity.

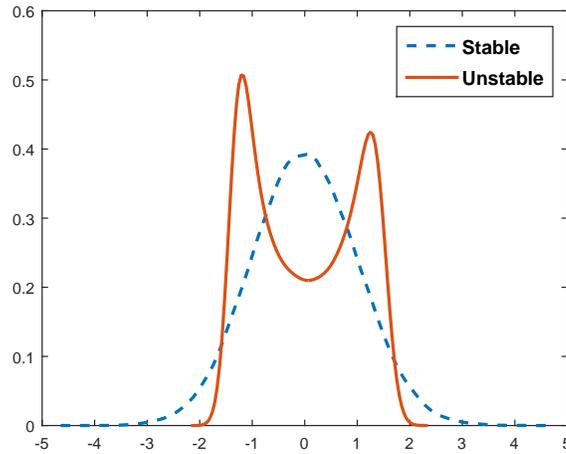
5 Results and Discussion

This section presents details of the analyses completed by using the pressure time-series data to infer the underlying reduced-order Markov model. Time-series data is first normalized by subtracting the mean and dividing by the standard deviation of its elements; this step corresponds to bias removal and variance normalization. Data from engineering systems is typically oversampled to ensure that the underlying dynamics can be captured. Due to coarse-graining from the symbolization process, an over-sampled time-series may mask the true nature of the system dynamics in the symbolic domain (e.g., occurrence of self loops and irrelevant spurious transitions in the Markov chain). Time-series is first down-sampled to find the next crucial observation. The first minimum of auto-correlation function generated from the observed time-series is obtained to find the uncorrelated samples in time. The data sets are then down-sampled by this lag. To avoid discarding significant amount of data due to downsampling, down-sampled data using different initial conditions is concatenated. Further details of this preprocessing can be found in [15].

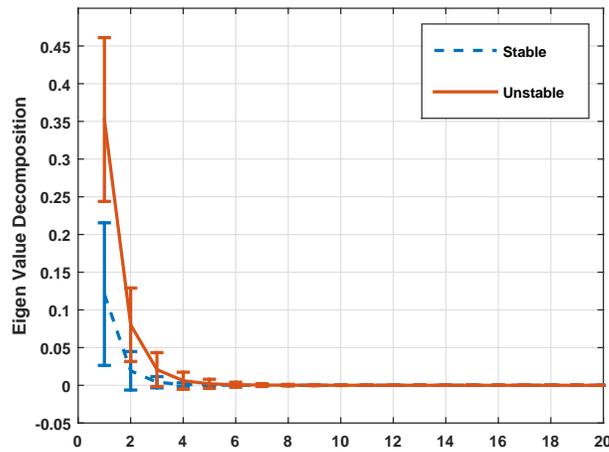
The continuous time-series data set is then partitioned using maximum entropy partitioning (MEP), where the information rich regions of the data set are partitioned finer and those with sparse information are partitioned coarser. In essence, each cell in the partitioned data set contains approximately an equal number of data points under MEP. A ternary alphabet with $\mathcal{A} = \{0, 1, 2\}$ has been used to symbolize the continuous combustion instability data. As discussed in Section 4, sets of time-series data from different phases have been analyzed, as the combustion process emerges from stable through the transient to the unstable region.

Figure 4 demonstrates the observed changes in the behavior of the data as the combustion operating condition changes from stable to unstable. As seen, there is a change in the empirical distribution which changes from a unimodal-Gaussian to a multi-modal Gaussian (bi-modal in Figure 4) as the operating condition changes from stable to unstable. Selected 150 samples of pressure data from the stable and unstable phases each are analyzed and compared.

First, the expected size of temporal memory is compared during the two phases. There are changes in the Eigen value decomposition rate for the 1-step stochastic



(a) Probability density function for the pressure time-series data



(b) Spectral decomposition of the stochastic matrix for 1-step Markov model

Fig. 4: (a) shows the change in the empirical density calculated for the pressure time-series data as the process deviates from the stable operating condition to unstable operating condition; (b) shows the spectral decomposition of the 1-step stochastic matrix for the data under stable and unstable operating conditions.

matrix calculated from the data during the stable and unstable behavior, irrespective of the combustor length and inlet velocity. During stable conditions, the Eigen values very quickly go to zero as compared to the unstable operating condition. This suggests that the size of temporal memory of the discretized data increases as the system move to the unstable operating condition. This indicates that under the stable

operating condition, the discretized data behaves as symbolic noise as the predictive power of Markov models remain unaffected even if the order of the Markov model is increased. On the other hand, the predictive power of the Markov models can be increased by increasing the order of the Markov model during unstable operating condition, indicating more deterministic behavior. An $\varepsilon = 0.05$ is chosen to estimate the depth of the Markov models for both the stable and unstable phases. Correspondingly, the depth was calculated as 2 and 3 for the stable and unstable conditions (see Figure 4).

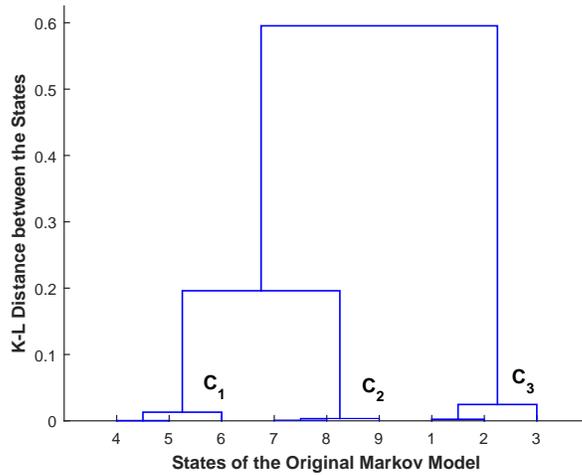
The corresponding $D(\varepsilon)$ is used to construct the Markov models next. First a PFSA whose states are words over \mathcal{A} of length $D(\varepsilon)$ is created and the corresponding maximum-likely parameters (\mathbf{M} and $\mathbf{\Pi}$) are estimated. Then, the hierarchical clustering algorithm using K-L distance is used to cluster and aggregate the states. It is noted that individual models are created for every sample of data, i.e., every sample is partitioned individually so that the symbols will have different meaning for every sample. Consequently, each sample will have a different state-space when viewed in the continuous domain. Thus, the mean behavior of the samples is not shown during any operating regime as the state-space would be inconsistent (even though the cardinality could be the same).

Figure 5 shows the hierarchical cluster tree that details the structure of the state-space for the PFSA with depth $D(\varepsilon)$ for a typical sample during stable and unstable behavior. The cluster tree also suggests the symbolic noise behavior of the data during the stable regime (the states are very close to each other based on the K-L distance). However, clearly a coarse clustering of states in the model during the unstable behavior would lead to significant information loss (as the states are statistically different). However, to compare the two Markov models, the cardinality of the final models are kept the same. The algorithm is terminated with 3 states in the final Markov model during the stable as well as the unstable regime. The parameters of the final Markov model are then estimated using the PFSA models of depth $D(\varepsilon)$ using the dynamic Bayesian network approach explained in Section 3.2.

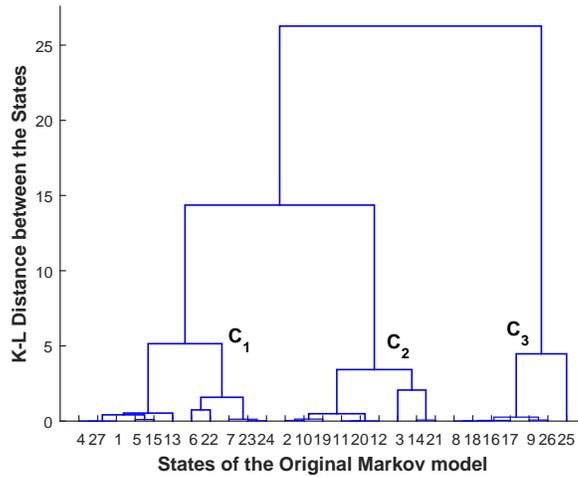
Figure 6 presents some results to show the class separability and changes in the Markov models as the states are aggregated. As the model is computed individually for every sample, comparing the stochastic matrices directly is not consistent. Instead a measure is introduced to model the complexity of the Markov model for every sample as follows: $\delta = \max_{q_i, q_j \in \mathcal{Q}} d(q_i, q_j)$ (where d is defined in equation (5)).

Essentially, the measure δ represents the maximum divergence between the symbol emission probabilities from the states of the Markov model created. Then, the statistics of δ obtained for the Markov models are compared during stable and unstable conditions. In Figure 6a, it could be seen that the measure δ is clearly able to separate the stable and unstable conditions with the original model.

Figure 6b shows the results with the final aggregated model with just 3 states. As seen in Figure 6b, there is some information loss upon model reduction; however, there is still good class separability. Another point to note is that while there is significant change in the set with unstable operating condition, there isn't much change in the behavior of the measure during stable operating condition.



(a) Hierarchical cluster tree of the states during stable behavior



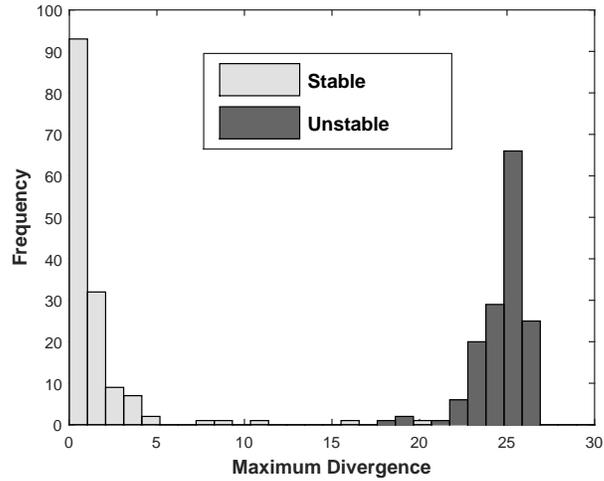
(b) Hierarchical cluster tree of the states during unstable behavior

Fig. 5: (a) shows hierarchical cluster tree for the states of the original Markov model obtained during stable combustion process; (b) shows the same during unstable behavior.

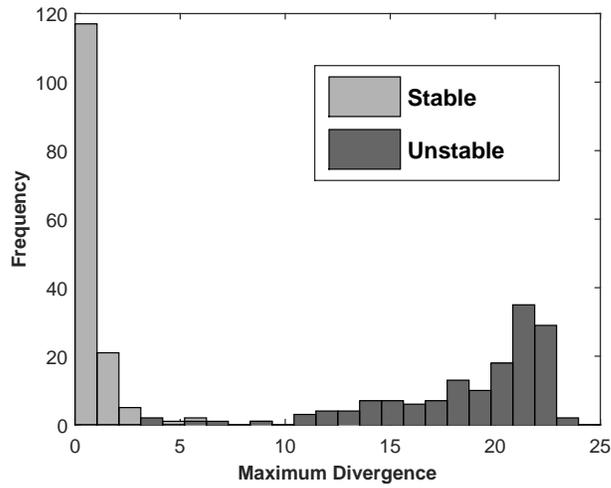
6 Conclusions and Future Work

This chapter has presented a methodology for Markov modeling of time-series data for dynamic data-driven application systems (DDDAS) [3,4]. The technical approach is based on the concepts of symbolic dynamics, where the memory size of the discretized time-series data is estimated to infer the equivalence class of states

based on KL distance. The proposed concepts have been tested on experimental data from a swirl-stabilized combustor apparatus used to study unstable thermoacoustic phenomena during the combustion process. The proposed approach affords the complexity of inferring the time-series data based on a Markov model. Use of Bayesian methods to infer models with various end objectives (e.g., class separability and clustering) is a topic of future research. Another important topic of future



(a) Histogram for the measure δ for the original model



(b) Histogram for the measure δ for the reduced model

Fig. 6: (a) shows the class separability with the original bigger model; (b) shows changes after state aggregation.

work is consistency analysis of the spectral method for memory estimation of the considered class of Markov models.

Acknowledgments

The authors would like to thank Professor Domenic Santavicca and Mr. Jihang Li of Center for Propulsion, Penn State for kindly providing the experimental data for combustion used in this work. Benefits of technical discussion with Dr. Shashi Phooha at Penn State are also thankfully acknowledged.

References

1. P. B. Graben, "Estimating and improving the signal-to-noise ratio of time series by symbolic dynamics," *Physical Review E*, vol. 64, no. 5, p. 051104, 2001.
2. C. S. Daw, C. E. A. Finney, and E. R. Tracy, "A review of symbolic analysis of experimental data," *Review of Scientific Instruments*, vol. 74, no. 2, pp. 915–930, 2003.
3. F. Darema, "Dynamic data driven applications systems: New capabilities for application simulations and measurements," in *5th International Conference on Computational Science - ICCS 2005*, Atlanta, GA; United States, 2005.
4. S. Sarkar, S. Chakravarthy, V. Ramanan, and A. Ray, "Dynamic data-driven prediction of instability in a swirl-stabilized combustor," *International Journal of Spray and Combustion*, vol. 8, no. 4, pp. 235–253, 2016.
5. A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection," *Signal Processing*, vol. 84, no. 7, pp. 1115–1130, July 2004.
6. K. Mukherjee and A. Ray, "State splitting and merging in probabilistic finite state automata for signal representation and analysis," *Signal Processing*, vol. 104, pp. 105–119, 2014.
7. I. Chattopadhyay and H. Lipson, "Abductive learning of quantized stochastic processes with probabilistic finite automata," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, p. 20110543, 2013.
8. C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
9. C. R. Shalizi and K. L. Shalizi, "Blind construction of optimal nonlinear recursive predictors for discrete sequences," in *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, ser. UAI '04, 2004, pp. 504–511.
10. Y. Seto, N. Takahashi, D. K. Jha, N. Virani, and A. Ray, "Data-driven robot gait modeling via symbolic time series analysis," in *American Control Conference (ACC), 2016*. IEEE, 2016, pp. 3904–3909.
11. K. Deng, Y. Sun, P. G. Mehta, and S. P. Meyn, "An information-theoretic framework to aggregate a markov chain," in *American Control Conference, 2009. ACC'09*. IEEE, 2009, pp. 731–736.
12. B. C. Geiger, T. Petrov, G. Kubin, and H. Koepl, "Optimal kullback–leibler aggregation via information bottleneck," *Automatic Control, IEEE Transactions on*, vol. 60, no. 4, pp. 1010–1022, 2015.
13. M. Vidyasagar, "A metric between probability distributions on finite sets of different cardinalities and applications to order reduction," *Automatic Control, IEEE Transactions on*, vol. 57, no. 10, pp. 2464–2477, 2012.
14. Y. Xu, S. M. Salapaka, and C. L. Beck, "Aggregation of graph models and markov chains by deterministic annealing," *Automatic Control, IEEE Transactions on*, vol. 59, no. 10, pp. 2807–2812, 2014.

15. A. Srivastav, "Estimating the size of temporal memory for symbolic analysis of time-series data," *American Control Conference, Portland, OR, USA*, pp. 1126–1131, June 2014.
16. D. K. Jha, A. Srivastav, K. Mukherjee, and A. Ray, "Depth estimation in Markov models of time-series data via spectral analysis," in *American Control Conference (ACC), 2015*. IEEE, 2015, pp. 5812–5817.
17. D. K. Jha, "Learning and decision optimization in data-driven autonomous systems," Ph.D. dissertation, The Pennsylvania State University, 2016.
18. D. K. Jha, A. Srivastav, and A. Ray, "Temporal learning in video data using deep learning and Gaussian processes," *International Journal of Prognostics and Health Monitoring*, vol. 7, no. 22, p. 11, 2016.
19. S. Sarkar, D. K. Jha, A. Ray, and Y. Li, "Dynamic data-driven symbolic causal modeling for battery performance & health monitoring," in *Information Fusion (Fusion), 2015 18th International Conference on*. IEEE, 2015, pp. 1395–1402.
20. J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing SAX: a novel symbolic representation of time series," *Data Mining and Knowledge Discovery*, vol. 15, no. 2, pp. 107–144, October 2007.
21. D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding*. Cambridge University Press, 1995.
22. S. Garcia, J. Luengo, J. A. Saez, V. Lopez, and F. Herrera, "A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 99, no. PrePrints, 2012.
23. V. Rajagopalan and A. Ray, "Symbolic time series analysis via wavelet-based partitioning," *Signal Processing*, vol. 86, no. 11, pp. 3309–3320, November 2006.
24. R. Xu and D. Wunsch, "Survey of clustering algorithms," *Neural Networks, IEEE Transactions on*, vol. 16, no. 3, pp. 645–678, 2005.
25. V. N. Vapnik, *Statistical learning theory*. Wiley New York, 1998, vol. 1.