Bayesian Inference of Flame Impulse Responses

Matthew Yoko[a,*], Wolfgang Polifke[b]

[a] *Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, United Kingdom.*
[b] *Department of Engineering Physics and Computation, Technical University of Munich, Boltzmannstr. 15, Garching, 85748, Germany.*

Abstract

The impulse response of a flame to acoustic velocity perturbations is a key quantity for predicting thermoacoustic stability, but its identification from sparse, noisy observations requires solving an ill-posed inverse convolution problem. Standard system identification methods typically proceed in two stages: they first estimate an impulse response from the data, using hand-tuned choices of regularization, model order, and sampling parameters, and then post-process that estimate to extract physically meaningful quantities such as time delays. This process provides no principled mechanism for incorporating prior physical knowledge during inference. In this paper, we instead reformulate the identification problem within a Bayesian framework by imposing from the outset a physically motivated distributed time-delay model in terms of a given number of Gaussian pulses. The corresponding model parameters are inferred directly from the data, providing both the parameters and the impulse response in one step. The number of pulses best supported by the data is then identified using Bayesian model comparison, which balances data fit against model complexity to identify the simplest model capable of explaining the data. The framework is demonstrated on broadband-forced large eddy simulation data from a turbulent swirl-stabilized burner. Bayesian model comparison selects a three-Gaussian impulse response for this flame, consistent with physical interpretations in previous work. Compared with system identification, the Bayesian approach produces impulse responses with fewer spurious features and enables straightforward enforcement of a known low-frequency gain. Finally, we show that the Bayesian approach is robust to significant reductions in signal length, making it appealing for impulse response identification from costly simulations, where there is an incentive to minimize simulation duration.

Thermoacoustic instability, Flame impulse response, System identification, Bayesian inference

∗ Corresponding author: Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, United Kingdom.

Novelty and significance statement

In this paper, we introduce a Bayesian framework for identifying flame impulse responses from sparse, noisy input-output data. The key novelty is that a physically motivated distributed time-delay model is imposed *during* inference, so that the impulse response and its underlying delay parameters are identified simultaneously rather than in two separate steps. This replaces the *ad hoc* choices required in standard system identification with a principled probabilistic formulation. Physical constraints such as causality, smoothness, characteristic convective timescales, and a known low-frequency gain are encoded directly through the model structure and parameter priors. A further novelty is that model order selection is handled automatically through Bayesian model comparison, rather than by manual inspection. These features make the method more interpretable, more physically consistent, and more robust when only short data records are available, which is especially important for costly simulations of turbulent reacting flows.

1. Introduction

The linear response of a premixed flame to an arbitrary acoustic velocity disturbance is entirely described by the flame's impulse response, or equivalently its transfer function [1, 2]. For a given perturbation in velocity u', the resulting perturbation in heat release rate q' can be expressed as the convolution of u' with the impulse response h:

$$q' = u' * h = \int_{-\infty}^{\infty} u'(t-\tau)h(\tau)\mathrm{d}\tau \qquad (1)$$

Given a flame's impulse response and an acoustic model of the system, it is therefore possible to predict the thermoacoustic stability of a system using linear analysis. The impulse response, however, is generally unknown and not directly measurable. A core challenge in thermoacoustics is to determine the impulse

e-mail: mjy29@cam.ac.uk (Matthew Yoko)

response or transfer function of a flame from sparse, noisy observations of u' and q', either from experiments or from numerical simulations. This requires solving an inverse convolution problem:

$$\text{given } (u', q'), \text{ find } h \text{ such that } q' = u' * h. \quad (2)$$

Inverse convolution problems are generally ill-posed, meaning that small errors in the data can lead to large errors in the estimated impulse response. Therefore, care must be taken in how the problem is formulated and solved.

A common approach is to excite the system with a broadband signal, such as filtered white noise, and record the time-domain response of the heat release rate [3, 4]. By capturing M discrete samples of the input-output data (u_m, q_m) at uniform intervals Δt, the discrete convolution with an impulse response of length L can be expressed as:

$$q_m = \sum_{l=0}^{L-1} u_{m-l} h_l \quad (3)$$

where $q_m = q'(m\Delta t)$, $u_{m-l} = u'((m-l)\Delta t)$, and $h_l = h(l\Delta t)$. Note that we expect the output to only depend on past inputs, i.e. the flame cannot preempt a velocity perturbation. This requires special treatment for the entries of u_{m-l} for which $m - l < 0$. We discuss this in detail in Section 4.

The discrete convolution in eq. (3) can be expressed in matrix form as:

$$\boldsymbol{q} = \boldsymbol{U}\boldsymbol{h} \quad (4)$$

where $\boldsymbol{q} = [q_0, q_1, \ldots, q_{M-1}]^\top$, $\boldsymbol{h} = [h_0, h_1, \ldots, h_{L-1}]^\top$, and $\boldsymbol{U} \in \mathbb{R}^{M\times L}$ is the Toeplitz matrix formed from the discrete input signal $\boldsymbol{u} = [u_0, u_1, \ldots, u_{M-1}]^\top$. The impulse response can then be estimated by solving the regularized least squares problem:

$$\hat{\boldsymbol{h}} = (\boldsymbol{U}^\top \boldsymbol{U} + \lambda \boldsymbol{I})^{-1} \boldsymbol{U}^\top \boldsymbol{q} \quad (5)$$

where $\boldsymbol{I}$ is the identity matrix and λ is the regularization parameter, which must be hand-tuned to provide sufficient smoothing without excessively biasing the result.

After fitting $\hat{\boldsymbol{h}}$, the residual $\boldsymbol{e} = \boldsymbol{q} - \boldsymbol{U}\hat{\boldsymbol{h}}$ can be analysed to obtain approximate uncertainty bounds on the impulse response, which can then be propagated to the transfer function [5, 6].

While variants of this approach have been widely used [6–13], it has several drawbacks. The matrix $\boldsymbol{U}^\top\boldsymbol{U}$ can be ill-conditioned depending on the properties of the input signal, leading to numerical instability in the inversion, and amplification of small errors. While regularization can ease this, the choice of regularization parameter λ is arbitrary and can bias the estimated impulse response. Most importantly, this method requires that all information about the impulse response is provided by the data, with no way to incorporate prior knowledge about the physics of the system.

In general, we have some broad but useful prior knowledge about the expected form of the impulse response. For example, we know that the impulse response must be smooth, and that it typically consists of a few characteristic time delays related to the physical processes in the flame [1]. We know that these time delays are typically governed by convective transport processes, providing a natural timescale for the impulse response. We know that the impulse response must be causal (i.e. all time delays must be positive), and we know that its integral describes the steady-state gain of the flame, which can often be prescribed based on physical arguments [14].

In this work, we reformulate this inverse convolution problem as a Bayesian inference problem, allowing us to combine the observed data with prior knowledge about the expected form of the impulse response. By proposing a physically-motivated parametric model for the impulse response, such as a distributed time delay model [1], we can encode physical constraints such as smoothness, causality and characteristic timescales. This leads to a more robust and interpretable estimation procedure that is less sensitive to noise and numerical instability.

This paper is structured as follows. In Section 2, we introduce the physically motivated model for the impulse response. In Section 3, we formulate the inverse problem in a Bayesian framework, defining the prior, likelihood, and posterior distributions. In Section 4, we describe a robust solution procedure for the Bayesian inverse problem. In Section 5, we present results from applying this framework to large eddy simulation (LES) data, before concluding in Section 6.

2. Physically motivated model of the impulse response

We have prior knowledge that flame dynamics are typically governed by convective processes with a few characteristic time delays, and often some diffusion or dispersion leading to these delays being distributed in time [1]. Several studies have proposed modelling the flame impulse response as a sum of Gaussian delays to capture this behaviour [1, 8, 15–19], so we adopt this as the parametric model for the impulse response:

$$h(t; N, \boldsymbol{a}) = \sum_{i=1}^{N} n_i \, (2\pi\sigma_i^2)^{-1/2} \exp\left[-\frac{(t-\tau_i)^2}{2\sigma_i^2}\right] \quad (6)$$

where N is the number of Gaussian pulses in the model and $\boldsymbol{a}$ is the vector of model parameters, which comprise: n_i, the coefficient, τ_i, the time delay, and σ_i, the width of each Gaussian pulse.

We will refer to this as the N-n-τ-σ model of the impulse response. By exploiting this prior knowledge of the physics, the problem stated in eq. (2) has re-

duced to:

$$\text{given } (u', q'), \text{ find } \{N, \boldsymbol{a}\} \text{ such that } q' = u' * h(t; N, \boldsymbol{a}) \tag{7}$$

We have therefore reduced the inverse convolution problem, which requires estimating an arbitrary function h in a high-dimensional function space, into a low-dimensional parameter estimation problem. In this parameter estimation problem, we have one discrete parameter N (the number of Gaussian delays) and $P = 3N$ continuous parameters (the coefficients, time delays, and widths of each Gaussian), each of which must be estimated from the data.

If we have prior knowledge about the low-frequency limit of the flame transfer function, which is often the case [14], the N-n-τ-σ model can be easily modified to enforce this constraint. The low-frequency limit of the transfer function is given by the time-integral of the impulse response, which for the N-n-τ-σ model is simply the sum of the coefficients. For a low-frequency gain of G, the coefficients must satisfy the constraint $\sum_{i=1}^{N} n_i = G$. One convenient implementation is to eliminate one coefficient, for example $n_N = G - \sum_{i=1}^{N-1} n_i$. The choice of eliminated coefficient is arbitrary and has no physical significance or impact on the results.

3. A probabilistic formulation of the inverse problem

We now formulate the inverse problem in eq. (7) in a Bayesian framework, allowing us to combine the measured data with prior physical knowledge about the parameters. We treat the discrete and continuous parameters separately.

The N-n-τ-σ model defines a *family* of models indexed by the model order N. For a fixed model order, our aim is to identify the continuous parameters that are most consistent with both the observed data and our prior physical expectations. This gives a best estimate of the parameters as well as a measure of how uncertain those estimates are. We then repeat this procedure for all plausible model orders and use Bayesian model comparison to determine which model order is most strongly supported by the data.

3.1. Bayesian parameter inference

In the Bayesian framework, our uncertainty in the unknown parameters $\boldsymbol{a} = \{\boldsymbol{n}, \boldsymbol{\tau}, \boldsymbol{\sigma}\}$ is represented by a probability density function (pdf) [20, Ch. 2]. Before any data is observed, this uncertainty is described by a *prior* pdf $p(\boldsymbol{a}|N)$, which encodes the parameter values we regard as plausible for a model of order N.

After observing the data $\boldsymbol{q}$, this prior knowledge is updated to a *posterior* pdf $p(\boldsymbol{a}|\boldsymbol{q}, N)$, which describes which parameter values are most plausible given both the data and the prior information [20, Ch. 3]. This update is governed by Bayes' theorem:

$$p(\boldsymbol{a}|\boldsymbol{q}, N) = \frac{p(\boldsymbol{q}|\boldsymbol{a}, N)\, p(\boldsymbol{a}|N)}{p(\boldsymbol{q}|N)} \tag{8}$$

where the *likelihood* $p(\boldsymbol{q}|\boldsymbol{a}, N)$ measures how well a given set of parameters explains the observed data, and the *evidence* $p(\boldsymbol{q}|N)$ can be viewed simply as the normalization constant that ensures the posterior pdf integrates to one. Its further role in comparing model orders is discussed in Section 3.2. We now define the prior and likelihood in turn, and then show how the posterior is approximated.

3.1.1. The prior

For a model of order N, we define prior pdfs over the plausible values of each of the continuous parameters that define the impulse response. At a high level, the prior is intended to encode broad physical expectations about the coefficients, delays, and widths of the Gaussian pulses, while discouraging unphysical parameter values. In practice, it is more robust to define the prior over a transformed set of parameters, which is discussed in detail in Section 4.1. At this stage, it is sufficient to note that the prior is chosen to be Gaussian:

$$p(\boldsymbol{a}|N) = \mathcal{N}(\boldsymbol{a}; \boldsymbol{a}_0, \boldsymbol{\Sigma}_{\boldsymbol{a}}) \tag{9}$$

where $\boldsymbol{a}_0$ is the prior mean and $\boldsymbol{\Sigma}_{\boldsymbol{a}}$ is the prior covariance matrix, whose diagonal entries represent the prior uncertainty in each parameter and off-diagonal entries represent correlations between parameters. We typically do not have strong prior knowledge about how the parameters are correlated, so we set the off-diagonal entries to zero.

In general, we know little about the parameter values *a priori*, but we are able to bound them within reasonable ranges based on physical arguments. If we are able to estimate a minimum and maximum expected value for each parameter, this is sufficient to define a Gaussian prior with a mean at the center of the range and a standard deviation such that 99.7% of the probability mass lies within the range (i.e., plus/minus three standard deviations). This leads to broad but useful priors that can be applied to a wide range of flames without requiring detailed prior knowledge. This will be explored further in Section 4.

3.1.2. The likelihood

The likelihood quantifies how compatible a candidate parameter vector is with the observed data. To define it, we specify a model for the discrepancy between the measured heat release rate signal, $\boldsymbol{q}$, and the signal predicted by the impulse-response model, $\hat{\boldsymbol{q}} = \boldsymbol{U}\,\boldsymbol{h}(\boldsymbol{a})$. In this paper, we neglect noise in $\boldsymbol{u}$ and represent the remaining mismatch in the output equation:

$$\boldsymbol{q} = \boldsymbol{U}\,\boldsymbol{h}(\boldsymbol{a}) + \boldsymbol{\varepsilon} \tag{10}$$

This is appropriate when the input signal is obtained from a measurement or reconstruction that suppresses uncorrelated fluctuations, such as the multi-microphone method in experiments or the area-integral of the volume flux through a reference plane

in LES, which is the case demonstrated in Section 5. In this model, the discrepancy term $\boldsymbol{\varepsilon}$ represents not only measurement and acquisition noise in $\boldsymbol{q}$, but also model-form error of the restricted N-n-τ-σ family, small departures from the linear time-invariant assumption, turbulent fluctuations, and numerical or post-processing artefacts.

We assume this discrepancy is approximately Gaussian with zero mean, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_q)$. This is a natural choice when the total discrepancy is the sum of many small contributions, and when only its mean and covariance can be specified [21, §19.2.1]. Under this assumption the likelihood is:

$$p(\boldsymbol{q}|\boldsymbol{a}, N) = \mathcal{N}(\boldsymbol{q}; \hat{\boldsymbol{q}}(\boldsymbol{a}), \boldsymbol{\Sigma}_q). \tag{11}$$

In the simplest case we take $\boldsymbol{\Sigma}_q = \sigma_q^2 \boldsymbol{I}$, corresponding to temporally uncorrelated discrepancy, where σ_q can either be prescribed or estimated from the data as described in Section 3.3. In general, the discrepancy may be temporally correlated, in which case a parametric model for $\boldsymbol{\Sigma}_q$ can be introduced, but this is left for future work. In the current paper, we minimize the effect of temporal correlation by downsampling the data below the correlation length.

If noise in the measured input is substantial, as may occur with pointwise velocity measurements in turbulent flows, a noise model for the input can be introduced. This leads to a modified Gaussian likelihood with a transformed covariance, which depends on the parameters. However, we do not pursue this extension in this paper.

We now expand the likelihood function in order to gain a deeper understanding of its properties:

$$p(\boldsymbol{q}|\boldsymbol{a}, N) = Z \exp\left[-\frac{1}{2}(\boldsymbol{q} - \hat{\boldsymbol{q}}(\boldsymbol{a}))^\top \boldsymbol{\Sigma}_q^{-1} (\boldsymbol{q} - \hat{\boldsymbol{q}}(\boldsymbol{a}))\right] \tag{12}$$

where $Z = |2\pi\boldsymbol{\Sigma}_q|^{-1/2}$ is the normalizing constant. In this form, it is clear that the likelihood depends on the weighted residual between the observed and predicted heat release rate signals. Parameter values therefore have high likelihood when they reproduce the data to within the level of discrepancy encoded by $\boldsymbol{\Sigma}_q$. Because the model prediction $\hat{\boldsymbol{q}}(\boldsymbol{a})$ depends nonlinearly on the parameters, the likelihood is generally non-Gaussian as a function of $\boldsymbol{a}$. This motivates the posterior approximation introduced in the following subsection.

3.1.3. Approximating the posterior

Equation (8) shows that if the prior and likelihood were both Gaussian in $\boldsymbol{a}$, the posterior would also be Gaussian. In our case, however, the nonlinear dependence of the forward model on $\boldsymbol{a}$ means that the posterior does not have a closed-form expression, so we approximate it in two steps: first, we identify the parameters that lie at the peak of the posterior, and then we estimate the width of the posterior around this peak [20, Ch. 27]. The first step gives the best estimate of the parameters, while the second gives an estimate of their uncertainty.

The parameters that lie at the peak of the posterior pdf are called the *maximum a posteriori* (MAP) estimate, and are given by:

$$\boldsymbol{a}^* = \arg\max_{\boldsymbol{a}} p(\boldsymbol{a}|\boldsymbol{q}, N). \tag{13}$$

Equivalently, we can minimize the negative log of the unnormalized posterior, which is more convenient for numerical optimization:

$$\boldsymbol{a}^* = \arg\min_{\boldsymbol{a}} [\mathcal{J}] \tag{14}$$

where the cost function $\mathcal{J}$ is defined as:

$$\begin{aligned} \mathcal{J} &= -\log p(\boldsymbol{q}|\boldsymbol{a}, N) - \log p(\boldsymbol{a}|N) \\ &= \frac{1}{2}(\hat{\boldsymbol{q}}(\boldsymbol{a}) - \boldsymbol{q})^\top \boldsymbol{\Sigma}_q^{-1} (\hat{\boldsymbol{q}}(\boldsymbol{a}) - \boldsymbol{q}) \\ &\quad + \frac{1}{2}(\boldsymbol{a} - \boldsymbol{a}_0)^\top \boldsymbol{\Sigma}_a^{-1} (\boldsymbol{a} - \boldsymbol{a}_0) + K \end{aligned} \tag{15}$$

where K denotes normalization terms that do not depend on $\boldsymbol{a}$. We now see the convenience of using Gaussian prior and likelihood functions: the negative log-posterior reduces to a sum of squared errors, weighted by the inverse covariances. We note, however, that the choice of Gaussian distributions is not driven by this convenience, but rather because it is the most natural choice for both the prior and the noise model, as discussed previously.

The MAP estimate can then be found efficiently using gradient-based optimization, with the gradient of $\mathcal{J}$ given by:

$$\nabla_{\boldsymbol{a}} \mathcal{J} = \boldsymbol{J}_{\boldsymbol{a}}^\top \boldsymbol{\Sigma}_q^{-1} (\hat{\boldsymbol{q}}(\boldsymbol{a}) - \boldsymbol{q}) + \boldsymbol{\Sigma}_a^{-1} (\boldsymbol{a} - \boldsymbol{a}_0) \tag{16}$$

where $\boldsymbol{J}_{\boldsymbol{a}} = \nabla_{\boldsymbol{a}} \hat{\boldsymbol{q}} \in \mathbb{R}^{M \times P}$ is the Jacobian of the forward model predictions with respect to the parameters. For a simple forward model such as our sum of Gaussian pulses, the gradient can be computed analytically or by using automatic differentiation. For more complex forward models, adjoint methods may be more appropriate, as they can be used to avoid computing this Jacobian directly [22].

Having found the peak of the posterior pdf, we now estimate its width using the Laplace approximation. We begin by performing a second order Taylor expansion of $\mathcal{J}$ around the MAP estimate $\boldsymbol{a}^*$, noting that $\mathrm{d}\mathcal{J}/\mathrm{d}\boldsymbol{a} = 0$ at $\boldsymbol{a}^*$:

$$\mathcal{J}(\boldsymbol{a}) \approx \mathcal{J}(\boldsymbol{a}^*) + \frac{1}{2}(\boldsymbol{a} - \boldsymbol{a}^*)^\top \boldsymbol{H} (\boldsymbol{a} - \boldsymbol{a}^*) \tag{17}$$

where $\boldsymbol{H} = \nabla_{\boldsymbol{a}} \nabla_{\boldsymbol{a}} \mathcal{J}|_{\boldsymbol{a}=\boldsymbol{a}^*} \in \mathbb{R}^{P \times P}$ is the Hessian of $\mathcal{J}$ evaluated at the MAP estimate. By inspection of eq. (17) (and comparison to eq. (15)), we see that to second order, the posterior pdf is a Gaussian distribution around the MAP estimate with inverse-covariance:

$$\boldsymbol{\Sigma}_{\boldsymbol{a}^*}^{-1} = \boldsymbol{H} \approx \boldsymbol{J}_{\boldsymbol{a}}^\top \boldsymbol{\Sigma}_q^{-1} \boldsymbol{J}_{\boldsymbol{a}} + \boldsymbol{\Sigma}_a^{-1} \tag{18}$$

Equation (18) shows that the local posterior precision is the sum of two contributions. The first, $\boldsymbol{J}_a^\top \boldsymbol{\Sigma}_q^{-1} \boldsymbol{J}_a$, represents the information provided by the data through the sensitivity of the model predictions to the parameters. The second, $\boldsymbol{\Sigma}_a^{-1}$, represents the information already supplied by the prior. As the data become more informative, the first term grows and the posterior becomes more sharply peaked around the MAP estimate, such that the posterior probability falls to near-zero before the effect of the nonlinearity becomes significant, and the true posterior is well-approximated by a Gaussian distribution. In Section 5 we will show with an LES case study that, even with moderate amounts of data, the true posterior is very close to Gaussian.

We now have the tools required to infer the impulse-response parameters for a fixed model order N and to estimate their uncertainty. In summary, we define (i) physically motivated prior pdfs over the parameters, and (ii) a Gaussian likelihood based on the forward convolution model and a model for the measurement discrepancy. We then combine these using Bayes' theorem to obtain the posterior parameter pdf, which we approximate using MAP estimation and the Laplace approximation. At this point we have, for a given model order N, an improved estimate of the parameters $\boldsymbol{a}^*$ and their uncertainty $\boldsymbol{\Sigma}_{\boldsymbol{a}^*}$. We now turn to the problem of determining which model order is most strongly supported by the data.

3.2. Bayesian model selection

In order to determine which model order is most strongly supported by the data, we apply Bayes' theorem to the discrete quantity N [20, Ch. 28]:

$$p(N|\boldsymbol{q}) \propto p(\boldsymbol{q}|N)p(N) \tag{19}$$

where $p(N)$ is the prior probability assigned to model order N. We typically have no reason to prefer one model order over another *a priori*, so we assign equal prior probability to all models. Model comparison then reduces to comparing the evidence $p(\boldsymbol{q}|N)$, which was introduced in eq. (8) as the normalization constant for the posterior parameter pdf.

The evidence is the probability of the observed data under model order N, averaged over all parameter values that the prior regards as plausible:

$$p(\boldsymbol{q}|N) = \int p(\boldsymbol{q}|\boldsymbol{a}, N)p(\boldsymbol{a}|N)\, d\boldsymbol{a}. \tag{20}$$

A model receives a large evidence only if it fits the data well for parameter values that are both plausible under the prior and not excessively fine-tuned. A very flexible model is therefore not automatically favoured. Although it may achieve an excellent fit at one parameter setting, its evidence is reduced if most of its parameter space fits the data poorly.

For nonlinear models, the integral in eq. (20) is generally intractable, so we again use the Laplace approximation. Substituting the Gaussian approximation from Section 3.1 into eq. (20) and performing the integration gives:

$$p(\boldsymbol{q}|N) \approx p(\boldsymbol{q}|\boldsymbol{a}^*, N)p(\boldsymbol{a}^*|N)|2\pi\boldsymbol{\Sigma}_{\boldsymbol{a}^*}|^{1/2} \tag{21}$$

This expression makes the trade-off between model fit and flexibility explicit. The first factor, $p(\boldsymbol{q}|\boldsymbol{a}^*, N)$, is the likelihood evaluated at the MAP point. This measures how well the model fits the data, and is called the best-fit likelihood (BFL). The remainder, $p(\boldsymbol{a}^*|N)|2\pi\boldsymbol{\Sigma}_{\boldsymbol{a}^*}|^{1/2}$, forms the Occam factor (OF), which penalizes models that are more flexible than necessary to fit the data. The model with the largest evidence is therefore the simplest model that explains the data adequately.

To interpret the Occam factor further, we substitute the Gaussian prior and take the logarithm:

$$\begin{aligned}\log \mathrm{OF} = &-\frac{1}{2}(\boldsymbol{a}^* - \boldsymbol{a}_0)^\top \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\boldsymbol{a}^* - \boldsymbol{a}_0) \\ &-\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{\boldsymbol{a}}|}{|\boldsymbol{\Sigma}_{\boldsymbol{a}^*}|}\end{aligned} \tag{22}$$

The first term penalizes models whose MAP parameters lie far from the prior mean. If the prior encodes physical expectations, as proposed in Section 3.1.1, this penalizes models that fit the data only by using physically implausible parameter values.

The second term involves the ratio of the prior and posterior covariance determinants. The determinant of the parameter covariance matrix measures the volume of the parameter space occupied by the distribution, so this ratio measures how much narrower the posterior is than the prior. If the posterior is much narrower than the prior, the parameters have been finely tuned in order for the model to fit the data. This indicates a fragile model whose predictions are highly sensitive to small perturbations in the parameters, which is penalized accordingly.

A complementary interpretation follows by substituting eq. (18) into the second term of eq. (22) and applying the matrix determinant lemma:

$$-\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{\boldsymbol{a}}|}{|\boldsymbol{\Sigma}_{\boldsymbol{a}^*}|} = -\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{\boldsymbol{q}} + \boldsymbol{J}_{\boldsymbol{a}}\boldsymbol{\Sigma}_{\boldsymbol{a}}\boldsymbol{J}_{\boldsymbol{a}}^\top|}{|\boldsymbol{\Sigma}_{\boldsymbol{q}}|} \tag{23}$$

where $\boldsymbol{J}_{\boldsymbol{a}}\boldsymbol{\Sigma}_{\boldsymbol{a}}\boldsymbol{J}_{\boldsymbol{a}}^\top$ is the covariance of the model predictions induced by the prior uncertainty in the parameters. The numerator therefore measures the volume of output space that the model can access given the prior, while the denominator measures the volume occupied by the data alone. This term therefore penalizes excessively flexible models whose prior uncertainty projects onto a large region of output space relative to the region occupied by the data. Such models are capable of explaining many different datasets, and are penalized for their lack of specificity.

3.3. Uncertainty estimation

As discussed in Section 3.1.2, the likelihood depends on the noise variance σ_q^2, which sets the ex-

pected scale of the discrepancy between the observed data and the model predictions. In practice, this discrepancy reflects not only measurement noise, but also model-form error and any remaining fluctuations not captured by the impulse-response model. We are rarely able to prescribe this variance *a priori*, so we estimate it from the data.

We do this by choosing the value of σ_q^2 that maximizes the marginal likelihood in eq. (21) [23]. In other words, we select the noise level that makes the observed data most probable under the model. For the case $\boldsymbol{\Sigma_q} = \sigma_q^2 \boldsymbol{I}$, this gives:

$$\sigma_q^2 = \frac{1}{M-\gamma}(\boldsymbol{q} - \hat{\boldsymbol{q}}(\boldsymbol{a}^*))^\top (\boldsymbol{q} - \hat{\boldsymbol{q}}(\boldsymbol{a}^*)) \tag{24}$$

$$\gamma = P - \mathrm{Trace}\left(\boldsymbol{\Sigma_{a^*}}\boldsymbol{\Sigma_a^{-1}}\right) \tag{25}$$

where γ is the effective number of parameters that are well-determined by the data, taking values between 0 and P. Equation (24) can therefore be interpreted as the mean squared error (MSE) of the fit, corrected for the number of parameter directions that have been used to explain the data.

This correction is necessary because a flexible model can reduce the MSE by fitting part of the noise. In the limiting case $\gamma \to M$, the model has one effective degree of freedom per data point and can reproduce the data exactly. A naive MSE estimate would then suggest zero noise, conflating an over-parameterized model with noise-free data. By contrast, eq. (24) diverges in this limit, correctly indicating that we cannot infer M effective parameters *and* the noise level from only M data points. This distinction is especially important when working with short signals (small M), which is one of the objectives of this work. A full derivation of eq. (24) is provided in the supplementary material.

In principle, estimating σ_q^2 in this way requires nested optimization: an inner loop computes the MAP parameters for a fixed σ_q^2, and an outer loop updates σ_q^2 to maximize the marginal likelihood. In practice, we avoid this expensive nesting and instead update $\boldsymbol{a}$ and σ_q^2 within a single iterative procedure, converging them simultaneously. We have found this joint update to be robust and substantially cheaper for all cases considered.

3.4. Summary

The Bayesian inference framework described above is summarized at a high level in figure 1, and in more detail in Algorithm 1.

Algorithm 1 Bayesian inference of flame impulse response

```
Input data vectors (u, q)
Select range of model orders N to consider
for each model order N do
    while ||∇_a J|| > ε_tol do
        Use eq. (16) to step towards the MAP estimate a*
        Use eq. (24) to update noise estimate
    end while
    Use eq. (18) to compute Σ_a*
    Use eq. (21) to estimate evidence p(q|N)
end for
Select model order N with highest evidence
```

4. Robust solution of the Bayesian inverse problem

In order to ensure success of the approximate inference framework described above, care must be taken in how the impulse response is parameterized and how the optimization problem is solved. We now describe our choices for these aspects of the problem.

4.1. Parameterization and the prior

A naive parameterization of the impulse response would be to use the raw parameters $\boldsymbol{a} = \{\boldsymbol{n}, \boldsymbol{\tau}, \boldsymbol{\sigma}\}$ directly. However, this has several drawbacks. First, the scales of the parameters can vary widely, leading to ill-conditioning in the optimization problem and poor convergence. A more robust parameterization would therefore normalize the scales of the parameters, keeping them all within a similar range.

The second drawback is that the time delays are symmetric under reordering, meaning that the parameter vector $\boldsymbol{\tau} = [\tau_1, \tau_2]$ describes the same impulse response as $\boldsymbol{\tau} = [\tau_2, \tau_1]$. This leads to a multi-modal posterior pdf, where each mode corresponds to a different ordering of the time delays. Our inference framework will find only one of these peaks, and ascribe all probability to the identified peak. The second requirement is therefore that we impose a specific ordering on the delays.

The third drawback is that the time delays $\boldsymbol{\tau}$ and widths $\boldsymbol{\sigma}$ must be strictly positive to ensure causality and physical realism. Using the raw parameters would therefore require imposing hard constraints on the optimization problem, which can lead to numerical instability and convergence issues, and likely lead to a non-Gaussian posterior. We must therefore find a more robust parameterization that naturally enforces positivity.

To normalize the scales of the parameters, we exploit our knowledge that the characteristic timescales of the system are related to perturbations being convected along the flame. For a flame with length L_f, and bulk velocity u, we define the convective timescale $T_c = L_f/u$. We then normalize the time

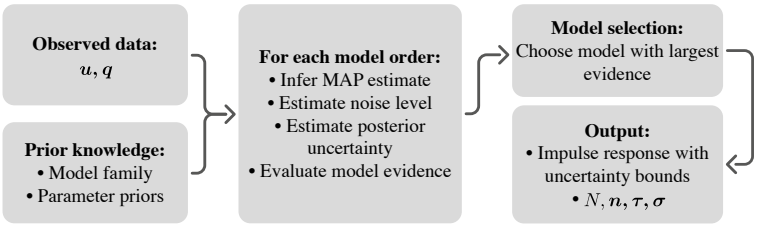


Fig. 1. Schematic overview of the Bayesian inference procedure.

delays and widths by this timescale:

$$\tilde{\tau}_i = \frac{\tau_i}{T_c}, \quad \tilde{\sigma}_i = \frac{\sigma_i}{T_c} \tag{26}$$

This normalization ensures that all the parameters have similar orders of magnitude, improving the conditioning of the optimization problem. Note that if the characteristic timescale is difficult to define, which might be the case for some flames, T_c can be treated as a hyperparameter that is inferred by maximizing the model evidence, as described for the noise variance in Section 3.3.

To remove the symmetry under reordering, we introduce a new set of parameters $\boldsymbol{\alpha}$, defined as the differences between consecutive time delays:

$$\alpha_i = \tilde{\tau}_i - \tilde{\tau}_{i-1}, \quad i = 1, \dots, N \tag{27}$$

where we define $\tilde{\tau}_0 = 0$. If we also constrain $\alpha_i > 0$, then the time delays are automatically ordered as $\tau_1 < \tau_2 < \ldots < \tau_N$. This avoids symmetry under reordering, removing the symmetric peaks in the posterior pdf.

Finally, in order to enforce positivity of the parameters, we introduce a logarithmic transformation for both α_i and σ_i:

$$\gamma_i = \log \alpha_i, \quad i = 1, \dots, N \tag{28}$$

$$\beta_i = \log \tilde{\sigma}_i, \quad i = 1, \dots, N \tag{29}$$

This transformation smoothly maps the positive real line to the entire real line, allowing us to define an unconstrained Gaussian probability density over $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ while keeping the physical parameters $(\boldsymbol{\tau}, \boldsymbol{\sigma})$ strictly positive.

Using these transformed parameters, we can define a broadly applicable Gaussian prior pdf that naturally enforces the physical constraints. While the prior can be adjusted if more specific knowledge about the parameters is available, a choice which should apply to a broad range of flames is:

$$p(n_i) = \mathcal{N}(n_i; 0, 1) \tag{30}$$

$$p(\gamma_i) = \mathcal{N}(\gamma_i; 0, 0.5) \tag{31}$$

$$p(\beta_i) = \mathcal{N}(\beta_i; -1.8, 0.5) \tag{32}$$

To assist with interpretation, the resulting priors over the physical parameters are illustrated in Figure 2. We see that the prior over n encodes an expectation that n_i will lie between ± 3. This is motivated by several previous studies where n has either been fixed to ± 1, or found to be close to ± 1 [15, 16, 18], but provides additional range for flexibility. The Gaussian prior over γ_i results in a log-normal prior over α_i (and equivalently τ_i), which enforces positivity. The chosen width encodes an expectation that the time delays will be spaced between roughly 0.1 and 4 convective timescales, which is in line with observations from several previous studies. Similarly, the Gaussian prior over β_i implies a log-normal prior over the nondimensional width σ_i / T_c. The prior encodes an expectation that the dispersion time scale, $6\sigma_i$, is $\mathcal{O}(T_c)$. We therefore set the prior mean: $\tilde{\sigma}_i \approx 1/6 \approx \exp(-1.8)$. These priors therefore enforce broad but useful physical knowledge about the expected form and range of the parameters, while remaining flexible enough to adapt to a wide range of flames.

4.2. Solution of the optimization problem

Several gradient-based optimizers could be used to minimize $\mathcal{J}$ in equation (15), given the gradients in equation (16). We recognize that the cost function has the form of a nonlinear least-squares problem, which is most efficiently solved using Gauss-Newton-type optimizers, such as the Levenberg-Marquardt (LM) algorithm [24, 25]. This algorithm uses a cheap approximation of the second-order gradients of the cost function to achieve rapid convergence, which happens to be the same approximation as that given in eq. (18).

Importantly, we note that gradient-based optimizers are local optimizers, and the nonlinearity of the forward model means that the cost function may have multiple local minima, even if the posterior is essentially unimodal. This is because we minimize the negative log-posterior, which may have local minima in regions that correspond to negligibly low posterior probability, i.e. there may be significant valleys in the cost landscape, that are not necessarily significant modes in the posterior pdf. To ensure we find the correct MAP point, we use a multi-start approach, where we run the optimizer from several initial conditions sampled randomly from the prior pdf. We then select the solution with the lowest cost as the MAP point, allowing us to robustly find the global minimum. In practice, the random starts required to reliably find the global minimum scales with the number of parameters. We propose 10 random starts per parameter, which works well on the problems considered in this work. The cost of further random starts is moderate, as they can be run in parallel.

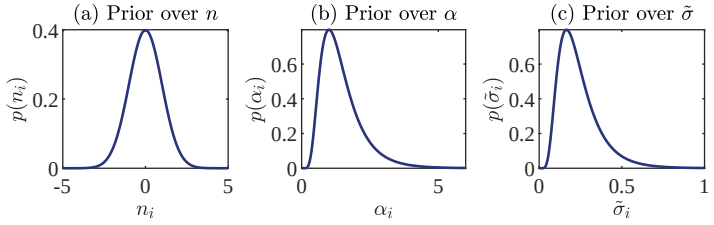


Fig. 2. Prior probability density functions for the impulse response parameters: (a) coefficient n_i, (b) nondimensionalized time delay gap α_i, (c) nondimensionalized width $\tilde{\sigma}_i$.

4.3. Edge effects and the valid convolution region

In Section 1 we introduced the discrete convolution model in eq. (3), which we repeat below for convenience:

$$q_m = \sum_{l=0}^{L-1} u_{m-l} h_l.$$

At the start of a finite record ($m < L-1$), the convolution reaches back to input samples that were not measured (e.g. $u_{-1}, u_{-2}, \ldots$). Using these start-up samples for inference therefore requires an assumption about the unknown input prehistory. A common choice in system identification is to implicitly assume zero prehistory, which can bias the inferred impulse response, particularly when the record is short relative to L.

To avoid this bias, we restrict inference to the valid region ($m \geq L-1$), for which the convolution is fully supported by measured inputs. In practice, this means discarding the first $(L-1)$ output samples and evaluating the likelihood on the remaining ($M_v = M-(L-1)$) samples. In probabilistic terms, this is equivalent to assigning the discarded samples arbitrarily large uncertainty to reflect the impact of the unobserved input prehistory. One could instead treat the unknown input prehistory as an additional uncertain variable and retain the full record, but the expected benefit is small, so we do not pursue this in the current paper.

4.4. Choosing the impulse response support

Choosing the expected length of the impulse response T_h (or equivalently its number of samples L) has two important consequences for the inference problem. First, it determines how much of the measured record can be used because we evaluate the likelihood only on the valid region, as discussed above. Increasing T_h therefore reduces the amount of data from which the impulse response can be inferred.

Second, it controls the effective flexibility of the model. Allowing the impulse response to extend to larger delays gives the model more freedom to explain features in the output, and this increased flexibility should be penalized in model comparison through the Occam factor in the evidence.

We therefore treat the support in one of two ways. If the goal is parameter estimation for a fixed model order, T_h can be prescribed from physical considerations (e.g. an expected maximum convective delay).

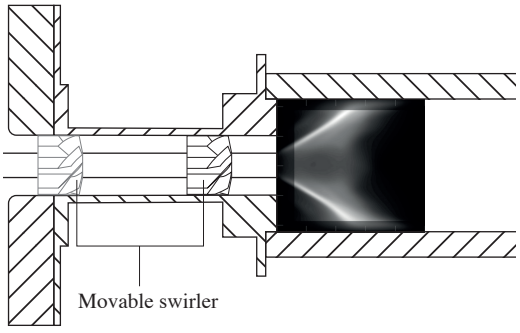


Fig. 3. Schematic of the BRS burner, showing the swirler in the forward position (black) and aft position (grey).

This allows the available data to be used as efficiently as possible. However, if the goal is to rank model orders N using the evidence, then T_h must scale consistently across N and consider all possible impulse responses of order N that are supported by the parameter priors. For a given set of parameter priors, higher-order models admit longer plausible impulse responses. Holding T_h fixed would therefore underestimate the flexibility of higher-order models, and bias the comparison towards them. For model ranking, we therefore set a support $T_h(N)$ using a conservative bound implied by our prior, i.e. $T_h(N)$ is chosen as a conservative upper bound such that 99% of impulse responses drawn from the prior for order N lie within $[0, T_h(N)]$. This produces a fair comparison across N, but requires more data to reliably rank models.

5. Results

We demonstrate the method on an LES dataset from the BRS burner, which is a laboratory-scale combustor with a turbulent, swirl-stabilized flame. The BRS burner has previously been used to demonstrate various methods for inferring flame impulse responses and transfer functions [5, 8, 10, 26]. The rig is illustrated in Figure 3, which shows that the swirler can be moved axially to study the effect on the flame response.

Komarek & Polifke [8] performed compressible RANS simulations of the BRS burner, in which the inlet to the combustor was modelled as a boundary con-

dition rather than resolving the swirler. In doing so, the authors could apply independent axial and circulatory fluctuations to the incoming flow and measure the flame response. By inspection, they found that the impulse response of the flame to velocity perturbations is well described by a sum of three Gaussian pulses. The first has a positive peak and corresponds to the response of the flame to perturbations in axial velocity. The remaining two have one positive and one negative peak with coefficients that sum to zero, and correspond to the response of the flame to circulation fluctuations. This three-Gaussian model was later adopted by Ghani et al. [26] to demonstrate a method for inferring the parameters of the impulse response model from pressure time-series data.

In this study, we use data from Eder et al. [10], who performed both compressible and incompressible LES of the BRS burner at a thermal power of 30 kW with the swirler in the forward position. The inlet was subject to bandwidth-limited broadband forcing, with a cut-off frequency of 630 Hz, and an amplitude of 10% of the bulk velocity. The input-output signals were sampled at a frequency of 1 MHz for 500 ms. The initial 30 ms were discarded to remove simulation transients, and the input-output signals were normalized and down-sampled to 3.3 kHz. A sample of the input-output data is shown in Figure 4.

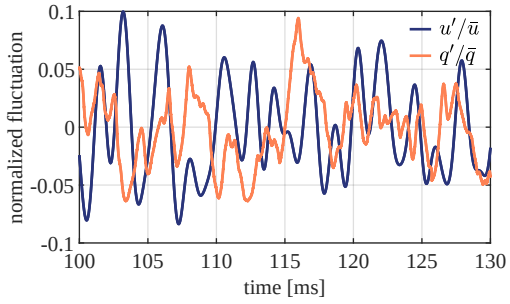


Fig. 4. Example input-output data from LES of a turbulent swirl flame: (blue) velocity perturbation, (orange) heat release rate fluctuation.

Eder et al. [10] applied system identification to the input-output data and found good agreement between the compressible and incompressible simulations, and with experimental data. They used built-in MATLAB function `impulseest` with cubic-spline regularization, signal pre-whitening, and a model order of 39. The appropriate down-sampling and model order need to be hand-tuned to achieve good results, and effectively constrain the length of the impulse response. In this case, a time increment of 0.3 ms and a model order of 39 constrains the impulse response to be 11.7 ms long.

We now apply the Bayesian inference framework described above to the same data, and compare the results with those obtained from system identification, and with experimental data. Unlike previous work, we do not assume a three-Gaussian model for the impulse response *a priori*, but rather use the model selection framework to determine the most likely number of Gaussian pulses given the data. We first assess a baseline case, where we replicate the signal properties used by Eder et al. [10], and then explore the effect of reducing the signal length on both Bayesian inference (BI) and system identification (SI).

5.1. Baseline results

The required inputs for the proposed framework are (i) the input and output signals, (ii) the range of model orders to consider, and (iii) an estimate of the convective timescale T_c. As a baseline, we use the full 470 ms recording, which we downsample to 3.3 kHz, as per Eder et al. [10]. We rank models with $N = 1 - 5$ Gaussian pulses, which bounds the three-Gaussian model that has previously been used for this burner. For the 30 kW operating condition studied, the inlet velocity was 11.3 m/s, which we use as the characteristic velocity scale. The flame length was taken to be the distance from the flame base to the point of impingement on the chamber wall, which was approximately 50 mm. This results in a convective timescale of $T_c = 4.4$ ms.

Like Eder et al., we find that the results from the compressible and incompressible simulations are practically identical, so we only report the incompressible case, as this is more appealing due to the lower computational cost of the simulations.

The three model ranking metrics described in Section 3.2 are shown for each of the five models in Figure 5. Recall that the model with the highest marginal likelihood is the simplest model capable of describing the data, and should be selected as the most likely model. We see that this is the three-Gaussian model, which is consistent with the conclusion that Komarek & Polifke arrived at by applying isolated perturbations at the flame base [8].

The impulse responses inferred by both SI and the most likely model identified using BI with and without a constraint on the low-frequency limit (LFL) are shown in Figure 6. We see that the SI and BI results are similar, but the BI impulse response contains fewer spurious undulations, which are not supported by our physical understanding of the system. These features change with the regularization and sampling parameters used for SI, indicating that they are artefacts of the method rather than robust features of the data. In the BI response, these artefacts are naturally filtered out by the prior information we provide (both in terms of the model structure and the parameter priors). Further, we see that enforcing the low-frequency limit has a negligible effect on the inferred impulse response.

The model parameters identified for this flame are summarized in Table 1. Like Komarek & Polifke [8], we find one positive peak associated with the axial

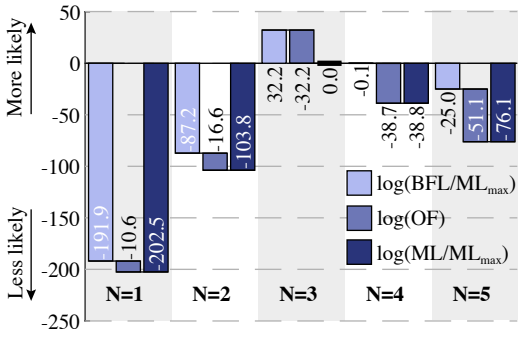


Fig. 5. Model ranking metrics for the baseline case, as described in Section 3.2. The best-fit likelihood (BFL) rewards models that fit the data well, while the Occam factor (OF) penalizes models that are too flexible. The marginal likelihood (ML) balances these two effects, and is used to select the most likely model. The best-fit likelihood and marginal likelihood are shown relative to the marginal likelihood of the three-Gaussian model, which is the most likely model.

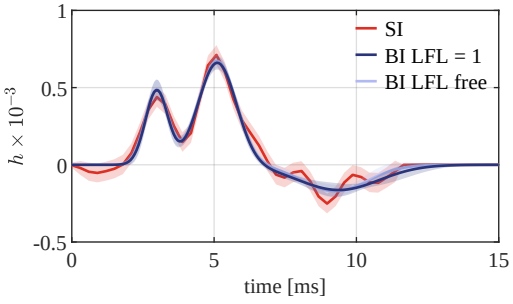


Fig. 6. Impulse responses inferred from LES data using system identification (SI) and Bayesian inference (BI) with and without enforcing the low-frequency limit (LFL).

response, and a positive-negative pair whose coefficients sum to zero, consistent with the flame response to circulatory velocity perturbations. However, unlike Komarek & Polifke, who studied a different operating condition with a thermal power of 70 kW, we find that the positive peak of the circulatory response arrives before the axial perturbation. This is likely due to changes in the relative convective velocities of the two types of perturbations with operating condition, but further analysis is required to fully understand this.

Next, we compare the resulting flame transfer functions (FTFs) in Figure 7. We see that the FTFs obtained from both SI and BI are similar, indicating that the spurious peaks filtered out by BI do not significantly contribute to the FTF. We see that, without enforcing the low-frequency limit, the FTFs from both SI and BI take on a non-physical value slightly above 1, which is easily corrected in the BI framework, but not in the SI framework. Finally, we see that the FTFs from both methods match experimental data well. While this is more of a comment on the quality of the LES than the identification method, it adds confidence that the inferred impulse responses are physically meaningful.

Table 1.
Inferred impulse response parameters: coefficient n_i, nondimensionalized time delay $\tilde{\tau} = \tau/T_c$, and nondimensionalized width $\tilde{\sigma} = \sigma/T_c$. Parameters are shown for the baseline case without enforcing the low-frequency limit. Each parameter value is shown along with the standard deviation of the posterior distribution.

i	n_i	$\tilde{\tau}_i$	$\tilde{\sigma}_i$
1	0.5 ± 0.03	0.61 ± 0.007	0.09 ± 0.008
2	1.1 ± 0.04	1.05 ± 0.006	0.14 ± 0.007
3	-0.5 ± 0.05	1.89 ± 0.026	0.25 ± 0.036

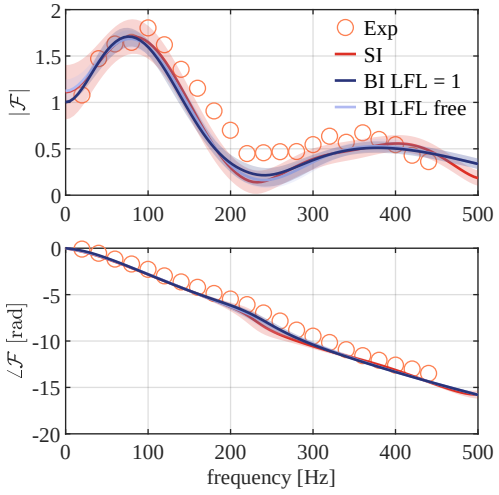


Fig. 7. Flame transfer functions obtained from the impulse responses shown in Figure 6, compared with experimental data.

5.2. *Effect of signal length*

Given the computational cost of reacting LES, there is a strong incentive to minimize the simulation time required to identify the flame response. We therefore assess how reducing the available data length affects both system identification and Bayesian inference.

For this demonstration, we will enforce a fixed support of $T_h = 3.5T_c$ for both SI and BI. As discussed in Section 4, applying a fixed support allows us to use the data more efficiently, but biases model comparison. We therefore assimilate the data into only the

three-Gaussian model rather than performing model ranking like in the baseline case.

Starting from the baseline input-output data, we truncate the valid portion of the signals and retain only the first 20%, 15% and 5%, corresponding to total simulation durations (including the time to flush out the effect of the input prehistory, but excluding the time to flush out simulation transients) of approximately 105 ms, 85 ms and 35 ms, respectively. We then apply SI and BI to these shorter records, downsampling to the same sampling frequency of 3.3 kHz and using the same range of model orders and convective timescale as in the baseline case. We retain the same broad priors over the parameters, as described in Section 4.1, for all cases.

The inferred impulse responses and corresponding flame transfer functions are shown in Figures 8 and 9. As the signal length decreases, we see that the two methods handle the reduction in information differently. SI requires stronger regularization to remain stable, which smooths the impulse response and degrades the FTF. By contrast, BI regularizes through the model structure and parameter priors, preserving temporal resolution and producing more accurate FTFs.

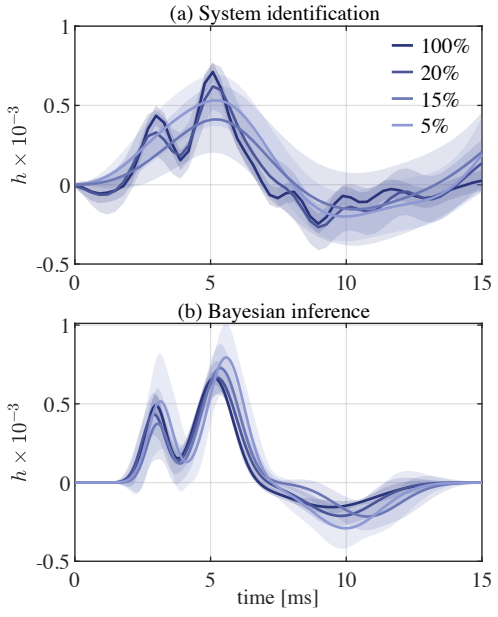


Fig. 8. Impulse responses inferred from truncated LES data using (a) system identification and (b) Bayesian inference with a fixed support of $T_h = 3.5T_c$. We show results from the baseline case (100%), and from cases where the valid portion of the data is truncated to 20%, 15% and 5% of the original length.

For a quantitative comparison, we compute the relative gain and phase errors of the truncated-signal FTFs as $\|x - x_{\text{ref}}\|/\|x_{\text{ref}}\|$, where the full-signal BI result is used as the reference. The results are summarized in Table 2. The errors for SI grow more rapidly than for BI as the signal length decreases. It is worth noting that the phase error for BI remains relatively low, even for the shortest signal, which is important for applications such as stability analysis.

Table 2.
Relative error of the gain and phase of the FTF for the truncated cases, using the full-signal BI results as a reference.

Truncated length	Gain error [%]		Phase error [%]	
	SI	BI	SI	BI
20%	15.3	8.5	6.2	2.2
15%	32.3	18.5	82.5	3.3
5%	40.7	26.1	83.7	7.6

The strong performance of BI in the low-data regime is largely due to the use of physical knowledge to restrict the space of allowable solutions before inference is performed. Rather than searching over arbitrary impulse responses, as is done in SI, the inference is confined to a physically motivated family of distributed time-delay models. This substantially reduces the amount of data required to identify a plausible response. It is useful to contrast this with physics-informed neural networks (PINNs), which are currently a popular approach to combine data with physical information. In many PINN formulations, the unknown is represented by a high-capacity neural network and the physics is imposed through soft penalty terms in the loss function, often with weights that must be chosen by the user. The search therefore remains over a broader class of functions, with the physics guiding the optimization rather than explicitly restricting the admissible solution family. From this perspective, generic PINN formulations would not necessarily be expected to show the same data efficiency as the present approach.

5.3. Assessing the Laplace approximation

In Section 3, we introduced the Laplace approximation, which replaces the posterior parameter distribution by a Gaussian centred at the MAP estimate. Although nonlinearities in the forward model can skew the true posterior, the approximation is expected to be accurate when the posterior is sufficiently concentrated around the MAP point. We assess this here by comparing the Laplace approximation with a posterior estimated using Markov Chain Monte Carlo (MCMC). MCMC makes no assumptions about the shape of the posterior, but is computationally expensive. In this case we use the Metropolis-Hastings algorithm [27], which we run for 500,000 iterations, with a burn-in of 50,000 iterations. We apply the comparison to the three truncated signals introduced above so that we can examine how the accuracy of the

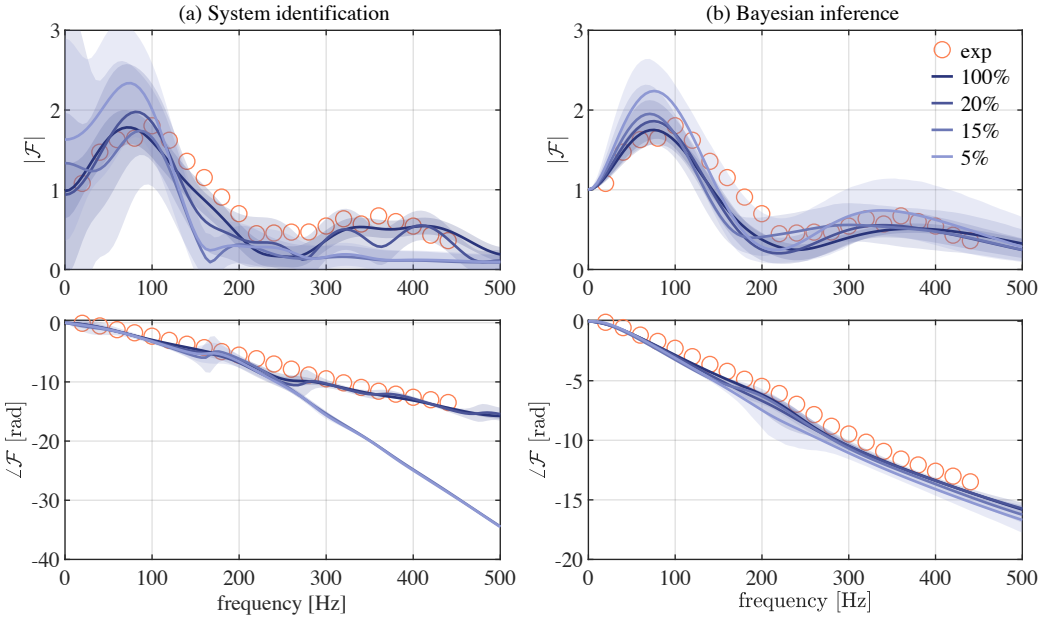


Fig. 9. Flame transfer functions obtained from the impulse responses shown in Figure 8, compared with experimental data.

Laplace approximation changes as the amount of data is reduced.

Figure 10 compares the *marginal* posterior distributions obtained with MCMC and the Laplace approximation (here, "marginal" refers to the distribution of each parameter after integrating out all other parameters). For clarity, each marginal distribution is normalized so that we can compare the *shape* of the posterior across the different data lengths. For the shortest signal (5% of the original valid record), the MCMC marginals exhibit moderate skewness, which is not captured by the Laplace approximation. Importantly, this discrepancy primarily affects the uncertainty quantification, while the MAP values remain well predicted. We see that the 20% and 100% signals are effectively identical, and the Laplace approximation tracks the MCMC results closely. This supports the claim that, once a moderate amount of information is available, the posterior concentrates around the MAP in a sufficiently small region for the second order Taylor expansion underpinning the Laplace approximation to be accurate.

To provide a more complete view, Figure 11 shows the joint posterior for the 15% case. Off-diagonal panels show pairwise joint distributions, while the diagonal panels reproduce the corresponding marginal distributions. As above, the Laplace approximation provides an excellent match to the MCMC posterior, and also reproduces the parameter correlations. The observed correlations are relatively weak in this case, suggesting that the parameters of the three-Gaussian model are uniquely identifiable from the available input-output data at this signal length.

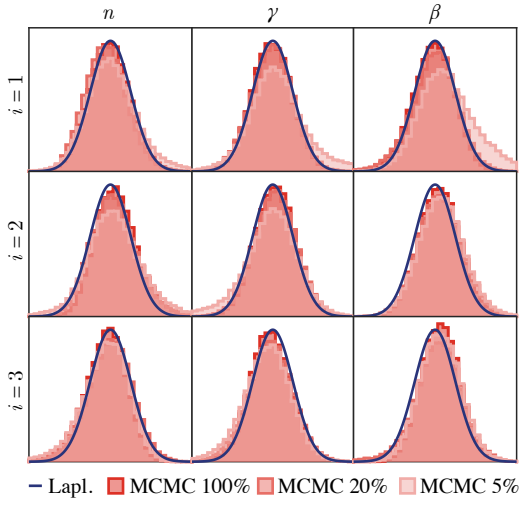


Fig. 10. Marginal posterior distributions of the model parameters obtained from MCMC sampling and the Laplace approximation, for the 100%, 20%, and 5% signals. Each marginal distribution is normalized to facilitate comparison of the shape of the posterior across data lengths.

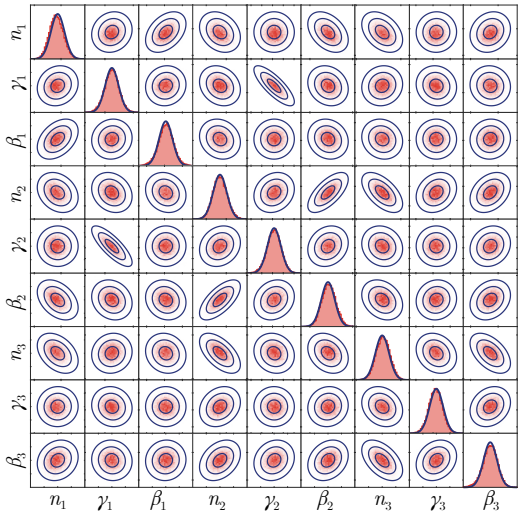


Fig. 11. Joint posterior distribution of the model parameters obtained from MCMC sampling (red) and the Laplace approximation (blue), for the 15% truncated signal. Off-diagonal panels show pairwise joint distributions, while diagonal panels show the corresponding marginals. The joint distributions are shown as heatmaps for MCMC and as ellipses of 1-3 standard deviations for the Laplace approximation.

6. Conclusions

This work reformulates the inverse problem of identifying a flame impulse response from recordings of velocity and heat release rate within a Bayesian framework. The impulse response is constrained to a physically motivated distributed time delay model represented as a sum of N Gaussian pulses. For a given N, the parameters of the pulses are inferred from the data with quantified uncertainty bounds using Bayesian parameter inference. This is repeated for several values of N, and the most likely model order is selected using Bayesian model comparison.

This formulation replaces ad hoc choices in standard system identification, such as the model order, sampling rate, and regularization model and strength. The Bayesian framework instead provides a principled trade-off between data fit and model complexity, while producing impulse responses that are directly interpretable in terms of convective delays and dispersive broadening.

On broadband-forced LES of the BRS swirl burner, this framework selected a three-Gaussian impulse response, which is consistent with prior physical interpretations of this configuration. Relative to standard system identification, the Bayesian approach produced an impulse response with fewer spurious features, while giving a comparable flame transfer function over the forced frequency range. It also enables straightforward enforcement of a known low-frequency limit, which is not easily achieved with standard system identification.

When the available record length is reduced, performance degrades mainly through increased posterior uncertainty rather than a qualitative change in the inferred impulse response. By contrast, standard system identification requires excessive regularization in the low-data regime, resulting in significant loss of temporal resolution in the impulse response.

Several extensions of the current approach are natural. The present likelihood uses a white Gaussian noise model on the output only. Incorporating temporally correlated discrepancy and/or uncertainty in the measured input would broaden applicability. More generally, applying the method across operating conditions would allow the inferred delay parameters to be related more directly to underlying flame-flow physics and to reveal systematic, uncertainty-aware trends. Similarly, applying the method to a range of forcing amplitudes would allow flame describing functions to be constructed more cheaply.

The software implementing the proposed framework is available at `https://github.com/mattyoko/bayesian-flame-impulse-response/tree/paper-released`. The repository includes scripts to reproduce the figures and analyses reported in this paper. We hope that making the code available will facilitate application of the method to other flames and support systematic studies of how the inferred impulse response parameters relate to underlying flame-flow physics.

CrediT authorship contribution statement

M.Y.: Conceptualization, Formal analysis, Methodology, Software, Writing - original draft.
W.P.: Conceptualization, Writing - review & editing.

Declaration of competing interest

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

Acknowledgments

The authors thank Korbinian Niebler for preparing and transferring the LES data, and Alexander Eder who performed the simulations. This work was supported by the Leverhulme Trust [grant number ECF-2025-609].

Supplementary material

The supplementary material contains additional details on the derivation of the noise variance estimate.

References

[1] W. Polifke, Modeling and analysis of premixed flame dynamics by means of distributed time delays, Progress in Energy and Combustion Science 79 (2020).

[2] T. Schuller, T. Poinsot, S. Candel, Dynamics and control of premixed combustion systems based on flame transfer and describing functions, Journal of Fluid Mechanics 894 (2020).

[3] W. Polifke, A. Poncet, C. O. Paschereit, K. Döbbeling, Reconstruction of acoustic transfer matrices by instationary computational fluid dynamics, Journal of Sound and Vibration 245 (3) (2001) 483–510.

[4] W. Polifke, Black-box system identification for reduced order model construction, Annals of Nuclear Energy 67 (2014) 109–128.

[5] S. Guo, C. F. Silva, A. Ghani, W. Polifke, Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis, Journal of Engineering for Gas Turbines and Power 141 (February) (2019) 1–10.

[6] J. F. Radack, B. Dharmaputra, B. Schuermans, N. Noiray, Quantifying uncertainties in the input – output identification of Flame Transfer Functions, Combustion and Flame 281 (August) (2025) 114398.

[7] A. Giauque, T. Poinsot, F. Nicoud, Validation of a flame tranfer function reconstruction method for complex turbulent configurations, 14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference) (May) (2008) 5–7.

[8] T. Komarek, W. Polifke, Impact of swirl fluctuations on the flame response of a perfectly premixed swirl burner, Journal of Engineering for Gas Turbines and Power 132 (6) (2010) 1–7.

[9] L. Tay-Wo-Chong, W. Polifke, Large eddy simulation-based study of the influence of thermal boundary condition and combustor confinement on premix flame transfer functions, Journal of Engineering for Gas Turbines and Power 135 (2) (2013) 1–9.

[10] A. J. Eder, C. F. Silva, M. Haeringer, J. Kuhlmann, W. Polifke, Incompressible versus compressible large eddy simulation for the identification of premixed flame dynamics, International Journal of Spray and Combustion Dynamics 15 (1) (2023) 16–32.

[11] T. Wang, M. Guo, W. Han, J. Li, On the effect of injector temperature on the flame transfer functions of laminar premixed flames, International Journal of Hydrogen Energy 194 (November) (2025) 152371.

[12] E. Ke, C. Ji, M. Wang, D. Pan, T. Zhu, Prediction of flame transfer function and combustion instability on a partially premixed swirling combustor by the system identification and CFD methods, Aerospace Science and Technology 151 (June) (2024) 109275.

[13] M. Bothien, D. Lauper, Y. Yang, A. Scarpato, Reconstruction and Analysis of the Acoustic Transfer Matrix of a Reheat Flame From Large- Eddy Simulations, Journal of Engineering for Gas Turbines and Power 141 (February) (2019) 1–9.

[14] W. Polifke, C. Lawn, On the low-frequency limit of flame transfer functions, Combustion and Flame 151 (3) (2007) 437–451.

[15] B. Schuermans, V. Bellucci, F. Guethe, F. Meili, P. Flohr, C. O. Paschereit, A detailed analysis of thermoacoustic interaction mechanisms in a turbulent premixed flame, Proceedings of the ASME Turbo Expo 2004 1 (January 2014) (2004) 539–551.

[16] S. Bade, M. Wagner, C. Hirsch, T. Sattelmayer, B. Schuermans, Design for thermo-acoustic stability: Modeling of burner and flame dynamics, Proceedings of the ASME Turbo Expo 1 B (November) (2013) 1–7.

[17] E. Æsøy, J. G. Aguilar, S. Wiseman, M. R. Bothien, N. A. Worth, J. R. Dawson, Scaling and prediction of transfer functions in lean premixed, Combustion and Flame 215 (2020) 269–282.

[18] A. Blondé, B. Schuermans, K. Pandey, N. Noiray, Effect of Hydrogen Enrichment on Transfer Matrices of Fully and Technically Premixed Swirled Flames, Journal of Engineering for Gas Turbines and Power 145 (12) (2023) 1–11.

[19] K. Moon, R. Martin, B. Schuermans, N. Noiray, Transfer functions of lean fully- and technically-premixed jet-stabilized turbulent hydrogen flames, Proceedings of the Combustion Institute 40 (1-4) (2024) 105256.

[20] D. J. C. MacKay, Information Theory, Inference, and Learning Algorithms, Cambridge University Press, 2003.

[21] E. Jaynes, Probability theory: the logic of science, Cambridge University Press, 2003.

[22] A. Kontogiannis, S. V. Elgersma, A. J. Sederman, M. P. Juniper, Bayesian inverse Navier-Stokes problems: joint flow field reconstruction and parameter learning, Inverse Problems 41 (1) (2025).

[23] D. J. MacKay, Comparison of approximate methods for handling hyperparameters, Neural Computation 11 (5) (1999) 1035–1068.

[24] J. J. Moré, D. C. Sorensen, Computing a trust region step, SIAM J. Sci. Stat. Comput. (December) (1981).

[25] H. B. Nielsen, Damping parameter in Marquardt's method, IMM, 1999.

[26] A. Ghani, A. Albayrak, From Pressure Time Series Data to Flame Transfer Functions: A Framework for Perfectly Premixed Swirling Flames, Journal of Engineering for Gas Turbines and Power 145 (1) (2023) 1–9.

[27] W. K. Hastings, Monte Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika 57 (1) (1970) 97–109.