Physics-informed data-driven prediction of premixed flame dynamics with data assimilation

By H. Yu†, T. Jaravel, J. Labahn, M. Ihme, M. P. Juniper† AND L. Magri†‡

We propose an on-the-fly statistical learning method to make a qualitative reduced-order model of the dynamics of a premixed flame quantitatively accurate. This physics-informed data-driven method is based on the statistically optimal combination of (i) a reduced-order model of the dynamics of a premixed flame with a level-set method, (ii) high-quality data, which can be provided by experiments and/or high-fidelity simulations, and (iii) assimilation of the data into the reduced-order model to improve the prediction of the dynamics of the premixed flame. The reduced-order model learns the state and the parameters of the premixed flame on the fly with the ensemble Kalman filter, which is a Bayesian filter used in the data assimilation of high-dimensional dynamical systems, e.g. in weather forecasting. The proposed method and algorithm are applied to two test cases with relevance to reacting flow and instability. First, the capabilities of the framework are demonstrated in a twin experiment, where the assimilated data is produced from the same model as that used in prediction. Second, the assimilated data is extracted from a high-fidelity reacting-flow direct numerical simulation (DNS). The results are analyzed by using Bayesian statistics, which provide the uncertainties of the calculations. This method opens up new possibilities for on-the-fly optimal calibration of computationally cheap reduced-order models when experimental data becomes available, for example, from sensors.

1. Introduction

Thermoacoustic instabilities are a persistent challenge in the design of jet and rocket engines. Velocity and pressure oscillations inside the combustion chamber interact with the flame and cause unsteady heat release. If moments of higher heat release coincide with moments of higher pressure (and lower heat release with lower pressure), acoustic oscillations can grow. This can lead to large-amplitude oscillations causing structural damage in the jet or rocket engine (Lieuwen & Yang 2005; Culick 2006).

A simple example of a thermoacoustic system is a heat source inside a duct. Depending on the type of heat source, this thermoacoustic system displays varying degrees of nonlinearity. For an electric heater or a diffusion flame, the nature of the nonlinearity is relatively simple: after a linear growth phase, the acoustic oscillations saturate and form a limit cycle. For a premixed flame, the nature of the nonlinearity becomes more complicated because of the formation of cusps and pinched-off fuel-air pockets. In this case, the thermoacoustic system displays rich nonlinear dynamics, e.g. chaos through period doubling or the Ruelle-Takens-Newhouse route (Kabiraj *et al.* 2012; Kashinath *et al.* 2014). Rich nonlinear dynamics have also been observed in more realistic settings, e.g. a gas-turbine model combustion chamber.

† Department of Engineering, University of Cambridge, UK
‡ Institute for Advanced Study, Technical University of Munich, Germany (visiting fellow)

The time-accurate calculation of thermoacoustic instabilities is challenging for several reasons, in particular:

(*a*) Aleatoric uncertainty. Under realistic conditions, a thermoacoustic system is subject to stochastic noise, which cannot be exactly replicated in a simulation. In the worst-case scenario, stochastic noise may trigger thermoacoustic instabilities before the limit of linear stability is reached.

(*b*) Epistemic uncertainty. Thermoacoustic instabilities involve hydrodynamic, chemical and acoustic effects among others. A simulation usually relies on some modeling assumptions, e.g. simplified governing equations, a (relatively) low spatial resolution or a reduced chemical mechanism. The result is a model that may lack relevant degrees of freedom or have inaccurate parameters.

(*c*) Extreme sensitivity. The long-term behavior of a thermoacoustic system, both qualitative and quantitative, may be highly sensitive to uncertain parameters such as boundary conditions and the operating regime.

We propose to address these challenges by augmenting reduced-order models with data from numerical experiments in the form of high-fidelity simulations using methods based on the theory of stochastic processes, namely data assimilation and parameter estimation based on a Bayesian approach. Data assimilation gives an optimal estimate of the true state of a system given experimental observations. Parameter estimation uses the data to find a maximum-likelihood set of parameters for the model. The theory of stochastic processes (Jazwinski 2007) offers a novel approach, which has not been explored yet in the context of thermoacoustics.

2. Level-set methods and data assimilation

The thermoacoustic system under investigation is a ducted premixed flame. In the past, its nonlinear dynamics have been successfully characterized using dynamical system techniques and continuation analysis on a reduced-order model as reviewed by Juniper & Sujith (2018). The acoustics are governed by linearized one-dimensional momentum and energy equations. The heat release rate perturbations are governed by the kinematics of the flame surface. Dowling (1999) showed that the kinematics of the flame surface are the major source of nonlinearity in a ducted premixed flame.

A data-driven framework for the time-accurate calculation of a ducted premixed flame using the reduced-order model requires two components: (i) a computational method to predict the motion of the flame surface (Section 2.1) and (ii) a statistical algorithm to find the optimal estimate from a model prediction and experimental observations (Section 2.2).

2.1. *Hamilton-Jacobi equation*

The laws of motion of a surface are given by

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{u} - s_L\mathbf{n}, \tag{2.1}$$

where $\mathbf{x}$ is the position of one point on the surface, and $\mathbf{n}$ is the normal vector at this point. $\mathbf{u}$ is the velocity field of the underlying medium, and s_L is the speed of the surface relative to the underlying medium. If we assume hyperbolicity, the laws of motion are equivalent to the following Hamilton-Jacobi equation (Yu *et al.* 2018):

$$\frac{\partial G}{\partial t} + (\mathbf{u}\cdot\mathbf{n} - s_L) = 0, \tag{2.2}$$

$$\text{subject to } G(\mathbf{x}(t),t) - G(\mathbf{x}(0),0) = 0, \tag{2.3}$$

$$\text{and } \sqrt{\nabla G \cdot \nabla G} - 1 = 0. \tag{2.4}$$

In combustion, Eq. (2.2) is also known as the G-equation. The normal vector $\mathbf{n}$ depends on the computation of partial derivatives in space, which makes Eq. (2.2) a partial differential equation (PDE). Equations (2.3) and (2.4) are constraints, which make the solution to the G-equation unique away from the surface. The solution $G(\mathbf{x},t)$ to the Hamilton-Jacobi equation is the so-called generating function. The solution $\mathbf{x}(t)$ to Eq. (2.3) gives the location of the surface at every time t.

There are several consequences to solving the Hamilton-Jacobi equation (Eq. (2.2)-(2.4)) instead of the laws of motion (Eq. (2.1)). Firstly, the laws of motion require a parameterization of the surface. The quality of the parameterization quickly deteriorates when the metric on the surface or its topology changes. The Hamilton-Jacobi equation avoids these issues by embedding the surface as a level set of the generating function, which is defined over the entire domain. Secondly, the Hamilton-Jacobi equation needs to be in theory solved over the entire domain even though surfaces are one-codimensional (i.e. surfaces have one less dimension than the volume they inhabit). This is more computationally expensive but has the advantage that surfaces can be identified by their generating functions as vectors in a state space. A well-defined state space is a prerequisite for data assimilation, which will be described in the next section.

The G-equation is solved using a computationally inexpensive narrow-band level-set method (Peng *et al.* 1999): The computational domain is discretized using a weighted essentially non-oscillatory (WENO) scheme and a total-variation diminishing (TVD) version of the Runge-Kutta scheme. This gives third-order accuracy in time and up to fifth-order accuracy in space. Cusps and pinched-off fuel-air pockets are reliably captured. The generating function is reconstructed from the solution to the G-equation using a fast-marching method (Sethian 1996).

2.2. *Ensemble Kalman filter*

The Kalman filter provides a statistically optimal estimate of the true state of a system from a model prediction and experimental observations. In general, the model prediction is represented by a vector ψ^f in the state space (where 'f' denotes 'forecast'). The experimental observations are collected in a vector $\mathbf{d}$ in the observation space. The model prediction is mapped from the state space to the observation space through a linear measurement operator $\mathbf{M}$. The prediction uncertainty and the experimental errors are represented by covariance matrices $\mathbf{C}^f_{\psi\psi}$ and $\mathbf{C}_{\epsilon\epsilon}$ respectively. The statistically optimal estimate of the state ψ^a and its uncertainty $\mathbf{C}^a_{\psi\psi}$ (where 'a' denotes 'analysis') are given by (Evensen 2009)

$$\psi^a = \psi^f + \left(\mathbf{M}\mathbf{C}^f_{\psi\psi}\right)^T \left[\mathbf{C}_{\epsilon\epsilon} + \mathbf{M}\mathbf{C}^f_{\psi\psi}\mathbf{M}^T\right]^{-1} \left(\mathbf{d} - \mathbf{M}\psi^f\right), \tag{2.5}$$

$$\mathbf{C}^a_{\psi\psi} = \mathbf{C}^f_{\psi\psi} - \left(\mathbf{M}\mathbf{C}^f_{\psi\psi}\right)^T \left[\mathbf{C}_{\epsilon\epsilon} + \mathbf{M}\mathbf{C}^f_{\psi\psi}\mathbf{M}^T\right]^{-1} \left(\mathbf{M}\mathbf{C}^f_{\psi\psi}\right). \tag{2.6}$$

For the ducted premixed flame, the state vectors ψ^f and ψ^a are the generating functions $G(\mathbf{x},t)$ from Section 2.1. After discretization, the state vectors have $\mathcal{O}(10^5)$ entries and the covariance matrices $\mathcal{O}(10^{10})$ entries. The computation and manipulation of the covariance matrices make the Kalman filter computationally infeasible. Alternatively, the prediction uncertainty may be approximated by an ensemble ψ^f_i of size N. The

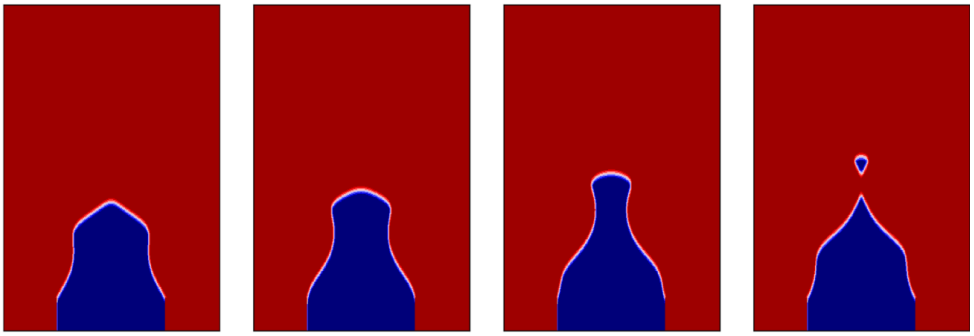

Figure 1: Solutions of G-equation over one cycle. The fuel-air mixture leaves the burner at the bottom. The infinitely thin flame surface separates the burnt (red) from the unburnt (blue) gas.

statistically optimal estimates ψ_i^a and $\mathbf{C}_{\psi\psi}^a$ from the ensemble Kalman filter are given by (Evensen 2009)

$$\psi_i^a = \psi_i^f + \left(\mathbf{M}\mathbf{C}_{\psi\psi}^f\right)^T \left[\mathbf{C}_{\epsilon\epsilon} + \mathbf{M}\mathbf{C}_{\psi\psi}^f\mathbf{M}^T\right]^{-1} \left(\mathbf{d} - \mathbf{M}\psi_i^f\right), \tag{2.7}$$

$$\overline{\psi} = \frac{1}{N}\sum_{i=1}^{N}\psi_i \quad , \quad \mathbf{C}_{\psi\psi} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\psi_i - \overline{\psi}\right)\left(\psi_i - \overline{\psi}\right)^T. \tag{2.8}$$

For combined state and parameter estimation, the state vectors are augmented by appending the parameters of interest to the discretized generating functions.

3. Results

Solutions of the G-equation for the ducted premixed flame are shown in Figure 1. The flame is attached to the burner lip, while the perturbations are convected from the base of the flame to the tip. If the perturbations are sufficiently large, a fuel-air pocket pinches off. In our reduced-order model, the convection of perturbations is mainly governed by two non-dimensional parameters (Kashinath *et al.* 2013): (i) the parameter K, which governs the perturbation convection speed, and (ii) e_a, which governs the amplitude of the response of the flame surface to acoustic excitation. Both parameters enter the G-equation via the underlying velocity field $\mathbf{u}$. Neither parameter is accurately known a priori, which is a major source of uncertainty.

In the following sections, two types of problems are addressed: The forward problem quantifies the propagation from the input (the uncertainty in the parameters) to the output (the uncertainty in the location of the flame surface) (Section 3.1). In the inverse problem, data assimilation is used to identify the location of the flame surface, and reduce the uncertainty in the location and the parameters (Sections 3.2-3.3). In Section 3.1, no data is used. In Sections 3.2-3.3, synthetic and DNS data are used respectively.

3.1. *Forward uncertainty quantification*

In the ensemble Kalman filter, the probability distibutions over the discretized generating functions are evolved in time using a Monte-Carlo method. The marginal probability distribution for the k-th entry in the state vector representing the discretized generating

function is given by

$$\psi[k] \sim \mathcal{N}(\overline{\psi}[k], \mathbf{C}_{\psi\psi}[k,k]), \tag{3.1}$$

where $\mathcal{N}$ denotes the normal distribution. The mean $\overline{\psi}[k]$ and the variance $\mathbf{C}_{\psi\psi}[k,k]$ are computed from Eq. (2.8). Hence the likelihood for the flame surface to be found at the location k is given by

$$p[k] = \frac{1}{\sqrt{2\pi\mathbf{C}_{\psi\psi}[k,k]}} \exp\left(-\frac{\overline{\psi}[k]^2}{2\mathbf{C}_{\psi\psi}[k,k]}\right). \tag{3.2}$$

In Figure 2, the logarithm of the normalized likelihood is shown,

$$\log\left(\frac{p[k]}{p_0[k]}\right) = -\frac{\overline{\psi}[k]^2}{2\mathbf{C}_{\psi\psi}[k,k]}. \tag{3.3}$$

The zero-level set gives the maximum-likelihood location of the flame surface. The more negative the value at a location is, the less likely the flame surface is to be found there. As the perturbation is convected from the base of the flame to the tip, the high-likelihood volume for the location of the flame surface grows. The largest area exists during pinch-off, which represents maximal uncertainty.

3.2. *Twin experiment*

In a twin experiment, both the model predictions and the experimental observations are generated by solving the G-equation. For the model predictions, an ensemble of twenty G-equations is solved. The same initial condition is used for the whole ensemble, but a different set of parameters K and e_a is chosen for each simulation. They are sampled from two independent normal distributions with 20 % standard deviation respectively. For the experimental observations, a separate set of parameters K and e_a is chosen. The measurements are taken at all grid points where the absolute value of the discretized generating function is less than one grid size. This gives a cloud of grid points inside a narrow band around the zero-level set of the generating function. Corresponding to the zero-level set, the vector of measurements $\mathbf{d}$ is zero-valued. The measurement matrix $\mathbf{M}$ is a restriction operator, which maps the state space to the observation space and indicates the locations of the grid points inside the narrow band. The covariance matrix $C_{\epsilon\epsilon}$, which represents the experimental errors, is a diagonal matrix with the grid size squared on the diagonal.

In a first twin experiment, the ensemble Kalman filter is used to perform state estimation. In the middle row of Figure 2, the logarithm of the normalized likelihood (Eq. (3.3)) is shown. A qualitative comparison to the top row of Figure 2 shows that the high-likelihood volume for the location of the flame surface is significantly reduced. A quantitative measure for the uncertainty in the location of the flame surface is given by the root mean square (RMS) error, which is defined as the square-root of the trace of the covariance matrix of the ensemble:

$$\mathrm{tr}\left(\mathbf{C}_{\psi\psi}\right) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\psi_i - \overline{\psi}\right)^T\left(\psi_i - \overline{\psi}\right). \tag{3.4}$$

In Figure 3, the RMS error is plotted over time. At timestep 0, the error is zero because the same initial condition is used for the whole ensemble. In the forward problem, the RMS error grows until it reaches a high-uncertainty plateau. Peaks in the high-uncertainty plateau coincide with the moments when fuel-air pockets are pinched off.

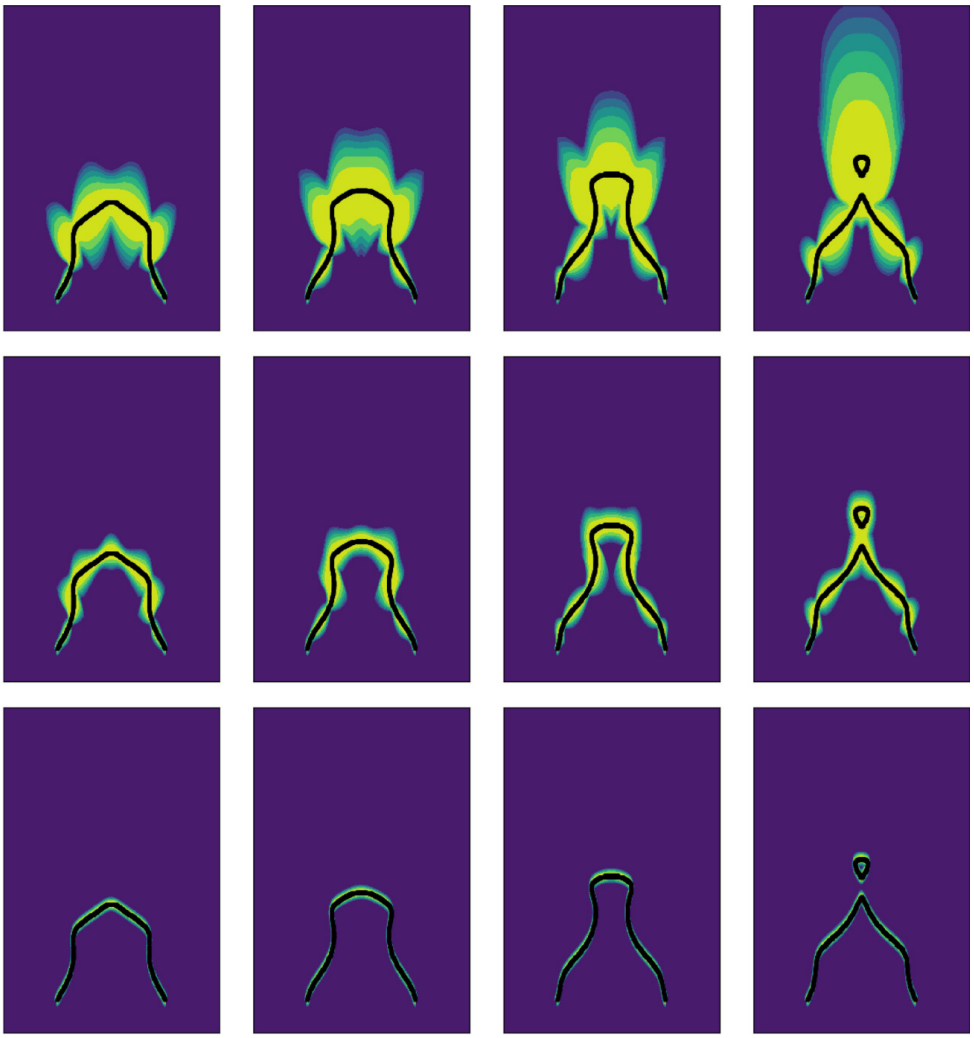

Figure 2: Snapshots of logarithm of normalized likelihood over one cycle for the forward problem (top row) and the inverse problems with either state estimation (middle row) or combined state and parameter estimation (bottom row) respectively. In the twin experiment, the experimental observations are extracted from a G-equation simulation (black line). High-likelihood (yellow) and low-likelihood (blue) volumes are shown.

With state estimation, the uncertainty is regularly capped, and the predictions of the location of the flame surface significantly improve.

In a second twin experiment, the ensemble Kalman filter is used to perform combined state and parameter estimation. In Figure 3, the RMS error is plotted over time. The initial behavior is similar to that of the twin experiment with state estimation. After the first data assimilation at timestep 1000, the uncertainty remains at a relatively constant level. At this point, the combined state and parameter estimation has updated the parameters to values close to the truth. Consequently, the subsequent uncertainty does not grow and the parameters do not improve. When the first fuel-air pocket is pinched off, at timestep

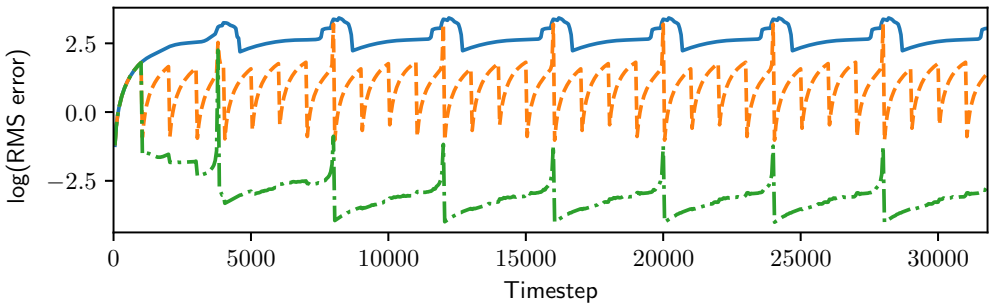

Figure 3: Logarithm of root mean square (RMS) error plotted over time for the forward problem (blue line) and the inverse problems with either state estimation (orange dash) or combined state and parameter estimation (green dash-dot).

4000, the uncertainty grows rapidly until the combined state and parameter estimation updates the state and the parameters again. The update step coincides with the moment of pinch-off, which is also a moment of high uncertainty. Thus the parameters can be found very accurately, which leads to a low-uncertainty plateau. High uncertainties at later pinch-off events are suppressed without much change in the parameters.

3.3. *Data from direct numerical simulation*

After the twin experiment, the assimilated data is extracted from a DNS performed with the compressible flow solver CharlesX. The computational mesh has 500,000 grid points, the grid size being 0.2 mm in the region of the flame. At the inlet of the domain, an ethylene-methane-air mixture comes out of the burner, surrounded by an air coflow. The reduced chemical mechanism includes 15 species and 5 quasi-steady state species. At the outlet of the domain, a sponge region is implemented to suppress reflected acoustic waves. In Figure 4, snapshots of the DNS are shown. The flame is attached to the burner lip on the edge of flashing back. The DNS displays the same qualitative behavior as the G-equation in that perturbations are observed to travel from the base to the tip of the flame. In the DNS, there are no pinched-off fuel-air pockets. For the model predictions, an ensemble of twenty-four G-equations is solved. The same initial condition is used for the whole ensemble, but a different set of parameters K and e_a is chosen for each simulation. They are sampled from two independent normal distributions with 25 % and 5 % standard deviation respectively. The covariance matrix $C_{\epsilon\epsilon}$ is a diagonal matrix with nine times the grid size squared on the diagonal. In Figure 5, the logarithm of the normalized likelihood is shown for the forward problem and the inverse problems, the latter involving either state estimation or combined state and parameter estimation. As in Figure 2, data assimilation significantly improves the ability of the reduced-order G-equation model to capture the motion of the flame surface. This is quantitatively confirmed by the RMS error shown in Figure 6. While the motion of the flame surface is accurately captured towards the tip of the flame, the attachment of the flame to the burner lip is not.

Finally, we build on our understanding of the state and parameter uncertainties to attempt model uncertainty quantification. Unlike optimization-based approaches, Bayesian statistics do not focus on the minimization of one cost functional but revolve around probability distributions. As such, they can be analyzed using histograms for example.

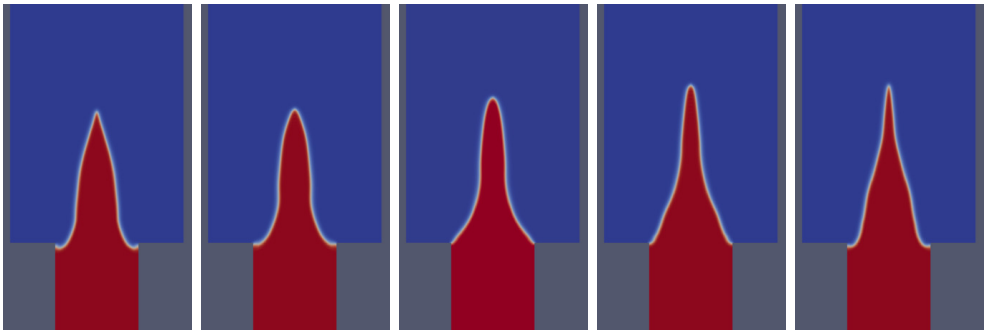

Figure 4: Ethylene mass fractions in direct numerical simulation (DNS) over one cycle. The fuel-air mixture leaves the burner at the bottom. The flame surface separates the burnt (blue) from the unburnt (red) gas.

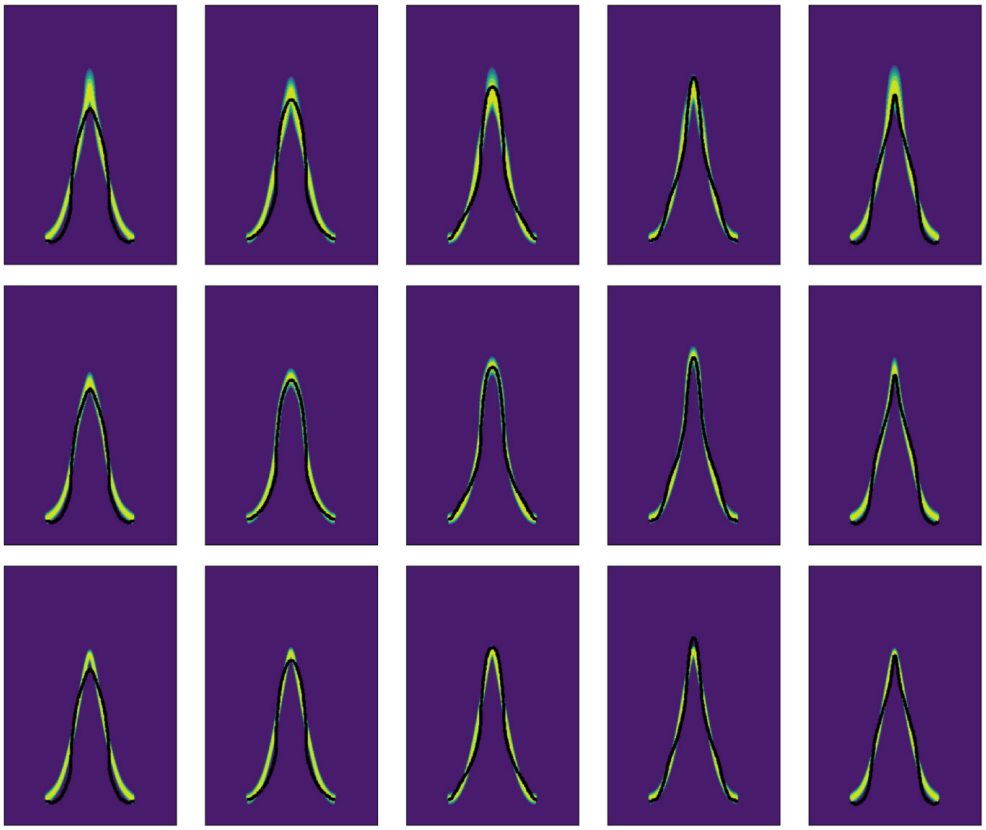

Figure 5: Snapshots of logarithm of normalized likelihood over one cycle for the forward problem (top row) and the inverse problems with either state estimation (middle row) or combined state and parameter estimation (bottom row) respectively. For the data assimilation, the experimental observations are extracted from a DNS (black line). High-likelihood (yellow) and low-likelihood (blue) volumes are shown.

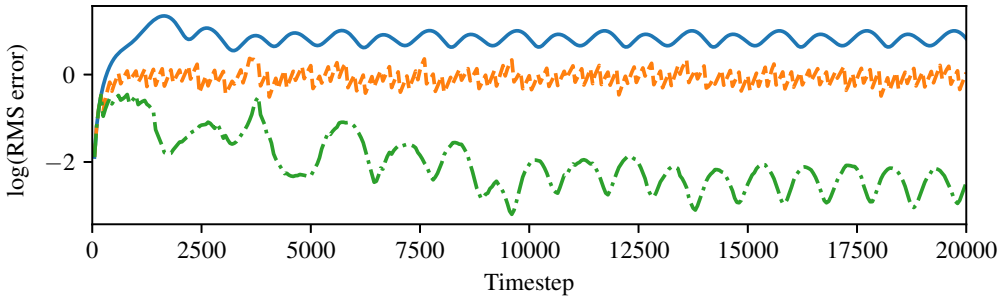

Figure 6: Logarithm of root mean square (RMS) error plotted over time for the forward problem (blue line) and the inverse problems with either state estimation (orange dash) or combined state and parameter estimation (green dash-dot).

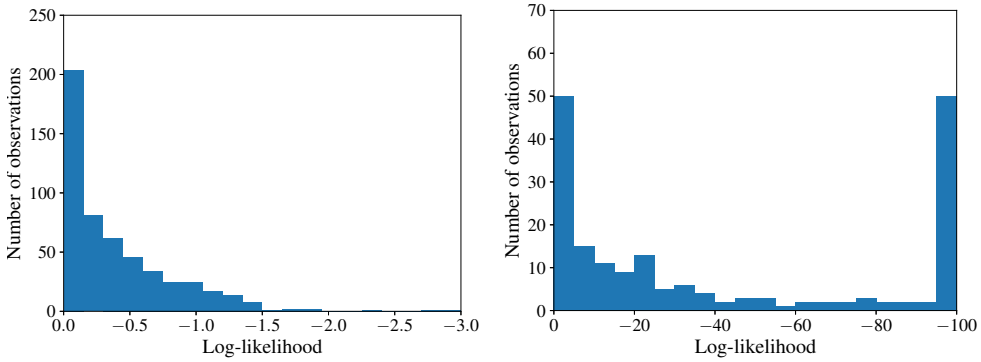

Figure 7: Histograms of experimental observations over logarithm of normalized likelihood (twin experiment left, assimilation of DNS data right). High-likelihood bins are located on the left of each histogram, low-likelihood bins on the right.

In Figure 7, a preliminary result is shown. In the twin experiment (left), most observation points are located in the high-likelihood volume. The likelihood of finding the flame surface in the low-likelihood volume decays exponentially. When assimilating DNS data (right), the histogram has a fat tail. The deviation from an exponentially decaying distribution is a piece of information non-existant in an optimization-based approach, and indicates model error. Model uncertainty is traditionally the most difficult to quantify, and Bayesian statistics promise to be a viable approach.

4. Conclusions

We propose an on-the-fly statistical learning method, based on data assimilation with the ensemble Kalman filter, to improve the prediction of the dynamics of premixed flames. The proposed framework can also be applied to other interface-tracking problems. First, the capabilities of the framework are demonstrated in a twin experiment, where the assimilated data is produced from the same model as that used in prediction. This guarantees that the assimilated data is consistent with the predictions, which is particularly useful to validate the algorithm. The uncertainties of the calculations are estimated by using

Bayesian statistics. It is shown that the major source of uncertainty are the unknown parameters while the pinch-off events are extremely sensitive to both parameters and initial conditions. Second, the assimilated data is extracted from a high-fidelity reacting-flow DNS. State estimation and combined state and parameter estimation again provide significant improvements in the model predictions. While the motion of the flame surface is well captured towards the tip of the flame, the attachment of the flame to the burner lip is not. Bayesian statistics reveal the reason to be neither state nor parameter uncertainties but deficiencies in the model. While the convection of perturbations along the flame surface is accurately modeled, the reduced-order G-equation model does not include the physical mechanisms necessary to model the flame attachment, i.e. heat loss to the burner and the shear layer in the wake of the burner. This physical insight into the significance of the pinch-off events and the flame attachment informs both the modeling assumptions as well as the design of experiments towards truly accurate prediction, for example in turbulent flows (Labahn *et al.* 2018).

REFERENCES

Culick, F. E. C. 2006 *Unsteady motions in combustion chambers for propulsion systems*. RTO AGARDograph AG-AVT-039, North Atlantic Treaty Organization.

Dowling, A. P. 1999 A kinematic model of a ducted flame. *J. Fluid Mech.* **394**, 51–72.

Evensen, G. 2009 *Data Assimilation*. Springer Berlin Heidelberg.

Jazwinski, A. H. 2007 *Stochastic Processes and Filtering Theory*. Dover Publications.

Juniper, M. P. & Sujith, R. 2018 Sensitivity and Nonlinearity of Thermoacoustic Oscillations. *Annu. Rev. Fluid Mech.* **50**, 661–689.

Kabiraj, L., Saurabh, A., Wahi, P. & Sujith, R. I. 2012 Route to chaos for combustion instability in ducted laminar premixed flames. *Chaos* **22**.

Kashinath, K., Hemchandra, S. & Juniper, M. P. 2013 Nonlinear thermoacoustics of ducted premixed flames: The influence of perturbation convection speed. *Combust. Flame* **160**, 2856–2865.

Kashinath, K., Waugh, I. C. & Juniper, M. P. 2014 Nonlinear self-excited thermoacoustic oscillations of a ducted premixed flame: bifurcations and routes to chaos. *J. Fluid Mech.* **761**, 399–430.

Labahn, J. W., Wu, H., Coriton, B., Frank, J. H. & Ihme, M. 2018 Data assimilation using high-speed measurements and LES to examine local extinction events in turbulent flames. *P. Combust. Inst.* .

Lieuwen, T. C. & Yang, V. 2005 *Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms, and Modeling*. American Institute of Aeronautics and Astronautics, Inc.

Peng, D., Merriman, B., Osher, S., Zhao, H. & Kang, M. 1999 A PDE-Based Fast Local Level Set Method. *J. Comput. Phys.* **155**, 410–438.

Sethian, J. A. 1996 A fast marching level set method for monotonically advancing fronts. *P. Natl. Acad. Sci. USA* **93**, 1591–1595.

Yu, H., Garita, F., Juniper, M. P. & Magri, L. 2018 Data assimilation and parameter estimation of thermoacoustic instabilities in a ducted premixed flame. *UK Fluids Conference 2018*, Manchester, UK.