Taylor & Francis
Taylor & Francis Group

Check for updates

Early Prediction of Lean Blowout from Chemiluminescence Time Series Data

Sudeepta Mondal[a,b], Somnath De[c], Achintya Mukhopadhyay[c], Swarnendu Sen[c], and Asok Ray[a,b]

[a]Department of Mechanical Engineering, The Pennsylvania State University, PA, USA; [b]Department of Mathematics, The Pennsylvania State University, PA, USA; [c]Department of Mechanical Engineering, Jadavpur University, Kolkata, India

ABSTRACT

Lean combustion systems are usually employed in gas-turbine engines, particularly in land-based applications, to reduce NOx emission. These combustion systems are often susceptible to lean blowout (LBO), which can be detrimental for operation and productivity of gas-turbine engines. An abrupt decrease in the equivalence ratio during a throttling operation, which is often encountered in power plants due to sudden decrease in load demand and also in aircraft engines at the time of landing, may lead to an unexpected LBO. From this perspective, online data-driven algorithms are deemed necessary for early prediction of potential transitions to near-blowout conditions. This procedure would provide the human operator/active controller with an appropriate lead time to alter the operating conditions so that the system can be brought back to a desired stable condition. The paper emulates pertinent conditions of LBO on a laboratory-scale apparatus of swirl-stabilized dump combustor with transient time series of CH* chemiluminescence data, where the objective is early prediction of LBO. The underlying algorithms are constructed based on a well-known statistical learning tool, called Hidden Markov Modeling (HMM), which can be used in the setting of supervised learning to discern near-blowout time series data from stable data. Being solely data-driven, the proposed methodology is model-free; it has been shown to be numerically efficient as well as sensitive to regime changes when the combustion system moves toward or away from LBO.

Introduction

With growing concerns over pollutant emission, lean combustion technology has been accepted as a promising alternative to alleviate high levels of NOx emissions (Correa 1993). Nevertheless, lean-premixed combustion can be susceptible to flame extinction due to the reduced flame speed against high flow rates of unburnt gases (Plee and Mellor 1979; Radhakrishnan, Heywood, Tabaczynski 1981), which is known as lean blowout (LBO). The effects of LBO are detrimental for both land-based and aircraft gas-turbine engines. In power plants, the penalty associated with LBO is very expensive when flame extinguishes abruptly during lower requirements of load (Meegahapola and Flynn 2015). The blowout phenomenon can promote the generator shutdown which leads to the frequency stability

issues (NER, June, 2008; FRC, Feb, 2008). A major risk develops in aircraft applications when the combustor faces a variation in power settings. During the landing period, based on lower load requirements, fuel-flow rate (FFR) needs to be decreased; but often, the reduction rate in fuel flow occurs rapidly. On the other hand, slower transience of the air-flow rate (AFR) due to compressor inertia makes the mixture so lean that the fuel-air mixture may reach the lower end of the flammability limit (Rosfjord and Cohen 1995). Thus, prior to landing, the flame may blow out and can result in an unexpected loss of power, and even worse, the engine shutdown. Therefore, early identification of the imminent lean blowout is essential to circumvent such situations.

Significant research on LBO has been carried out under National Jet Fuels Combustion Program (NJFCP), a multi-agency program in the United States, steered by the Federal Aviation Administration (FAA). Some of the significant initiatives taken under this program attempted to study the effects of different alternative fuels on the LBO signature (Peiffer, Heyne, Colket 2019; Zheng et al. 2019). For that purpose, Zheng et al. (2017) considered the performance near LBO as one of the major criteria for the evaluation of candidate fuels. The consideration of LBO behavior for using different candidate fuels was further explored by Esclapez et al. (2017a), where the main objective was to compare the behavior of flame dynamics in terms of OH* emissions using different fuels to evaluate the fuel performance on combustion stability. A few attempts were also made to explore the performance of LBO mechanisms through large eddy simulation (LES) (Cavaliere, Kariuki, Mastorakos 2013; Esclapez et al. 2017b). Recently, the investigation of fuel compositions on the blowout dynamics has been extended to the liquid fuels by Rock et al. (2020), where the pre-LBO signature is observed to be highly dependent on the fuel composition. LBO has been an important issue of concern in the gas turbine industry, and several investigations have been carried out to explore the flame dynamics near blowout regime for the last two decades. The studies have reported that the flame undergoes different oscillatory behaviors as the combustion approaches blowout (Chao et al. 2000; Hertzberg, Shepherd, Talbot 1991; Nair and Lieuwen 2005; Nicholson and Field 1948). It has also been observed that the height of the flame increases toward LBO (Chao et al. 2000; Savas and Gollahalli 1986). Therefore, researchers have also looked into the prediction of blowout through behavioral changes of the flame (Muruganandam et al. 2002; Yi and Gutmark 2007a). Moreover, to get direct information from the flame behavior, chemiluminescence-aided image techniques are often used (Herna´ndez and Ballester 2008; Rock et al. 2020; Sun et al. 2013).

On the other hand, continued investigations were emphasized on quantitative analysis of measurements (e.g., temperature (Kaluri, Malte, Novosselov 2018) and pressure (Barnum and Bell 1993))-based combustion properties to predict the lean blowout. Li et al. (2007) showed that the intensity of low frequency temperature fluctuations can be used for the prediction of lean blowout. Radhakrishnan and Heywood (1980) prescribed a statistical model with Monte-Carlo mixing with an overall rate equation for detection of lean ignition and blowout limits of uniform mixtures. Yi and Gutmark (Yi and Gutmark 2007b) proposed two parameters, normalized root mean square and normalized cumulative duration of OH* chemiluminescence to predict the lean blowout. Muruganandam et al. (2005) used a double threshold-based method for identifying the precursors seen close to LBO. The active control methodology, proposed by Muruganandam et al. (2005), is based on detecting an abrupt reduction in the mean value of the CH* chemiluminescence as a precursor to LBO detection. One immediate drawback of such a threshold-based approach is that the value of the threshold is most likely to

change with a variation in combustion process parameters, let alone experiments with a different combustor. Moreover, monitoring the mean of the signal and detecting a change in the mean is often data-intensive, and might be impractical when the system is imminent to LBO and a transition to blowout is possible in a time-scale of the order of milliseconds. The method, proposed by Yi and Gutmark (Yi and Gutmark 2007b), involves a normalized root-mean-squared (NRMS) metric of normalized chemiluminescence, which can be monitored for online prediction of LBO, where the threshold is subjective to the particular experimental data and is devised for prediction of LBO onset. Recently, De et al. (2019) studied different statistical tools and showed that the mean frequency based on Hilbert transform can serve as an LBO predictor for premixed as well as partially premixed flames. Apart from these statistical tools, techniques (Chaudhari et al. 2013; De et al. 2020b) based on flame color are new additions to the series of LBO predictors.

Apart from these, different nonlinear tools have been developed to study flame behaviors, for example, Recurrence Plots (RP) and Recurrence Quantification Analysis (RQA) (De et al. 2020a; Nair, Thampi, Sujith 2014; Unni and Sujith 2016), phase plots (Kabiraj et al. 2012; Mondal, Pawar, Sujith 2017b; Sen et al. 2018), Lyapunov exponent (Gotoda et al. 2012b; Kabiraj et al. 2015, 2012), permutation entropy (Gotoda et al. 2012a), translational error (Gotoda et al. 2014), Hurst exponent (Nair and Sujith 2014) and so on. The parameters have the ability to signify the notable changes of the flame dynamics. However, few of them are shown as suitable candidates for being implemented in LBO control applications (Domen et al. 2015; Gotoda et al. 2016).

In addition, different data-driven approaches (e.g., cross wavelet transform (XWT) (Dey et al. 2015), and Kullback-Leibler divergence (Sarkar et al. 2014) based anomaly measure employing finite memory Markov modeling) were used to predict the proximity to LBO. The investigation of Dey et al. (2015) also performed an accuracy test of the XWT-based method and showed that the strategy has ability to predict the approach to LBO. However, the work did not highlight on the computational performance and also on the suitability of the method in online LBO control application. Mukhopadhyay et al. (2013) established a strategy, which is based on symbolic time series analysis (STSA) of the CH⋆ emission of the flame collected through optical sensor. They showed that the anomaly measure, defined as the deviation of the current state vector from the reference state vector, can be used for LBO prediction. It is worth noting that data-driven methods, which have previously been employed for online prediction of LBO, are mostly based on amplitude-based thresholding heuristics applied to the chemiluminescence signals. Moreover, the sizes of time windows used in almost all of the available data-driven analysis for LBO detection in literature are typically in the order of multiples of seconds, which might incur significant data overhead for detection purposes, particularly in such a high-stake application where early detection of incipient LBO is of utmost priority for giving the controller an appropriate lead time to implement corrective measures to stabilize the system.

The current paper proposes a dynamic data-driven method that can be used for early prediction of LBO and also for regime identification when the combustion system stabilizes from being close to the LBO limit. In this context, "regime detection" refers to the classification of two different kinds of transient regimes in the combustion system: 1) transition from stable mode of operation toward LBO; and 2) transition from near LBO to being stable. The central idea in this work is to solely use the experimentally procured chemiluminescence data from a laboratory-scaled swirl-stabilized dump combustor to train

different latent-variable models, called Hidden Markov Models (HMMs), and to employ them for regime classification. HMMs have been successfully used as a statistical learning tool for speech recognition (Rabiner 1989); specifically, HMMs are used to solve an inference problem, whereby a doubly stochastic Markov chain with unobserved states is used to model the temporal behavior of the available data regimes (Zhong and Ghosh 2002). HMMs have also been used as online predictive monitoring techniques for a laboratory-scaled Rijke tube apparatus (Mondal et al. 2019), pulverized coal-based combustion systems in furnaces (Chen et al. 2011), and gas turbine engines (Menon et al. 2003). An HMM, trained on data from a particular regime of operation, is capable of modeling the inherent statistical behavior of that regime. With a bank of HMMs, pre-trained with respect to the regimes of *a-priori* known operation, it is possible to compute the likelihoods of observed data sequences with respect to each of the pre-trained models. This computation can be performed almost in real-time for monitoring purposes whereby sequences of chemiluminescence data windows can be processed to find the most likely regime to which it belongs, based on the likelihoods calculated by each HMM. Thus, if a combustion system approaches LBO, the likelihood of the "stable" HMM model would decrease with respect to a "blowout" HMM model. Since the HMM likelihoods are very sensitive to regime changes, this decrease would be rapid and serve as an online predictor of the imminent LBO. In addition to detecting the transience toward LBO regime, this HMM-based online predictor must make sure that the false alarm rates are minimized. That is, it must be able to discriminate the transience of the system if it starts becoming stable from the near LBO phase. The HMM-based prediction algorithm proposed in this paper has been shown to be robust to the two different types of transients, and have been demonstrated to identify the system regime with significant accuracy while exhibiting parsimonious data requirements.

The proposed LBO prediction methodology is free from subjective thresholding heuristics, because the regime classification methodology is based on choosing the HMM that has the maximum likelihood of generating a particular test data signature among all the HMMs that have been pre-trained with data from different regimes. So, the likelihood ratio-based metric devised in this paper in threshold-free due to its formulation which does not need to be changed across different types of experiments. Moreover, a comparison has been provided whereby the early LBO prediction has been demonstrated with significantly lower data requirements than that of the NRMS metric proposed by Yi and Gutmark (2007b), or the RQA-based metric (Recurrence Rate) proposed by Unni and Sujith (2016). This is accomplished with considerably higher prediction accuracy, particularly owing to the fact that HMM-based metrics are significantly more sensitive to regime changes than those based on traditional statistics (e.g., mean and RMS) of the chemiluminescence data. Moreover, a comparison of online computational expenses of the HMM-based metric with NRMS and RR has been presented.

The paper is organized in six sections including the present section. Section II describes the laboratory apparatus and the experimental procedure for validation of the theoretical results. Section III describes the technical approach for early prediction of the proximity to LBO from transient time series. Section IV presents the results of early LBO detection on the experimental data procured from the laboratory apparatus described in Section II along with pertinent discussions. The paper is summarized and concluded in Section V along with recommendations for future research. Appendix A summarizes the essential mathematical concepts that form the backbone for analysis of stochastic dynamical behavior of LBO.

Laboratory apparatus and experimental procedure

This section describes the laboratory apparatus and the experimental procedures for chemiluminescence data generation on which HMM technique has been applied for early detection of LBO.

Description of the laboratory apparatus

A swirl-stabilized dump combustor, detailed in the earlier reported work (De et al. 2019), has been used in this study, which represents a laboratory-scaled model of a generic gas-turbine combustor. The schematic of the combustor rig is shown in Figure 1. The apparatus has three major components: a premixing chamber (PC), a combustion chamber (CC), and an exhaust chamber (EC), which are coaxially aligned; and the entire body of the combustor is made of stainless steel (SS) type SS316 to resist corrosion at high temperatures. The fuel-air mixing takes place in the PC before the mixture enters into the CC. The fuel enters at a distance of 330 mm from dump plane, and the air inlets are located at a distance of 20 mm from the upstream end of the premixing chamber, as shown in Figure 1. The distance between air and fuel entrance points is 50 mm. There are four circumferential holes at respective entrances of both air and fuel to enhance the flow symmetry. To ensure good mixing, a swirler is placed at a distance of 15 mm upstream of the dump plane. The swirl number (SN) used here is 1.26 and the angle of swirler blade is 60° relative to the axial direction (De et al. 2019, 2020b).

The CC is fabricated to provide optical access to the swirling premixed flame. A quartz tube of 65 mm outer diameter and 200 mm length is held between two grooved flanges, which serves as the flame observation window. At the base of the CC, a provision is made for sparking the flame during the start of an experiment. The flow rate of fuel and air mixture is measured using Alicat Mass flow controllers (range: 0 to 250 SLPM, and 0 to 1500 SLPM, for liquid petroleum gas (LPG) and air, respectively). The uncertainty in the flow measurement is $\pm$ (0.8% Reading $\pm$ 0.2% of Full Scale). The equivalence ratio (ϕ) is defined as:

$$\phi = \frac{\dot{M}_{F,a}/\dot{M}_{A,a}}{\dot{M}_{F,s}/\dot{M}_{A,s}}$$

where $\dot{M}$ represents the mass flow rate; the subscripts A and F represent the air and fuel, respectively; and the subscripts s and a represent the stoichiometric and actual conditions, respectively, at the time of measurement of both fuel and air flow rates. For the ease in comparison, the results in Section 4 are described in terms of normalized equivalence ratio ϕ/ϕ_{LBO}. The blowout equivalence ratio, ϕ_{LBO} is defined as the lowest equivalence ratio at which the flame is observed. Any further reduction in fuel flow rate or increase in air flow rate may lead to the disappearance of the flame. Thus, the actual equivalence ratio at which the flame blows out is within a range of 0.04 (determined by the least counts of MFCs) of the last recorded value, ϕ_{LBO}.

Flame behavior

Considering the changes in flame dynamics in response to variations in the flow condition, the experimental period, including the initiation of flame to the onset of LBO, can be
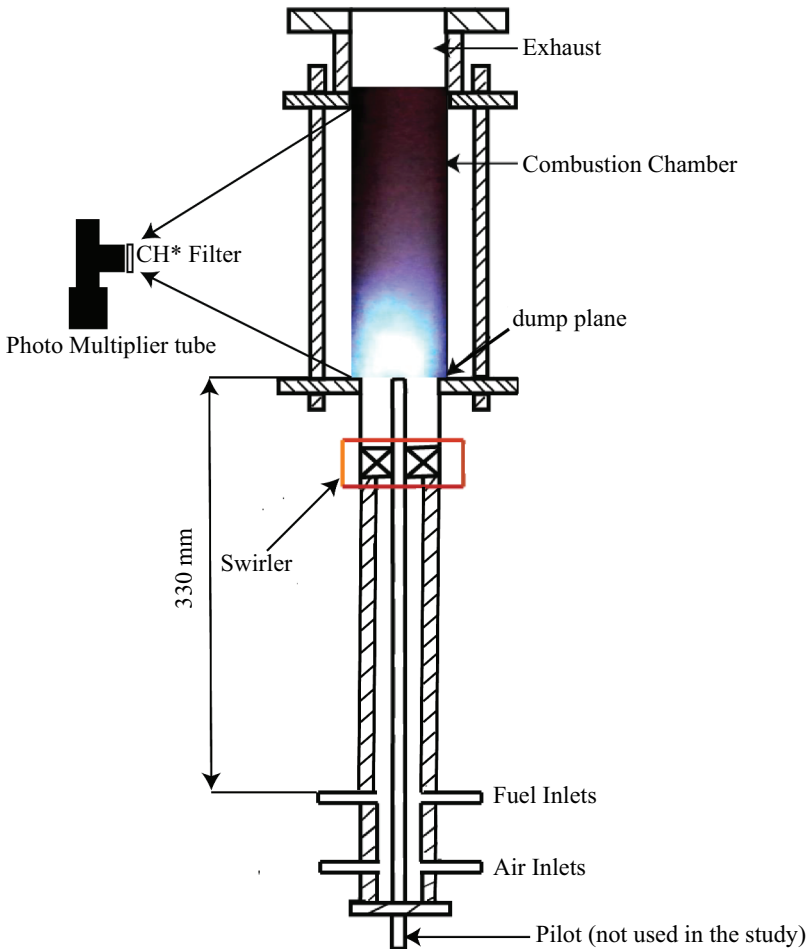
Figure 1. Schematic diagram of the experimental apparatus, incorporating a CH* filter mounted on the photo multiplier tube.

divided into three regimes, namely stable, transition, and near LBO. The start and end points of these regimes can be varied at different fixed FFRs by increasing the AFRs in steps. In this section, the typical flame behavior in the three regimes are described for the case shown in Figure 3, for which FFR has been kept fixed at 3.5 SLPM while AFR has been increased from 105 SLPM to 205 SLPM in steps of 5 SLPM after every 30 s. However, variations of the flame dynamics from higher equivalence ratio to LBO (by variation of AFR) for other cases with fixed FFRs exhibit similar characteristics.

1) *Stable Regime: Near Stoichiometric Condition*

At higher equivalence ratios, the flame looks symmetric where an outer conical envelop shines inside the combustion chamber (CC). The structure of the flame at $\phi/\phi_{LBO} = 1.52$ is shown in Figure 2(a). The base of the flame still anchors at the dump plane while changing ϕ/ϕ_{LBO} from 1.46 to 1.41 (Figure 2(b,c)).
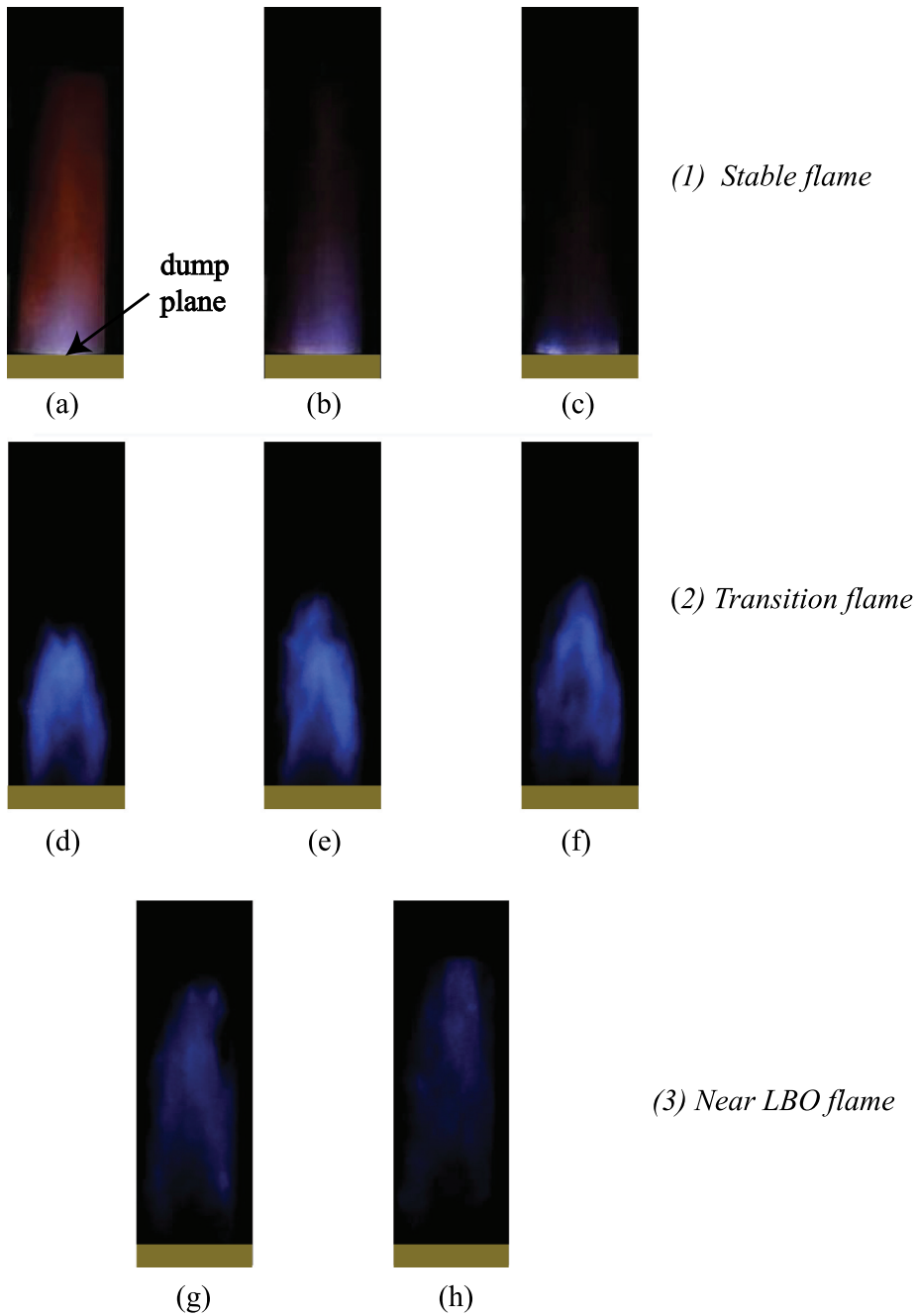
Figure 2. The display of flame behavior in the three regimes in response to variation in AFR at a fixed FFR of 3.5 SLPM. The dynamics of the stable flame (1) are shown at: (a) $\phi/\phi_{LBO} = 1.52$; (b) $\phi/\phi_{LBO} = 1.46$; and (c) $\phi/\phi_{LBO} = 1.41$. The transition flames (2) are presented at: (d) $\phi/\phi_{LBO} = 1.28$,; (e) $\phi/\phi_{LBO} = 1.21$; and (f) $\phi/\phi_{LBO} = 1.17$. The lean flames near LBO (3) are shown at: (g) $\phi/\phi_{LBO} = 1.08$; and (h) $\phi/\phi_{LBO} = 1.05$.

2) *Transition Regime: Flame Transition Condition*

As the equivalence ratio decreases, the effect of the slower reaction rate possibly alters the flame dynamics. At $\phi/\phi_{LBO} = 1.28$ the flame base is not attached to the burner inlet, rather it moves further downstream (Figure 2(d)). The flame at this stage stands inside the CC at a certain height from the dump plane. The weak attachment of the flame is further observed in the Figure 2(e,f). In this work, we define the *transition* or *bifurcation* point where the conical shape of the reddish flame disappears and instead of that, a prominent blue lifted or elongated flame is visible (De et al. 2019, 2020b)., 3) *Near Lean Blowout (LBO) Regime: Flame Blow-off Condition*

As the AFR is increased gradually, the flow reaches a stage when the air-fuel ratio becomes high enough to cause frequent pulsation of the flame base, accompanied by local extinction and re-ignition (Chaudhari et al. 2013; De et al. 2019, 2020b; Nair and Lieuwen 2005). As the equivalence ratio is further reduced, these events are observed more frequently (Nair and Lieuwen 2005). The detailed analysis of these pre-LBO events using flame imaging can be found in a previous work by some of the co-authors (De et al. 2019, 2020b), and is not presented here further for brevity. The images of the near LBO flames are shown in Figure 2(g,h).

The dynamics of the flame behavior at the three regimes (mentioned above) for constant FFR are exhibited in videos provided as Supplementary materials. V01.mp4, V02.mp4 and V03.mp4 show the flame behavior at the stable, transient and near LBO regimes respectively for the constant FFR = 3.5 SLPM case. Similar behavior of the flame at those regimes for constant AFR = 80 SLPM is observed in videos V04.mp4, V05.mp4 and V06.mp4 respectively.

Experimental procedure

In these experiments, the fuel is liquid petroleum gas (LPG) whose composition is 60% butane (C_4H_{10}) and 40% propane (C_3H_8) by volume (Gas, Accessed on June 20, 2020). The variation in the characteristics of a premixed flame is captured through chemiluminescence data, which are assumed to be proportional to the heat release rate (Hardalupas and Orain 2004; He et al. 2019; Liu et al. 2019). The chemiluminescence signal is obtained from the flame directly using a photo multiplier tube (PMT), as seen in Figure 1. Thus, the effort to understand the flame dynamics through flame emission is non-intrusive, where the sensor does not need to be placed inside the high-temperature combustion chamber. However, earlier literature had provided evidence of examining the flame behavior using heat release data (Keller and Saito 1987; Lawn 2000; Price, Hurle, Sugden 1969). A PMT (model 931A, Hamamatsu Corporation), fitted with a CH* filter (a narrow bandpass filter centered at 430 nm with a bandwidth of ± 10 nm), is kept in front of the CC in such a way that it can collect only CH* emission from the flame. The voltage output from the PMT is collected by interfacing a c-RIO data acquisition system (Serial No. 9073) with the computer through NI-LabVIEW 2015 at the sampling frequency of 4.096 kHz. The chemiluminescence data is captured as the equivalence ratio is programmed to decrease in small steps.

Lean blowout can be approached in two ways: either by keeping the fuel flow rate (FFR) constant and increasing the air flow rate (AFR), or by keeping the AFR constant and decreasing the FFR. Referring to Section IV, two types of LBO data have been collected during the experimental procedure. In the first type, as shown in Figure 3 and reported in Table 1, the fuel flow has been kept constant, whereas the AFR has been increased in steps of 5 standard
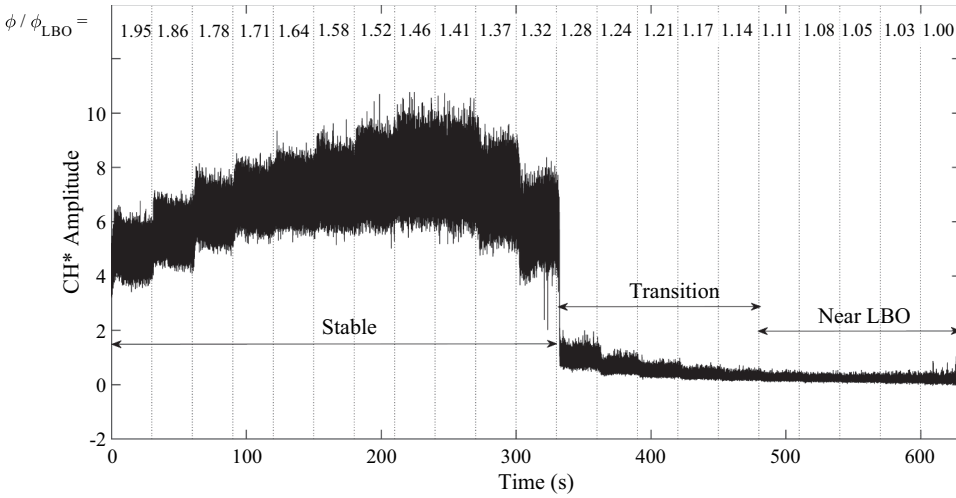
Figure 3. Experimental data collected over 630 s with FFR kept fixed at 3.5 SLPM and increasing the AFR from 105 SLPM to 205 SLPM in steps of 5 SLPM.

liters per minute (SLPM), which results in a data signature that moves from stable toward LBO phase, through a transience regime. The step-duration of AFR increase is controlled at 30 s using a program running in FlowVision, a software that controls the mass flow controllers for the fuel and air. With such a step change in AFR the change in equivalence ratio is about 0.04 near stoichiometry and 0.02 near LBO. On the other hand, in Figure 8, the equivalence ratio is gradually moved toward the lean zone by reducing FFR in steps of 0.1 SLPM from near stoichiometric condition after every 5 s (fastest ramping time allowable by the system hardware), while AFR is kept fixed at 110 SLPM. For the data used in Figure 5, the AFR has been kept fixed at 80 SLPM, while the FFR is reduced in steps of 0.1 SLPM till $\phi/\phi_{LBO} = 1.083$, after which the FFR is increased again in steps of 0.1 SLPM till $\phi/\phi_{LBO} = 1.42$. This results in a data signature that moves into the transience phase before LBO, after which the stability in the flame is recovered by a gradual increase in FFR.

The reason for demonstrating the online prediction performance with two different types of data is to check the efficacy of proposed loglikelihood ratio (LLR)-based method not only during the phase of imminent LBO, but also when the combustion system recovers from the LBO phase after a short excursion event. Unlike the monotonic transition toward LBO, this scenario may not need any corrective action. The distinction between the two types of events would save unnecessary control efforts along with the associated costs. The online

Table 1. Duration of transience to LBO and lead time in the prediction of imminent LBO for different cases.

FFR (SLPM)	2.5	2.8	3.0	3.2	3.5
Transient to LBO limit (s)	120 − 310	210 − 310	210 − 480	230 − 508	330 − 630
Lead time of prediction (s)	120	120	160	165	180

condition monitoring system should be sufficiently fast to discriminate between these two different types of transients in order to reduce the chances of false alarms.

Technical approach: early prediction from transient data

This section describes the technical approach for early prediction of the proximity to LBO in transient time series data. The problem is addressed in the setting of the two different types of experimental data, as discussed in Section II. From the first type of datasets, whereby the experiments were conducted such that the fuel-air mixtures were made leaner progressively (by varying AFR while keeping FFR fixed at values ranging from 2.5 SLPM to 3.5 SLPM), chunks of data belonging to each of the three regimes, namely, stable, transition and near-LBO, were extracted. Each of these datasets comprises data in multiples of 30 s windows at a sampling rate of 4096 Hz. From all the available time windows (in multiples of 30 s), randomly selected 40% of the data from each regime were used for training three HMMs: λ_i, for $i \in \{1, 2, 3\}$, where λ_1 corresponds to the stable regime of operation, λ_2 corresponds to the transition regime, and λ_3 corresponds to the near-LBO regime, using Baum-Welch algorithm as described in Appendix A. The pertinent mathematical formulation is described in Appendix A. Instead of using individual sequences to learn a particular HMM, the training is performed with ensemble data, e.g., ensemble of stable time series data across different experiments were used to learn λ_1. This ensures that an HMM for a particular regime learns the overall pattern of the chemiluminescence signature in that regime and does not overfit to a particular response signature. It is to be noted that the HMM training is performed only once, with one set of experimental data for which AFR has been increased in ramps of 30 s. Test data of similar type has been used in the analysis discussed in Section IVA which refers to different experiments of the same type whereby the AFR increment occurs in 30 s steps. But for the analysis in Section IVB and IVC, the type of experiments performed is different, with the FFR being changed after every 5 s instead of changing AFR in steps of 30 s. But even for a difference in experimental conditions, the same pre-trained HMMs have been found to generalize, as will be discussed in Section IV.

These learned HMMs serve as generative models for the respective datasets, which are based on the domain knowledge procured by the experiments. The HMM-based approach presented in this paper is completely data-driven, and is intended to learn the trend in the data available, which can be used to train the models. Unlike model-based approaches, it is not limited to simplifying assumptions of the thermophysics involved in the LBO phenomenon. Therefore, with a sufficiently dense set of experimental training data procured from different regimes, this approach can be used to learn the stochastic behavior of the system in the different regimes. Since these generative models are learnt from an ensemble of data belonging to these different regimes, the modeling is robust to issues like variations in operational conditions and differences in sensor noise levels for even a single mode of operation.

With the pre-trained models λ_i, for $i \in \{1, 2, 3\}$, the chemiluminescence data have been used from the test data set for early prediction of the onset of LBO. Given a window of time series data $\{y_1, \ldots, y_T\}$, the algorithm calculates the following log-likelihood ratio (LLR) using Forward Procedure:

$$\mathcal{L}_{k,m} = log\left[\frac{(p(y_{1:T})|\lambda_k)}{(p(y_{1:T})|\lambda_m)}\right] = log(p(y_{1:T})|\lambda_k) - log(p(y_{1:T})|\lambda_m) \stackrel{\Delta}{=} [L_k - L_m] \qquad (1)$$

where k, m = 1, 2, or 3, k ≠ m, and $(p(y_{1:T})|\lambda_k)$ denotes the probability that the observed chemiluminescence time series sequence is generated by model λ_k. This paper primarily focuses on the prediction of the onset to LBO during the transition phase. From the dataset that has been collected in the experiments, it has been seen that there is a distinct decrease in amplitude of the CH^* chemiluminescence as the system moves into the transition zone from the stable regime. This results in a sharp increase of $\mathcal{L}_{3,1}$ or $\mathcal{L}_{2,1}$, which unequivocally corresponds to an almost sure distinction of the LBO regime from the stable behavior. On the other hand, there is a slow and inconspicuous change in the amplitude and dynamics of the data from the transition to the near-LBO regime, which renders the problem of predicting the proximity to LBO in the transition zone even more challenging, particularly in a real-time scenario where the data windows available for analysis are short and limited. This problem is addressed by tracking the temporal behavior of appropriate LLRs as the data windows scan through the transition zone into the LBO limits.

Results and discussions on early LBO detection

This section presents the results of experimentation conducted on the laboratory apparatus described in Section II along with pertinent discussions in the following three subsections.

Early prediction of LBO in the transition regime

This subsection presents the results for prediction of LBO during the transition phase for five different experiments, where the FFRs were kept fixed at 2.5 SLPM, 2.8 SLPM, 3.0 SLPM, 3.2 SLPM and 3.5 SLPM, respectively. Representative results are shown with the chemiluminescence data obtained by increasing the AFR from 105 SLPM to 205 SLPM while keeping the FFR fixed at 3.5 SLPM. Similar results for the other cases are enlisted in Table 1. Time series data of chemiluminescence have been collected over 630 s with the FFR ratio kept fixed at 3.5 SLPM; the data profile is shown in Figure 3. The three regimes of data are marked: a) Stable regime from 0 to 330 s, b) Transition regime from 330 s to 480 s and c) Near LBO regime from 480 s to 630 s. The AFR has been increased from 105 SLPM to 205 SLPM in steps of 5 SLPM, where each step of increase is kept fixed for 30 s. For this case, the LBO happens at ∼630 s when the AFR is 205 SLPM.

Referring to Section III, early LBO prediction within the transient period relies on tracking the evolution of the LLR, $[L_3 - L_2]$, calculated with a fixed window length from the time series data. The evolution of $[L_3 - L_2]$ is shown in Figure 4 for a window length of 1000; that is, the LLR is calculated after every 1000 samples of chemiluminescence data recording. A positive $[L_3 - L_2]$ for a dataset indicates a higher likelihood of belonging to a near-LBO regime. This implies that the operator/controller will have the indication of a forthcoming LBO and thus may take appropriate actions to alleviate the problem. The lead time in prediction of LBO corresponds to the time from which $[L_3 - L_2]$ remains consistently positive. It is seen from Figure 4 that $[L_3 - L_2]$ profile becomes positive at 430 s, but it oscillates between positive and negative values from 430 s to 450 s. From 450 s onwards the $[L_3 - L_2]$ profile remains positive and starts to increase rapidly as the LBO limit is reached. The system moves from transient to near LBO limit for this case from 330 s to 630 s, and
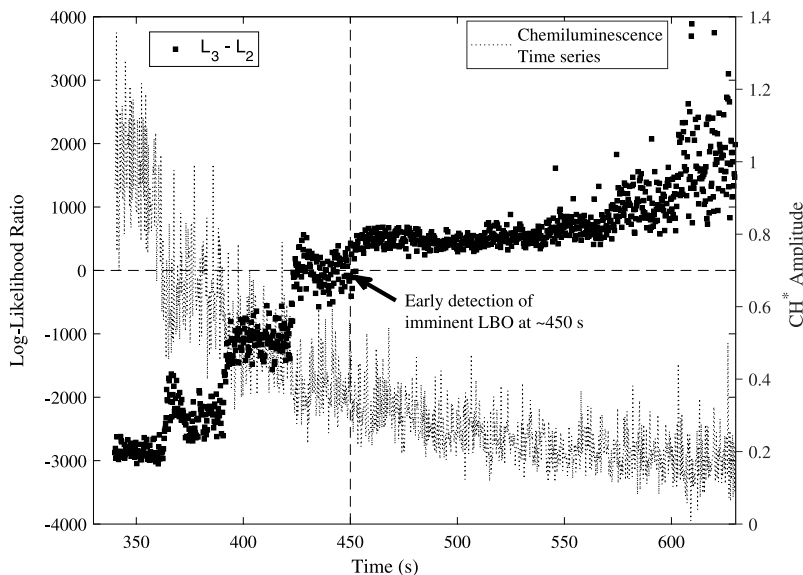
Figure 4. Evolution of $[L_3 - L_2]$ with the chemiluminescence signal from 330 s to 630 s for the case shown in Figure 3. LBO occurs at 630 s. $[L_3 - L_2]$ starts becoming positive at 450 s indicating an increased likelihood of approaching LBO.

hence, there is a lead time ($\sim$180 s) of imminent LBO prediction, during which control actions can be taken for flame stabilization. Table 1 enlists the duration from transient to LBO limit and the lead time of imminent LBO prediction by the tracking of $[L_3 - L_2]$ for all these cases. The least lead time has been obtained with FFR at 2.5 SLPM and the highest lead time with FFR at 3.5 SLPM. This observation follows the intuitive fact that there is a smaller lead time before LBO occurs with a lower FFR as less energy is released, making the flame inherently weaker.

### Classification of data towards and away from the LBO limit

The previous subsection dealt with datasets for which the flame moves from stable to LBO regime before extinguishing completely. Now, if appropriate actions are taken near the LBO limit so that the blowout does not happen and if the system gradually moves toward stability, then the HMM-based algorithm must indicate the recovery in its LLR signature. This subsection extends the proposed HMM-based approach to demarcate two regimes of data, which are close to LBO limit. Figure 5 depicts the experimental data obtained by decreasing the FFR to reduce the ϕ/ϕ_{LBO} from 1.42 to 1.08, after which the FFR is increased to increase ϕ/ϕ_{LBO} from 1.08 to 1.42 in steps of 5 s, as discussed in Section IIC.

The transition region in the time interval of 15 s to 40 s is of interest, where the temporal behavior of the chemiluminescence signal changes from moving toward LBO to going out of the blowout regime. A windowed rate of change, $d(L_3 - L_1)/dt$, is calculated to detect this regime change. A positive rate of change of $[L_3 - L_1]$ in a time window is equivalent to a data signature that corresponds to the flame approaching toward LBO limit. For extracting the LLR rate, a time window of 2.5 s has been chosen. The windows contain data in non-overlapping segments from 15 s to 40 s at intervals of 2.5 s. Within each of these 2.5 s time
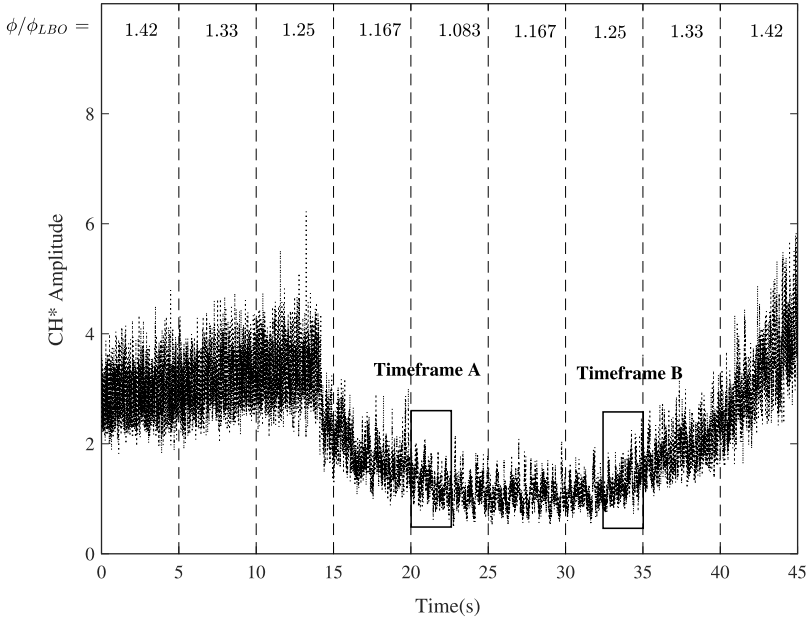
Figure 5. Dataset showing transition from near LBO toward stable combustion. FFR is gradually reduced to near LBO limit ($\phi/\phi_{LBO} = 1.083$), and then increased in steps of 0.1 SLPM till $\phi/\phi_{LBO} = 1.42$, while AFR is kept fixed at 80 SLPM.

windows, a 100-sample-based calculation of $[L_3 - L_1]$ is performed, from which a linear rate is extracted based on least-square regression. Figure 5 shows two timeframes: Timeframe A (from 20 s to 22.5 s) and Timeframe B (from 32.5 s to 35 s), which are plotted on top of another in Figure 6(a). It is seen that the chemiluminescence signature in these two time windows are very similar, but the $[L_3 - L_1]$ signatures in the two timeframes (Figure 6(b,c), respectively) show that there is a positive rate of increase of $[L_3 - L_1]$ in Timeframe A, and a decay of $[L_3 - L_1]$ in Timeframe B. Since $d(L_3 - L_1)/dt$ is estimated via the linear rate (θ) which is subject to uncertainties due to noisy observations within a short time window, statistical behavior of the estimates have been analyzed, and reported as $\theta = \mu \pm 2\sigma$ (corresponding to ~95% confidence band), where μ is the mean of the estimation and σ is the standard deviation. For Timeframe A, $\theta = 723 \pm 289 s^{-1}$, while for Timeframe B, $\theta = -1121 \pm 327 s^{-1}$.
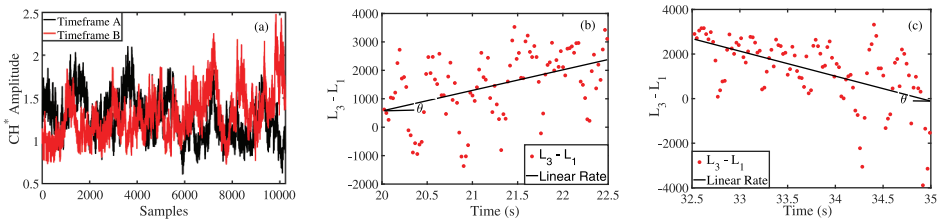
Figure 6. Comparison of results in Timeframe A and Timeframe B. The slope of the fitted line in (b) and (c) is estimated by $\theta = d(L_3 - L_1)/dt$ in Timeframes A and B, respectively.
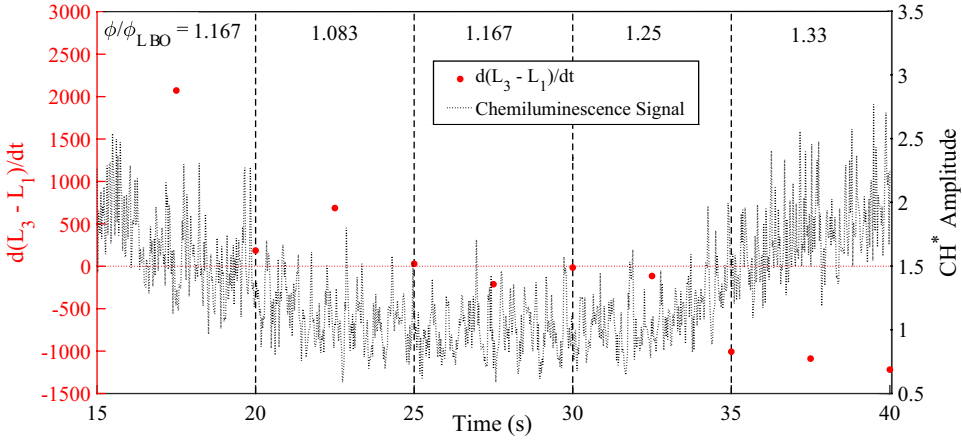
Figure 7. Evolution of rate of change of $[L_3 - L_1]$ with the chemiluminescence signal from 15 s to 40 s for the case in Figure 5. The rate of LLR change (i.e., $d(L_3 - L_1)/dt$) is calculated with a time window of 2.5 s.

The results of $d(L_3 - L_1)/dt$ variation are shown in Figure 7, where vertical dashed lines correspond to the time instants at which the FFR has been changed, indicating a change in ϕ/ϕ_{LBO}. With 2.5 s time windows the time rate of change $d(L_3 - L_1)/dt$ (filled red circles) remains positive till about 25 s, and starts to become negative after that. It is worth mentioning that the temporal behavior indicated by the rate of change of LLR is highly local to the time window being observed, i.e., if the chemiluminescence data show local temporal decay while globally moving toward a stabler regime, it is likely to show local increase in $d(L_3 - L_1)/dt$. Hence, an optimally long enough sample window is required before a data-driven decision can be taken in order to ascertain appropriate control actions, which are robust to the noisy local fluctuations in the temporal signal. So, for this experiment where step changes in ϕ/ϕ_{LBO} occur at 5 s intervals, a time window of ~2.5s has been found to be sufficient to discern the regime of operation of the system based on the chemiluminescence data signature. It is noted that this optimal time window can change from system to system and from one experiment to another because its calculation is essentially data-driven. With available experimental data from a particular system, such an analysis can be useful to have an estimate of how much time is required for tracking the rate of change of LLR for regime detection from online time series data.

Comparison with existing techniques of online LBO detection

Of the various data-driven methods studied till date that have focused on LBO detection from time series data, the prominent ones that have focused on online prognosis of LBO are the works by Yi and Gutmark (2007b) and Unni and Sujith (2016). In this section, we present a comparative study of the classification performance and computational cost of the proposed LLR-based LBO detection algorithm with the aforementioned method.

Yi and Gutmark (2007b) have proposed the tracking of normalized root mean square (NRMS) calculated from the filtered chemiluminescence time series signal as an online condition monitoring metric, where NRMS is defined as:

$$NRMS \triangleq \frac{\sqrt{\sum_{i=1}^{N} \frac{\xi_i^2}{N}}}{\bar{\xi}} \tag{2}$$

where ξ_i denotes the ith sample of the chemiluminescence after filtering the signal with a bandpass filter of bandwidth $2 - 100$ Hz, N denotes the number of samples in the time series data analyzed, and $\bar{\xi}$ denotes the mean of the chemiluminescence signal in the time period under analysis. The LBO warning methodology suggested in Yi and Gutmark (2007b) is based on the following inequality condition:

$$\frac{NRMS(\phi/\phi_{LBO})}{NRMS(\phi/\phi_{LBO} > 1.5)} > \alpha \tag{3}$$

The above inequality is based on an observed behavior of NRMS which remains almost constant when $\phi/\phi_{LBO} > 1.15$ and starts to increase as ϕ/ϕ_{LBO} drops below 1.15 when the flame moves toward LBO regime. There are a few caveats with the NRMS metric in Eq. (3). First of all, the NRMS mapping, including the threshold for classifying the proximity to LBO, needs to be updated for different operating conditions even for the same combustor. The value of α and the ϕ/ϕ_{LBO} limit till which NRMS remains approximately constant is very likely to change with a variation in experimental conditions. Moreover, for computing NRMS, because of the irregularity due to the chaotic nature of the signal near LBO, a large value of time window length N is preferred (at least $N = 10000$ with sampling rate = 5 kHz as reported by Yi and Gutmark (2007b)) for robust performance.

Recently, Unni and Sujith (2016) have adopted a methodology of Recurrence Quantification Analysis (RQA) (Iwanski and Bradley 1998) without phase space reconstruction as a metric that is capable of predicting incipient blowout in combustion systems via capturing the dynamic transitions in the system. Though Unni and Sujith (2016) have applied their metric on pressure time series data, the applicability of the method has been tested on chemiluminescence time series in this work. According to Equation 4, recurrence occurs at the point (t_i, t_j) in the recurrence plot, if $X(t_j)$ (the value of the time series at time t_j) is within the range $(X(t_i) - \epsilon, X(t_i) + \epsilon)$. Θ denotes the Heaviside step function and ϵ indicates the upper limit of the distance between two recurrent observations.

$$R_{ij} = \Theta\left(\epsilon - \left|X_{t_i} - X_{t_j}\right|\right) \qquad i,j = 1, 2, N \tag{4}$$

The threshold ϵ is considered here as the 10% of the standard deviation of the chemiluminescence fluctuations observed at the stoichiometric condition. The strategy of considering such constant threshold is borrowed from the work of Unni and Sujith (2016) where 40% of the pressure oscillations during limit cycle oscillations was taken as the threshold for the construction of recurrence matrix (R_{ij}) and measurement of RQA parameters. To compare the performance of our LLR-based method with that work, recurrence rate (RR) is considered, which is defined as:

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{ij} \tag{5}$$

where, $N = N_s - \tau F_s(m - 1)$ is the total number of phase space vectors in reconstructed space. N_s is the total number of time steps acquired from the chemiluminescence time series
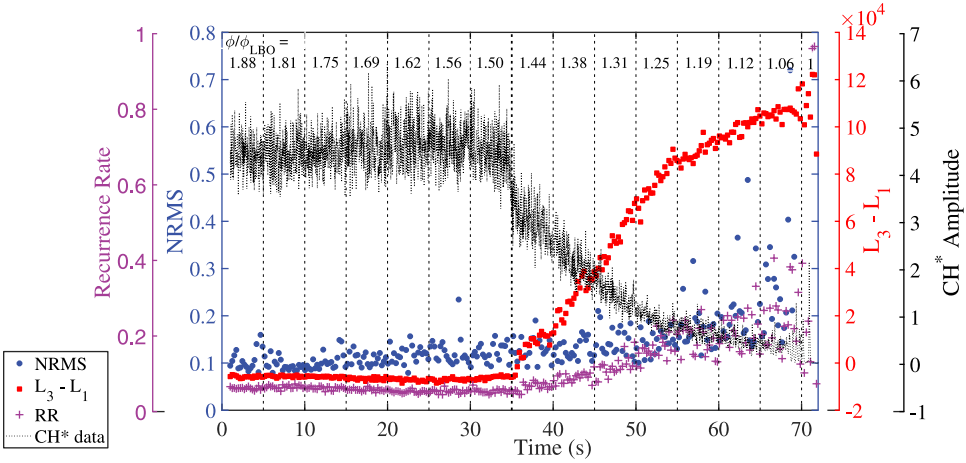
Figure 8. Comparison of normalized root mean square (NRMS) and recurrence rate (RR) with $L_3 - L_1$ norm with a scanning window-size of 1000 samples. The left vertical axes in blue and magenta indicate the ranges for NRMS and RR, respectively. The right vertical axes in red and black indicate the ranges for $L_3 - L_1$ and chemiluminescence amplitude, respectively.

obtained experimentally. τ is the optimum time delay obtained using autocorrelation function (Abarbanel et al. 1993), F_s is the sampling frequency, and m is the minimum embedding dimension (Kennel, Brown, Abarbanel 1992). In the current analysis, m is considered as 1 (i.e., one dimensional system) following the work done by Unni and Sujith (2016) as the process had been reported as effective for exploring the dynamic transitions of a turbulent combustor, although the computational information was not provided there.

The LLR-based methodology has been compared with NRMS and Recurrence Rate (RR) in this section for a comprehensive comparison between the two. Figure 8 shows a comparison between the three metrics. The metrics have been calculated with non-overlapping scanning windows of 1000 samples each. It can be clearly seen that even with a window-length of 1000 samples, the change in $[L_3 - L_1]$ is significantly higher than the corresponding changes in NRMS or RR across the transition from $\phi/\phi_{LBO} = 1.50$ to 1.44, where the chemiluminescence signal indicates a shift from the stable to the transition regime leading to LBO. In fact, while $[L_3 - L_1]$ changes from around -0.5×10^4 to 1×10^4, NRMS shows a noisy behavior between $0.1 - 0.2$ across the regime transition point. The behavior of RR is less noisy as compared to NRMS across the changepoint. This indicates that the LLR metric is more sensitive to regime change as compared to NRMS or RR, which is crucial for maximizing the chances of detecting near LBO transition in an online condition monitoring setting. Moreover, the LLR "threshold" to be chosen for $L_3 - L_1$ is equal to 0, as expected intuitively, since $L_3 - L_1 > 0$ indicates that the data window under consideration is most likely to be belonging to the near blowout phase, as compared to the stable phase. This allows for a decision-making algorithm that is free from the need of thresholding, regardless of the operating conditions/type of combustor chosen for analysis, where a positive log-likelihood ratio between the near-LBO and stable HMMs for a given data window should always be the condition to monitor for signaling the near LBO alerts to the control system.

The NRMS-based detection methodology as proposed in Equation 3 was not directly applicable in our case, since the ϕ/ϕ_{LBO} value till which NRMS remains almost constant was difficult to ascertain, along with the value of α. A modification of Eq. (3) has been made for LBO prediction based on a binary classification problem (Poor 1988), where the hypothesis H_0 corresponds to the stable regime, and H_1 corresponds to the transient combined with near LBO regime:

$$NRMS(\phi/\phi_{LBO}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{6}$$

for a specified threshold τ. A Similar idea of thresholding is needed for RR:

$$RR \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{7}$$

This τ is expected to vary from one experiment to another, and hence needs to be chosen on a case-by-case basis for online prognosis purposes. A commonly used criterion to choose τ is the receiver operating characteristic (ROC) curve that is obtained by varying τ as a trade-off between the probability of successful detection ($p_D \overset{\Delta}{=} p[\text{Selected Class} = H_1 | \text{True Class} = H_1]$) and the probability of false alarms ($p_F \overset{\Delta}{=} p[\text{Selected Class} = H_1 | \text{True Class} = H_0]$). Defining the threshold in the above manner provides more flexibility for the user, whereby an informed choice can be made on the value of the threshold on the basis of some predefined notion of the allowable misdetection and false alarm rates for the particular application.

Window sizes ranging from 200 to 1000 samples have been tested for comparing RR and NRMS-based detection performance with that of LLR. One commonly used method for comparing the performance of different classifiers is by comparing the area under the curve (AUC) for each classifier (Fawcett 2006). The classifier that has a higher AUC is generally considered to be a better classifier on average. It is seen in Table 2 that the AUC in the ROC curves for NRMS and RR consistently less than that of LLR. Moreover, the AUC for LLR with a window size of 200 is almost close to 1, which is not achieved even with a larger window size for the other two metrics. This indicates that using NRMS or RR as a metric for detecting LBO is less data-efficient than using the proposed LLR-based classifier.

Along with AUC, another metric that can be useful for the practical application of ROC curves is the optimal operating point (OOP) of the curve. It is noted that OOP is the point (p_F, p_D) on the ROC curve, where a line drawn from the top left corner (i.e., $p_F = 0, p_D = 1$) meets the curve with a slope:

$$S \overset{\Delta}{=} \left[\frac{Cost(H_1|H_0) - Cost(H_0|H_0)}{Cost(H_0|H_1) - Cost(H_1|H_1)} \right] \frac{|H_0|}{|H_1|} \tag{8}$$

Table 2. Comparison of AUC of ROC curve as a function of window size for the three classifiers.

Window size	LLR	RR	NRMS
200	0.999	0.978	0.615
400	0.999	0.971	0.729
600	1.000	0.975	0.796
800	0.995	0.967	0.832
1000	0.993	0.967	0.858

Table 3. Comparison of OOP (p_F, p_D) of ROC curve as a function of window size for the three classifiers.

Window size	LLR	RR	NRMS
200	(0.002,0.997)	(0.019,0.9168)	(0.237,0.422)
400	(0.002, 1.000)	(0.011, 0.927)	(0.178, 0.552)
600	(0.001, 0.996)	(0.013, 0.913)	(0.218, 0.671)
800	(0.005, 1.000)	(0.011, 0.931)	(0.167, 0.700)
1000	(0.003, 1.000)	(0.007, 0.913)	(0.111, 0.670)

where $|H_k|$ indicates the number of instances of H_k, $k = 1$ or 2, in the data; $Cost(H_1|H_0)$ indicates the penalty associated with a false alarm; $Cost(H_0|H_1)$ indicates the penalty associated with a misdetection; and $Cost(H_1|H_1)$ and $Cost(H_0|H_0)$ indicate the cost associated with correctly classifying the members belonging to the respective hypothesis, respectively. In this analysis, equal penalties have been associated with both false alarms and misdetections, while the cost of each correct classification is set to 0. These penalty weights are user-defined, and can be changed based on the requirements of the problem at hand (Mondal et al. 2017a). OOP is one of the different ways of choosing the best point to operate the ROC, which indicates better performance if p_F is low and p_D is high at that point. Table 3 lists the comparison of OOPs of ROC curves as a function of window sizes for the three metrics. The OOPs are listed as (p_F, p_D) pairs, for example, the OOP for LLR with a window size of 200 has a false alarm probability of 0.2% and a true detection probability of 99.7%. It is seen that OOPs of the ROC curves for LLR always have lower p_F and higher p_D than the corresponding ROC curves for NRMS and RR, with every window size.
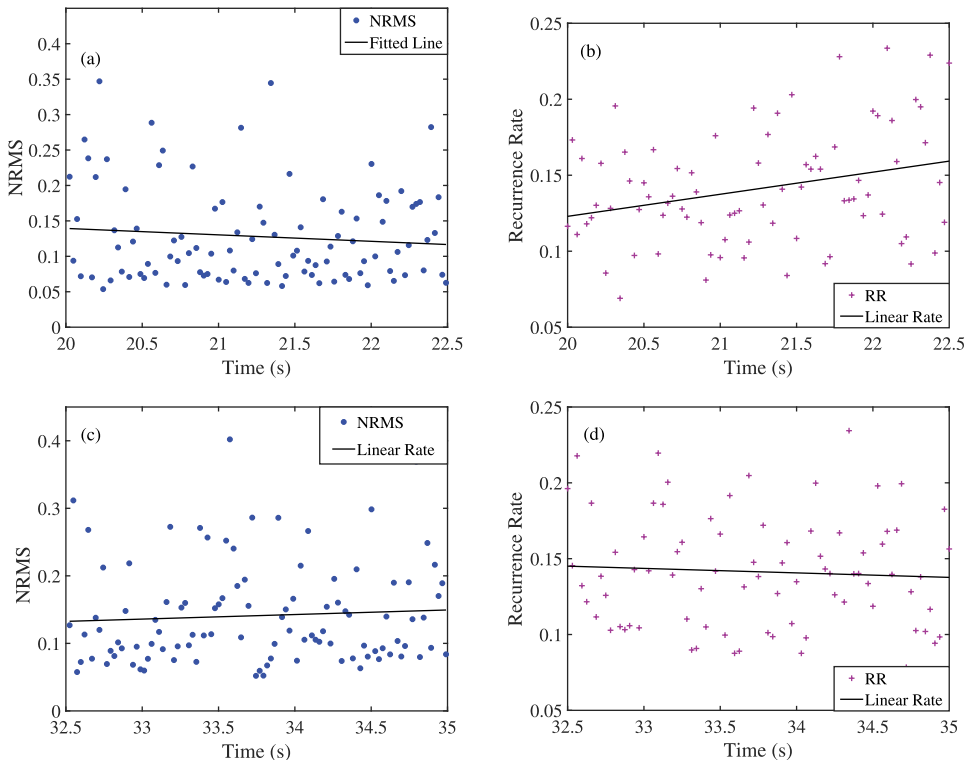
Figure 9. Metrics of variations of NRMS and RR in Timeframes A and B. Linear rates of increase/decrease of the metrics are estimated in the timeframes.

Between RR and NRMS, the comparison of the performance on the basis of AUC and OOP of the ROC curves shows that RR has an improved performance as compared to NRMS for all the window sizes explored, which can be attributed to the fact that RR is a more complicated analysis tool that takes into account nonlinear characteristics to understand the recurrent behavior of the state of a dynamical system. But the stochastic sequential behavior of the system is captured better in the probabilistic modeling architecture of HMM, which renders its performance to be better than RR or NRMS.

The regime detection capability of LLR with respect to two different transients as discussed in Section IVB has also been compared with those of RR and NRMS. Figure 9 shows the variations of NRMS and RR in the Timeframes A and B as shown in Figure 5. The 95% confidence intervals for the slopes of the variations of NRMS and RR in Timeframe A are -0.009 ± 0.0183 and 0.01453 ± 0.0109 respectively, while that in Timeframe B are 0.006759 ± 0.0202 and -0.0029 ± 0.0112, respectively. To compare the results with the estimated slopes using LLR as shown in Figure 6, the coefficient of variation (CV) $\triangleq \frac{\sigma}{|\mu|}$ of the slope parameter estimate has been computed for each of the metrics. CV is a measure of dispersion of the data relative to the mean, and can be used to compare the degrees of variation across datasets (Lovie 2005). The CV_{LLR} (CV using LLR metric) for Timeframe A is 0.2, while that for Timeframe B is 0.15, as shown in Figure 6. On the other hand, CV_{NRMS} for Timeframes A and B are 1.02 and 1.49, respectively, while that of CV_{RR} are 0.375 and 1.87, respectively. So, not only are the mean estimated slope values for LLR significantly greater than those of NRMS and RR (which is an indication of high sensitivity of LLR), the CV for LLR is also lower than that of NRMS and RR for each of Timeframe A and Timeframe B, which signifies that the slope estimation with LLR is more precise, and hence can be trusted more for online regime detection purposes.

Along with the performance metrics, it is necessary to compare the computational complexity of the three classifiers, since they are potential competitors as data-driven metrics for detecting incipient LBO in combustion systems. Table 4 lists the average online processing times of the three classifiers as a function of window sizes. The comparison has been carried out in a MATLAB-2015 environment running on Intel Xeon CPU E5-2620 v3 @ 2.40 GHz (24 cores) with a system memory of 256 GB.

It can be seen that NRMS, being the simplest in construction among the three classifiers, has the least time complexity, with it being almost constant over the studied time window range. Since the calculation of log-likelihood based on *Forward Procedure* has a time complexity to the order of $(Q^2 T)$, the computational time of LLR increases almost linearly with window size (since Q remains constant for a given HMM). The computational time of RR, on the other hand, has a higher order dependence on the window size as compared to the almost linear dependence of LLR, which is observed from the fact that the computational cost of RR

Table 4. Comparison of computational time (s) as a function of window sizes for three classifiers.

Window size	LLR	RR	NRMS
200	0.019	0.007	0.011
400	0.039	0.024	0.014
600	0.067	0.059	0.012
800	0.079	0.125	0.013
1000	0.088	0.132	0.013

is significantly higher than LLR's cost at the higher window sizes of 800 and 1000 samples. Thus HMMs, despite being more complex in its non-deterministic formulation as a stochastic model representing the sequential behavior of temporal data, has a competing computational cost at low window sizes as compared with RR, although the classification performance is superior to that of RR and NRMS. Moreover, a focal area of this paper is to devise a data-driven technique of LBO detection which performs well with very short data lengths and is free from empirical thresholding needs, all of which is satisfied with the HMM-based LLR metric. Moreover, the recent work of (Mondal et al. 2019) has shown the applicability of LLR in detecting the onset of thermoacoustic instabilities (TAI) in combustion systems with short time windows. TAI involves very different physics as compared to LBO, and the temporal behavior of the time series data involved in TAI is characteristically different than that of LBO. This shows the generalizability of the LLR metric in detecting different types of instabilities in combustion systems. Hence, the performance robustness of LLR, along with its competitive computational time complexity, make it a potentially suitable metric for incipient LBO detection and regime classification, which can significantly aid in near-real time detection and control of LBO in combustion systems.

Summary, conclusions, and future work

This paper has developed a methodology for early prediction of lean blowout (LBO) phenomena in combustors of aircraft and land-based gas-turbine systems in the framework of hidden Markov modeling (HMM) and hypotheses testing, which can be used in the setting of supervised learning to discern near-blowout time series data from stable data. The paper also addresses online classification of two different kinds of transient regimes in the combustion systems, which are: 1) transition from stable mode of operation toward LBO; and 2) transition from near LBO to being stable. In order to experimentally validate the underlying theory, pertinent conditions of LBO have been emulated on a laboratory-scale apparatus of swirl-stabilized dump combustor with transient time series of CH* chemiluminescence data. Being data-driven, the proposed methodology is model-free; it has been shown to be numerically efficient as well as sensitive to regime changes when the combustion system moves toward or away from LBO.

While there are many areas of theoretical and experimental research to enhance the work reported in this paper, the authors suggest the following topics of future research for early prediction of lean blowout (LBO) phenomena and classification of operational regimes in combustion systems:

(1) Usage of statistical modeling tools based on symbolic dynamics for online detection of LBO (Mukherjee and Ray 2014; Ray 2004). of the LBO detection framework for liquid fuel-based combustion systems.
(2) Development of analytical tools for prediction of precursors to LBO from changes in the flow patterns via: (i) synergistic combinations of data-driven and physics-based modeling; and (ii) high-fidelity measurements like high-speed imaging/advanced diagnostics.
(3) Experimental validation on laboratory apparatuses for different types of multi-physics experiments.

Nomenclature

λ	Hidden Markov model
Θ	Heaviside Step Function
θ	Rate of increase of anomaly metric
ϕ	Equivalence ratio
μ	Mean of estimated slope parameter
σ	Standard deviation of estimated slope parameter
τ	Classification threshold for binary hypothesis problem
LIST OF ACRONYMS	
AFR	Air flow rate
AUC	Area under the Curve
CC	Combustion chamber
CV	Coefficient of variation $\left(\frac{\sigma}{|\mu|}\right)$
ES	Exhaust chamber
FFR	Fuel flow rate
HMM	Hidden Markov model
LBO	Lean blow-out
LLR	Log likelihood ratio
MFC	Mass flow controller
NRMS	Normalized root mean square
OOP	Optimal operating point
PC	Premixing chamber
PMT	Photo multiplier tube
RMS	Root mean square
ROC	Receiver Operating Characteristics
RP	Recurrence plots
RQA	Recurrence quantification analysis
RR	Recurrence rate
SLPM	Standard liters per minute
SN	Swirl number
SS	Stainless steel
STSA	Symbolic time series analysis
XWT	Cross wavelet transform

References

Abarbanel, H. D. I., R. Brown, J. J. Sidorowich, and L. S. Tsimring. 1993. The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys* 65:1331–92. https://link.aps.org/doi/10.1103/RevModPhys.65.1331.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19 (6):716–23. doi:10.1109/TAC.1974.1100705.

Barnum, B., and R. C. Bell. 1993. Flame failure detection method. *Google Pat.* US5235802A.

Bishop, C. 2006. *Pattern recognition and machine learning.* Springer: New York, NY, USA.

Cavaliere, D. E., J. Kariuki, and E. Mastorakos. 2013. A comparison of the blow-off behaviour of swirl-stabilized premixed, non-premixed and spray flames. *Flow Turbul. Combust* 91 (2):347–72. doi:10.1007/s10494-013-9470-z.

Celeux, G., and J.-B. Durand. 2008. Selecting hidden markov model state number with cross-validated likelihood. *Comput. Stat* 23 (4):541–64. doi:10.1007/s00180-007-0097-1.

Chao, Y.-C., Y.-L. Chang, C.-Y. Wu, and T.-S. Cheng. 2000. An experimental investigation of the blowout process of a jet flame. *Proc. Combust. Inst.* 28 (1):335–42. http://www.sciencedirect.com/science/article/pii/S0082078400802283.

Chaudhari, R. R., R. P. Sahu, S. Ghosh, A. Mukhopadhyay, and S. Sen. 2013. Flame color as a lean blowout predictor. *Int. J. Spray Combust. Dyn.* 5 (1):49–65. doi:10.1260/1756-8315.5.1.49.

Chen, J., T.-Y. Hsu, -C.-C. Chen, and Y.-C. Cheng. 2011. Online predictive monitoring using dynamic imaging of furnaces with the combinational method of multiway principal component analysis and hidden markov model. *Ind Eng Chem Res* 50 (5):2946–58. doi:10.1021/ie100671j.

Correa, S. M. 1993. A review of NOx formation under gas-turbine combustion conditions. *Combust. Sci. Technol* 87 (1–6):329–62. doi:10.1080/00102209208947221.

De, S., A. Bhattacharya, S. Mondal, A. Mukhopadhyay, and S. Sen. 2019. Investigation of flame behavior and dynamics prior to lean blowout in a combustor with varying mixedness of reactants for the early detection of lean blowout. *Int. J. Spray Combust. Dyn.* 11:1756827718812519. doi:10.1177/1756827718812519.

De, S., A. Bhattacharya, S. Mondal, A. Mukhopadhyay, and S. Sen. 2020a. Application of recurrence quantification analysis for early detection of lean blowout in a swirl-stabilized dump combustor. *CHAOEH* 30 (4):043115. doi:10.1063/1.5131231.

De, S., A. Biswas, A. Bhattacharya, A. Mukhopadhyay, and S. Sen. 2020b. Use of flame color and chemiluminescence for early detection of lean blowout in gas turbine combustors at different levels of fuel–air premixing. *Combust. Sci. Technol* 192 (5):933–57. doi:10.1080/00102202.2019.1604514.

Dey, D., R. R. Chaudhari, A. Mukhopadhyay, S. Sen, and S. Chakravorti. 2015. A cross-wavelet transform aided rule based approach for early prediction of lean blow-out in swirl-stabilized dump combustor. *Int. J. Spray Combust. Dyn.* 7 (1):69–90. doi:10.1260/1756-8277.7.1.69.

Domen, S., H. Gotoda, T. Kuriyama, Y. Okuno, and S. Tachibana. 2015. Detection and prevention of blowout in a lean premixed gas-turbine model combustor using the concept of dynamical system theory. *Proc. Combust. Inst.* 35 (3):3245–53. http://www.sciencedirect.com/science/article/pii/S1540748914003241.

Esclapez, L., P. C. Ma, E. Mayhew, R. Xu, S. Stouffer, T. Lee, H. Wang, and M. Ihme. 2017a. Fuel effects on lean blow-out in a realistic gas turbine combustor. *Combust. Flame* 181:82–99. doi:10.1016/j.combustflame.2017.02.035. http://www.sciencedirect.com/science/article/pii/S0010218017300822.

Esclapez, L., P. C. Ma, E. Mayhew, R. Xu, S. Stouffer, T. Lee, H. Wang, and M. Ihme (2017b). Large-eddy simulations of fuel effect on gas turbine lean blow-out. In *55th AIAA Aerospace Sciences Meeting*, Grapevine, Texas, USA, (p. 1955).

Fawcett, T. 2006. An introduction to ROC analysis. *Pattern Recognit Lett* 27 (8):861–74. doi:10.1016/j.patrec.2005.10.010.

(Accessed on June 20, 2020). Gas India technical specifications. http://www.gasindia.in/technical-specification.html/. (Feb, 2008). *FRCC system disturbance and under frequency load shedding event report.* FRCC. (June, 2008). *Turbine combustor lean blowout.* NERC Industry Advisory.

Gotoda, H., M. Amano, T. Miyano, T. Ikawa, K. Maki, and S. Tachibana. 2012a. Characterization of complexities in combustion instability in a lean premixed gas-turbine model combustor. *Chaos* 22 (4):043128. doi:10.1063/1.4766589.

Gotoda, H., K. Hayashi, R. Tsujimoto, S. Domen, and S. Tachibana. 2016. Dynamical properties of combustion instability in a laboratory-scale gas-turbine model combustor. *J. Eng. Gas Turbines Power* 139 (4):041509. doi:10.1115/1.4034700.

Gotoda, H., T. Ikawa, K. Maki, and T. Miyano. 2012b. Short-term prediction of dynamical behavior of flame front instability induced by radiative heat loss. *CHAOEH* 22 (3):033106. doi:10.1063/1.4731267.

Gotoda, H., Y. Shinoda, M. Kobayashi, Y. Okuno, and S. Tachibana. 2014. Detection and control of combustion instability based on the concept of dynamical system theory. *Phys. Rev. E* 89:022910. https://link.aps.org/doi/10.1103/PhysRevE.89.022910.

Hajek, B. 2015. *Random processes for engineers*. 1st ed. Cambridge, UK: Cambridge University Press.

Hardalupas, Y., and M. Orain. 2004. Local measurements of the time-dependent heat release rate and equivalence ratio using chemiluminescent emission from a flame. *Combust. Flame* 139 (3):188–207. http://www.sciencedirect.com/science/article/pii/S001021800400152X.

He, L., Q. Guo, Y. Gong, F. Wang, and G. Yu. 2019. Investigation of OH⋆ chemiluminescence and heat release in laminar methane–oxygen co-flow diffusion flames. *Combust. Flame* 201:12–22. doi:10.1016/j.combustflame.2018.12.009. http://www.sciencedirect.com/science/article/pii/S0010218018305261.

Herna´ndez, R., and J. Ballester. 2008. Flame imaging as a diagnostic tool for industrial combustion. *Combust. Flame* 155 (3):509–28. http://www.sciencedirect.com/science/article/pii/S0010218008001983.

Hertzberg, J., I. Shepherd, and L. Talbot. 1991. Vortex shedding behind rod stabilized flames. *Combust. Flame* 86 (1):1–11. http://www.sciencedirect.com/science/article/pii/001021809190051C.

Iwanski, J. S., and E. Bradley. 1998. Recurrence plots of experimental data: To embed or not to embed? *CHAOEH* 8 (4):861–71. doi:10.1063/1.166372.

Kabiraj, L., A. Saurabh, N. Karimi, A. Sailor, E. Mastorakos, A. P. Dowling, and C. O. Paschereit. 2015. Chaos in an imperfectly premixed model combustor. *CHAOEH* 25 (2):023101. doi:10.1063/1.4906943.

Kabiraj, L., A. Saurabh, P. Wahi, and R. I. Sujith. 2012. Route to chaos for combustion instability in ducted laminar premixed flames. *CHAOEH* 22 (2):023129. doi:10.1063/1.4718725.

Kaluri, A., P. Malte, and I. Novosselov. 2018. Real-time prediction of lean blowout using chemical reactor network. *Fuel* 234:797–808. doi:10.1016/j.fuel.2018.07.065. http://www.sciencedirect.com/science/article/pii/S0016236118312638.

Keller, J. O., and K. Saito. 1987. Measurements of the combusting flow in a pulse combustor. *Combust. Sci. Technol* 53 (2–3):137–63. doi:10.1080/00102208708947024.

Kennel, M. B., R. Brown, and H. D. I. Abarbanel. 1992. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* 45:3403–11. https://link.aps.org/doi/10.1103/PhysRevA.45.3403.

Lawn, C. 2000. Distributions of instantaneous heat release by the cross-correlation of chemiluminescent emissions. *Combust. Flame* 123 (1):227–40. http://www.sciencedirect.com/science/article/pii/S0010218000001292.

Li, H., X. Zhou, J. B. Jeffries, and R. K. Hanson. 2007. Active control of lean blowout in a swirl-stabilized combustor using a tunable diode laser. *Proc. Combust. Inst.* 31 (4):3215–23. http://www.sciencedirect.com/science/article/pii/S1540748906000071.

Liu, Y., J. Tan, H. Wang, and L. Lv. 2019. Characterization of heat release rate by OH⋆ and CH⋆ chemiluminescence. *Acta Astronaut.* 154:44–51. doi:10.1016/j.actaastro.2018.10.022. http://www.sciencedirect.com/science/article/pii/S0094576518314334.

Lovie, P. 2005. Coefficient of Variation. *Am Cancer Soc.* https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa107.

Meegahapola, L., and D. Flynn. 2015. Characterization of gas turbine lean blowout during frequency excursions in power networks. *IEEE Trans. Power Syst.* 30 (4):1877–87. doi:10.1109/TPWRS.2014.2356336.

Menon, S., O. Uluyol, K. Kim, and E. O. Nwadiogbu (2003). Incipient fault detection and diagnosis in turbine engines using hidden markov models.

Mondal, S., C. Bhattacharya, P. Chattopadhyay, A. Mukhopadhyay, and A. Ray (2017a). Prediction of thermoacoustic instabilities in a premixed combustor based on FFT-based dynamic characterization. In *53rd AIAA/SAE/ASEE Joint Propulsion Conference*. https://arc.aiaa.org/doi/abs/10.2514/6.2017–4892

Mondal, S., N. F. Ghalyan, A. Ray, and A. Mukhopadhyay. 2019. Early detection of thermoacoustic instabilities using Hidden Markov Models. *Combust. Sci. Technol* 191 (8):1309–36. doi:10.1080/00102202.2018.1523900.

Mondal, S., S. A. Pawar, and R. I. Sujith. 2017b. Synchronous behaviour of two interacting oscillatory systems undergoing quasiperiodic route to chaos. *CHAOEH* 27 (10):103119. doi:10.1063/1.4991744.

Mukherjee, K., and A. Ray. 2014. State splitting and merging in probabilistic finite state automata for signal representation and analysis. *Signal Process* 104:105–19. doi:10.1016/j.sigpro.2014.03.045.

Mukhopadhyay, A., R. R. Chaudhari, T. Paul, S. Sen, and A. Ray. 2013. Lean blow-out prediction in gas turbine combustors using symbolic time series analysis. *J. Propul. Power* 29 (4):950–60. doi:10.2514/1.B34711.

Murphy, K. 2012. *Machine learning: A probabilistic perspective*. 1st ed. Cambridge, MA, USA: The MIT Press.

Muruganandam, T., S. Nair, Y. Neumeier, T. Lieuwen, and J. Seitzman. 2002. Optical and Acoustic Sensing of Lean Blowout Precursors. In 38th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, Indianapolis, Indiana, USA.

Muruganandam, T. M., S. Nair, D. Scarborough, Y. Neumeier, J. Jagoda, T. Lieuwen, J. Seitzman, and B. Zinn. 2005. Active control of lean blowout for turbine engine combustors. *J. Propul. Power* 21 (5):807–14. doi:10.2514/1.7254.

Nair, S., and T. Lieuwen. 2005. Acoustic detection of blowout in premixed flames. *J. Propul. Power* 21 (1):32–39. doi:10.2514/1.5658.

Nair, V., and R. I. Sujith. 2014. Multifractality in combustion noise: Predicting an impending combustion instability. *J Fluid Mech* 747:635–55. doi:10.1017/jfm.2014.171.

Nair, V., G. Thampi, and R. I. Sujith. 2014. Intermittency route to thermoacoustic instability in turbulent combustors. *J Fluid Mech* 756:470–87. doi:10.1017/jfm.2014.468.

Nicholson, H., and J. Field. 1948. Some experimental techniques for the investigation of the mechanism of flame stabilization in the wakes of bluff bodies. *Symp. Combust. Flame Explosion Phenom.* 3 (1):44–68. http://www.sciencedirect.com/science/article/pii/S1062289649800080.

Peiffer, E. E., J. S. Heyne, and M. Colket. 2019. Sustainable aviation fuels approval streamlining: Auxiliary power unit lean blowout testing. *Aiaa J.* 57 (11):4854–62. doi:10.2514/1.J058348.

Plee, S., and A. Mellor. 1979. Characteristic time correlation for lean blowoff of bluff-body-stabilized flames. *Combust. Flame* 35:61–80. doi:10.1016/0010-2180(79)90007-5. http://www.sciencedirect.com/science/article/pii/0010218079900075.

Poor, H. V. 1988. *An introduction to signal detection and estimation*. New York, NY, USA: Springer-Verlag.

Price, R., I. Hurle, and T. Sugden. 1969. Optical studies of the generation of noise in turbulent flames. *Symp. (Int.) Combust.* 12 (1):1093–102. http://www.sciencedirect.com/science/article/pii/S008207846980487X.

Rabiner, L., and B.-H. Juang. 1993. *Fundamentals of speech recognition*. Upper Saddle River, NJ, USA: Prentice- Hall, Inc.

Rabiner, L. R. 1989. A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE* 77 (2):257–86. doi:10.1109/5.18626.

Radhakrishnan, K., and J. B. Heywood. 1980. Effects of combustor inlet conditions on flame stability. *Combust. Sci. Technol* 24 (5–6):165–78. doi:10.1080/00102208008952435.

Radhakrishnan, K., J. B. Heywood, and R. J. Tabaczynski. 1981. Premixed turbulent flame blowoff velocity correlation based on coherent structures in turbulent flows. *Combust. Flame* 42:19–33. doi:10.1016/0010-2180(81)90139-5. http://www.sciencedirect.com/science/article/pii/0010218081901395.

Ray, A. 2004. Symbolic dynamic analysis of complex systems for anomaly detection. *Signal Process* 84 (7):1115–30. doi:10.1016/j.sigpro.2004.03.011.

Rock, N., B. Emerson, J. Seitzman, and T. Lieuwen. 2020. Near-lean blowoff dynamics in a liquid fueled combustor. *Combust. Flame* 212:53–66. doi:10.1016/j.combustflame.2019.10.010. http://www.sciencedirect.com/science/article/pii/S0010218019304663.

Rosfjord, T. J., and J. M. Cohen. 1995. Evaluation of the transient operation of advanced gas turbine combustors. *J. Propul. Power* 11 (3):497–504. doi:10.2514/3.23870.

Ryde´n, T. 1995. Estimating the order of hidden markov models. *Statistics* 26 (4):345–54. doi:10.1080/02331889508802501.

Sarkar, S., A. Ray, A. Mukhopadhyay, R. R. Chaudhari, and S. Sen (2014). Early detection of lean blow out (LBO) via generalized D-markov machine construction. In *2014 American Control Conference*, Portland, Oregon, USA, (pp. 3041–3046).

Savas, O., and S. R. Gollahalli. 1986. Stability of lifted laminar round gas-jet flame. *J Fluid Mech* 165:297–318. doi:10.1017/S0022112086003105.

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Statist* 6 (2):461–64. doi:10.1214/aos/1176344136.

Sen, U., T. Gangopadhyay, C. Bhattacharya, A. Mukhopadhyay, and S. Sen. 2018. Dynamic characterization of a ducted inverse diffusion flame using recurrence analysis. *Combust. Sci. Technol* 190 (1):32–56. doi:10.1080/00102202.2017.1374952.

Sun, D., G. Lu, H. Zhou, and Y. Yan. 2013. Condition monitoring of combustion processes through flame imaging and kernel principal component analysis. *Combust. Sci. Technol* 185 (9):1400–13. doi:10.1080/00102202.2013.798316.

Unni, V. R., and R. I. Sujith (2016). Precursors to blowout in a turbulent combustor based on recurrence quantification. In *Proceedings of 52nd AIAA/SAE/ASEE Joint Propulsion Conference*. https://arc.aiaa.org/doi/abs/10.2514/6.2016-4649

Yi, T., and E. Gutmark (2007a). Effects of chemical kinetics and heat loss on near-LBO combustion dynamics - stability analysis. In *45th AIAA Aerospace Sciences Meeting and Exhibit*. https://arc.aiaa.org/doi/abs/10.2514/6.2007-388

Yi, T., and E. J. Gutmark. 2007b. Real-time prediction of incipient lean blowout in gas turbine combustors. *Aiaa J.* 45 (7):1734–39. doi:10.2514/1.25847.

Zheng, L., I. Ahmed, E. Ubogu, Y. Zhang, and B. Khandelwal (2017). Evaluation of the lean blowout limit of alternative fuels in a gas turbine combustor. In *Proceedings of the 23rd International Society for Air Breathing Engines (ISABE) Conference 2017*, Manchester, UK.

Zheng, L., J. Cronly, E. Ubogu, I. Ahmed, Y. Zhang, and B. Khandelwal. 2019. Experimental investigation on alternative fuel combustion performance using a gas turbine combustor. *Appl. Energy* 238:1530–42. doi:10.1016/j.apenergy.2019.01.175. http://www.sciencedirect.com/science/article/pii/S0306261919301448.

Zhong, S., and J. Ghosh (2002). HMMs and coupled HMMs for multi-channel EEG classification. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, Honolulu, Hawaii, USA, vol. 2, (pp. 1154–59).

Appendix A
Mathematical formulation: hidden markov modeling

While the details on hidden Markov modeling (HMM) are available in standard literature (Bishop 2006; Hajek 2015; Murphy 2012; Rabiner 1989; Rabiner and Juang 1993), this appendix summarizes the essential concepts that form the backbone for analysis of stochastic dynamical behavior of LBO in this paper. The applications of HMM in a different context of detecting thermoacoustic instabilities in combustion systems have been demonstrated by the authors in a recent work (Mondal et al. 2019). This paper extends the approach demonstrated in the work of (Mondal et al. 2019) for early detection of LBO. In general, HMMs are a class of stochastic processes, which involves a latent *hidden* state evolution. The hidden states are not directly observed, but can be inferred by drawing observations from another stochastic process that produces the sequence of observations (Rabiner and Juang 1993).
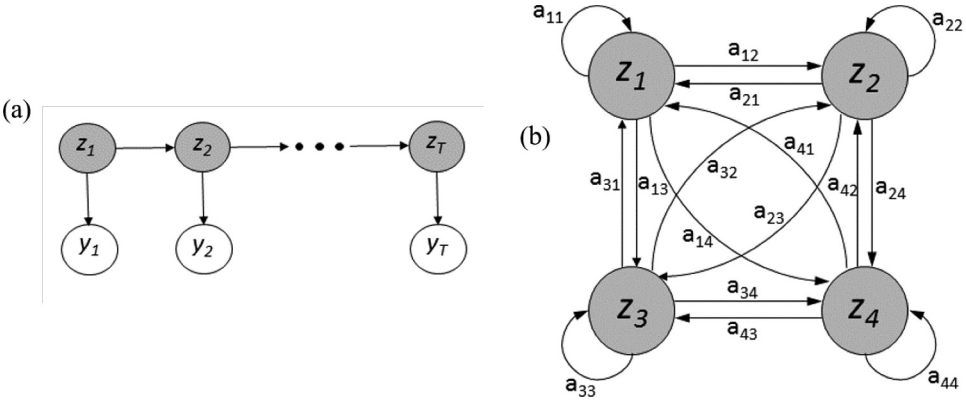
Figure 10. (a) Bayesian network model of an HMM showing conditional dependence of the observations with the hidden Markovian states. (b) State transitions in a 4 state HMM (Mondal et al. 2019).

HMMs differ from standard Markov models in the formulation that does not assume Markov property for the observations themselves.

Let a sequence of T discrete-time observations be represented as $\mathbf{Y} = \{y_1, y_2, \ldots, y_T\}$. Under the first-order Markov assumption, the joint probability of this observation sequence is formulated as: $p(\mathbf{Y}) = p(y_1) \prod_{t=1}^{T-1} p(y_{t+1}|y_t)$; and also the sequence $\mathbf{Z} = \{z_1, z_2, \ldots, z_T\}$ of hidden state vectors are assumed to follow Markov properties, i.e., given the state z_{t-1} at time $(t-1)$, the current state z_t is independent of all the states prior to $(t-1)$. It is noted that the observations $\mathbf{Y} = \{y_1, y_2, \ldots, y_T\}$ are conditionally dependent on the states.

Figure 10 depicts the above concept with a Bayesian network model of an HMM. A shaded node in Figure 10(a) denotes a hidden variable, an unshaded node signifies an observed variable, and an arrow denotes conditional dependence between variables. In Figure 10(b) the arrows denote the transitions from state z_i to z_j with probability a_{ij}, with the states belonging to a finite and discrete set, i.e. $z_t \in \{1, , Q\}$. The observation sequence $\{y_1, \ldots, y_T\}$ is assumed to be generated by a latent state sequence $\{z_1, \ldots, z_T\}$. In this setting, an HMM is represented as a triplet $\lambda = \{A, B, \pi\}$, where:

- $A \stackrel{\Delta}{=} \{a_{ij}\}$ is the state-transition probability: $a_{ij} = P(z_{t+1} = q_j | z_t = q_i)$.
- $B \stackrel{\Delta}{=} \{b_j(y_t)\}$ is the probability distribution of the observation given the state: $b_j(y_t) = P(y_t | z_t = q_j)$.
- $\pi \stackrel{\Delta}{=} \{\pi_i\}$ is the initial state distribution: $\pi_i = P(z_1 = q_i)$ at time $t = 1$.

Associated with the above hmm $\lambda = \{A, B, \pi\}$, the following two estimation tasks are considered in this paper:

1) Given the observed data $\mathbf{Y} = \{y_1, y_2, \ldots, y_T\}$ and the HMM λ, the first task is to compute the conditional distribution of the state, which is solved by the forward algorithm (Hajek 2015; Rabiner 1989).
2) Given the observed data, the second task is to obtain the maximum likelihood (ML) estimate of the HMM λ, which is solved by the backward algorithm and Baum-Welch/Expectation Maximization (EM) algorithm (Hajek 2015; Rabiner 1989).

In the above context, the forward, backward, and Baum-Welch (Expectation Maximization) algorithms are succinctly presented as follows.

A. the forward procedure

The forward variable $\alpha_t(i) \stackrel{\Delta}{=} p(y_1, y_2, \ldots, y_t; z_t = i | \lambda)$ is the probability that the partial observation sequence $\{y_1, y_2, \ldots, y_t\}$, $1 \le t < T$ ending with the state of the system being i at time t. This is solved inductively as follows:

$$1. \text{ Initialization step}: \quad \alpha_1(i) = \pi_i b_i(y_1), \quad 1 \leq i \leq Q \tag{9}$$

$$2. \text{ Induction step}: \quad \alpha_{t+1}(j) = \left[\sum_{i=1}^{Q} \alpha_t(i) a_{ij}\right] b_j(y_{t+1}), \quad 1 \leq t \leq T-1, \quad 1 \leq j \leq Q \tag{10}$$

$$3. \text{ Termination step}: \quad p(\mathbf{Y}|\lambda) = \sum_{i=1}^{Q} \alpha_T(i) \tag{11}$$

where T and Q are the same as defined earlier.

B. the backward procedure

The backward variable $\beta_t(i) \triangleq p(y_{t+1}, y_{t+2}, \ldots, y_T | z_t = i; \lambda)$ is the probability of the partial observation sequence $\{y_{t+1}, y_{t+2}, \ldots, y_T\}$ from time $t+1$ till the end time T and the state at time t is i and the HMM followed is λ. This is solved inductively as follows:

$$1. \text{ Initialization step}: \quad \beta_T(i) = 1, \quad 1 \leq i \leq Q \tag{12}$$

$$2. \text{ Induction step}: \quad \beta_t(i) = \left[\sum_{j=1}^{Q} a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)\right], \quad t = T-1, T-2, \ldots, 1, \quad 1 \leq i \leq Q \tag{13}$$

where the parameters T and Q are the same as defined earlier.

C. model learning: baum-welch algorithm

The model learning problem involves estimation of the parameters $\lambda = \{A, B, \pi\}$ so as to maximize the likelihood $p(\mathbf{Y}|\lambda)$. Baum-Welch algorithm is a procedure for recursive estimation of the HMM parameters. Given the model and the observation sequence, the intermediate variables ξ and γ are defined, for $1 \leq i \leq Q$, $1 \leq j \leq Q$ and $1 \leq t \leq T$, as:

$$\xi_t(i,j) \triangleq P(z_t = q_i, z_{t+1} = q_j | \mathbf{Y}, \lambda); \quad \gamma_t(i) \triangleq p(z_t = q_i | \mathbf{Y}, \lambda) \tag{14}$$

The variables γ_t and $\xi_t(i,j)$ are expressed in terms of the forward and backward variables α_t and β_t, defined earlier as:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{p(\mathbf{Y}|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{K} \alpha_t(i)\beta_t(i)}; \quad \xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_t(i) a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)} \tag{15}$$

Using the above relations, and the fact that $\gamma_t(i) = \sum_{j=1}^{K} \xi_t(i,j)$, it is possible to estimate $\{A, B, \pi\}$ as:

$$\hat{\pi}_i = \gamma_1(i); \quad \hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}; \quad \hat{c}_{jk} = \frac{\sum_{t=1}^{T} \tilde{\gamma}_t(j,k)}{\sum_{t=1}^{T} \sum_{k=1}^{M} \tilde{\gamma}_t(j,k)};$$

$$\hat{\mu}_{jk} = \frac{\sum_{t=1}^{T} \tilde{\gamma}_t(j,k) \cdot y_t}{\sum_{t=1}^{T} \tilde{\gamma}_t(j,k)}; \quad \hat{\Sigma}_{jk} = \frac{\sum_{t=1}^{T} \tilde{\gamma}_t(j,k) \cdot (y_t - \mu_{jk})^2}{\sum_{t=1}^{T} \tilde{\gamma}_t(j,k)} \tag{16}$$

where $\tilde{\gamma}_t(j,k)$ is the probability of being in state j at time t with the kth mixture component. That is,

$$\tilde{\gamma}_t(j,k) = \left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}\right] \left[\frac{c_{jk}\mathcal{N}(y_t, \mu_{jk}, \Sigma_{jk})}{\sum_{k=1}^{M} c_{jk}\mathcal{N}(y_t, \mu_{jk}, \Sigma_{jk})}\right] \tag{17}$$

It is noted that the term $\tilde{\gamma}_t(j,k)$ generalizes to $\gamma_t(j)$ in case of a single-component Gaussian density (i.e., $M = 1$) or a (discrete) probability mass function. The model $\hat{\lambda} = \{\hat{A}, \hat{B}, \hat{\pi}\}$ is recursively estimated as it converges to a local maximum of the likelihood function $p(\mathbf{Y}|\lambda^*)$, where λ^* is the ML estimate of the HMM (Hajek 2015; Rabiner 1989).

D. execution of hmm algorithms

A continuously-varying structure of the observation model has been used in this paper, where the observations $\{y_1, \ldots, y_T\}$ are one-dimensional and their emission probability follows a Gaussian mixture model having a univariate distribution for each of the mixture component:

$$p(y_t|z_t = j, \lambda) = \sum_{\ell=1}^{M} c_{j\ell} \mathcal{N}(y_t; \mu_{j\ell}, \Sigma_{j\ell}) \tag{18}$$

where M is the number of Gaussian mixture components; the weights $c_{j\ell}$ are such that $\sum_{\ell=1}^{M} c_{j\ell} = 1$ $j \in \{1, \cdots, Q\}$; and $\mathcal{N}(y_t; \mu_{j\ell}, \Sigma_{j\ell})$ represents Gaussian density function of y_t with mean $\mu_{j\ell}$ and covariance $\Sigma_{j\ell}$ associated with state j and mixture component ℓ. Although some research work has been reported on optimal identification of the HMM parameters Q and M (e.g., (Celeux and Durand 2008; Ryde´n 1995), Tobias), the Akaike Information Criterion(AIC)/Bayesian Information Criterion(BIC)-based techniques are commonly used to optimally select these model parameters (Akaike 1974; Schwarz 1978); this is accomplished by maximizing the likelihood of the data and minimizing the model complexity, thereby balancing the goodness of the fit against the complexity of the model to prevent over-fitting. The parameters Q and M have been chosen in this paper by following the BIC criterion. Consequently, the model learning problem is to find the optimal set of parameters for $\lambda = \{A, B, \pi\}$ in order to maximize $P(\mathbf{Y}, \mathbf{Z}|\lambda)$ by using an iterative procedure called *Baum-Welch Algorithm*, which is an application of Expectation-Maximization method for inferring HMM parameters.

Following the HMM $\lambda = \{A, B, \pi\}$, the corresponding joint distribution of the states and observations has the form:

$$p(\mathbf{Y}, \mathbf{Z}|\lambda) = p(z_{1:T})p(y_{1:T}|z_{1:T}) = p(z_1) \prod_{t=1}^{T-1} p(z_{t+1}|z_t) \prod_{t=1}^{T} p(y_t|z_t) \tag{19}$$

The calculation of $p(\mathbf{Y}, \mathbf{Z}|\lambda)$ according to Eq. (19) has computational complexity in the order of $(2TQ^T)$, which may become intractable for even small values of Q and T. A numerically efficient method, called the *Forward Procedure* has been used, which reduces computational complexity to the order of $(Q^2 T)$.