Using shock tube species time-histories in Bayesian parameter estimation: effective independent-data number and target selection

Huaibo Chen[a], Weiqi Ji[a], Séan J. Cassady[b], Alison M. Ferris[b], Ronald K. Hanson[b], Sili Deng[a,*]

[a] *Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b] *High Temperature Gasdynamics Laboratory, Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA*

Abstract

Species time-histories in shock tube experiments provide rich kinetic information for parameter estimation, but there are two problems in using these data in Bayesian approaches. First, the effective independent-data number is not equal to the number of data points in a curve, so brute multiplication of all data points in likelihood function can weaken the constraints from prior information. Second, taking all points of a curve as targets can lead to results different from that of taking several representative points in the curve. In this paper, we employed maximum a posteriori estimation combined with a neural network response surface to optimize a propane mechanism against multispecies time-histories of propane pyrolysis in a shock tube. Three methods of calculating the likelihood function are used: multiplying all points in a curve (C-160), taking the averaged likelihood in each point (C-1), and taking the likelihood of last points (LastP). The influence of effective independent-data number was studied by comparing C-1 and C-160. It was found that C-160 performed slightly better in fitting experimental data, but brute multiplication overtuned the rate constants beyond a reasonable range. The larger the effective independent-data number, the more severe the overtuning, leading to only a slight improvement of model predictions. The influence of target selection is investigated by comparing LastP and C-1. LastP outperformed C-1 slightly, which can be attributed to the fact that larger discrepancies observed between experimental data and model predictions of the last point can increase the weights of likelihood functions. This further implies that several critical points can represent the entire line for point estimation. This paper can provide a reference both for modelers about reasonable utilization of species time-histories, and for experimentalists about the importance of a detailed probability distribution of measurement error, as well as experiment design with emphasis on critical points.

Keywords: Optimization; Bayesian approach; Species time-histories; Uncertainty quantification; Neural network

*Corresponding author: silideng@mit.edu

1. Introduction

Although direct measurement and quantum computation have been widely used to determine reaction rate constants in kinetic models, many rate parameters are still estimated from a system level against experimental measurements, such as laminar flame speeds, ignition delay times, etc. In 2009, Davidson and Hanson have made the multispecies time-histories data available [1], which can provide rich kinetic information and hence are highly suitable for parameter estimation and model optimization.

Many tools have been developed for the determination and optimization of kinetic parameters. Frenklach and his coworkers [2] developed the solution mapping method, in which a weighted-least-squares objective function is minimized. Sheen and Wang [3] developed the Method of Uncertainty Minimization using Polynomial Chaos Expansions (MUM-PCE), which can propagate the uncertainty of experiments to the that of rate parameters with relatively small computational costs. Turányi et al. [4] proposed a new optimization method to integrate both indirect and direct measurements. Recently, the Bayesian approach has attracted much attention in combustion community [5, 6], which provides a probabilistic approach to quantify the uncertainty of parameters from both prior knowledge and experimental data. However, the computational cost of it (usually involves applying the Markov chain Monte Carlo algorithm) is generally large, so the application of Bayesian approach usually relies on creating response surfaces, which map the model parameters to predictions. It is noteworthy that the MUM-PCE is also a special case of Bayesian approach with simplifications [5].

In this work, we focus on two problems in the utilization of species time-histories in shock tubes for kinetic model optimization by Bayesian approaches. The first problem lies in the target selection: representative points on the species time-history curve [7, 8] or the entire curve [9, 10]. Both methods have been widely adopted by modelers, and yet, there is no comparison of the approaches in the literature on their influences on kinetic parameter estimation. The second problem is how to balance the information from the prior knowledge and new measurements. Specifically, how to determine the weights of the likelihood of different data points, indicated by the number of effective independent data adopted in the Bayesian approach.

Therefore, we conducted kinetic model optimization by Bayesian approach facilitated by adopting neural networks as response surfaces. For demonstration, we will not fully solve the Bayesian equation but obtain the optimal parameters by maximum a posterior (MAP) estimation. The elucidated influences of the target selection and number of independent effective data will guide experimentalists and modelers in generating and adopting data for model development.

2. Methodology

To investigate the influence of target selection on kinetic model optimization, we combined neural networks and maximum a posteriori (MAP) estimation to optimize a propane pyrolysis mechanism against species histories in shock tube measurements [11]. The trial mechanism is a C_3 sub-mechanism extracted from USC-Mech II [12] with 111 elementary reactions and 27 species; Cantera's format of the sub-mechanism is available in the Supplementary Materials. In the shock tube experiments previously reported in [11], the time-histories of mole fraction for eight species across five initial temperatures were reported. There are 160 data points in each profile, from 0.01 ms to 1.6 ms with an interval of 10 μs. The initial gas composition is 2% propane in argon. The thermodynamic conditions are initial temperatures of 1250 K, 1290 K, 1330 K, 1370 K, and 1410 K, respectively, at a constant pressure of 4 atm. The average experimental uncertainties (one standard deviation, σ) of mole fraction for each species are 0.0027 for H_2, 0.0012 for C_2H_2, 0.0011 for CH_4, 0.0013 for C_2H_4, 0.0012 for C_2H_6, 0.0011 for pC_3H_4, 0.0016 for C_3H_6, and 0.0015 for C_3H_8.

The optimization process includes two steps: first, a neural network response surface model representing the response of species evolution to the perturbation of pre-exponential factors was trained; then, the Bayesian approach was employed with the previous neural networks to optimize the kinetic parameters.

2.1 Bayesian approach for mechanism optimization

Bayesian approaches has been extensively adopted to optimize kinetic mechanisms against experimental data [5, 6]. The core is the Bayes' theorem,

$$f(k \mid d) = \frac{f(d|k)g(k)}{\int_K f(d|k')g(k')dk'}, \quad (1)$$

where $f(d|k)$ is the likelihood of d calculated by the model with parameter k, $g(k)$ is the prior distribution of the model parameter based on previous experience and knowledge, and K is the domain of $g(k)$. The optimal model parameter k^* based on the observed value d is the k that can make $f(k|d)$, the posterior distribution, reach its maximum value. In previous works [5, 6], numerical sampling and integration algorithms, such as Markov chain Monte Carlo (MCMC), were often utilized to solve the posterior distribution and obtain the optimal parameters. To focus on the discussion of the influence of calibration target selection rather than conduct a comprehensive uncertainty quantification, we only focused on finding the optimal parameters corresponding to the maximum value of the posterior distribution. Since the

denominator in Eq. (1) is independent of k, we only need to maximize $f(d|k)g(k)$ and the corresponding k is the optimal value. This approach is called maximum a posteriori (MAP) estimation [13].

For demonstration, we focused on the pre-exponential factors in each elementary reaction and left the other parameters in the kinetic mechanism, such as temperature exponents and activation energies, unchanged. The prior distributions of pre-exponential factors k_i are parameterized as [3]:

$$x_j = \frac{\ln (k_j/k_{j,0})}{\ln f_j}, \quad (2)$$

where $k_{j,0}$ is the nominal (unperturbed) value in the mechanism, of the pre-exponential factor for jth elementary reaction, f_j is the uncertainty factors for jth elementary reaction [3], and x_j is subject to the normal distribution with an expectation of 0 and a variance of 0.25. Generally, it is assumed that the prior distribution of k_j is mutually independent [3]. As for the $f(d|k)$, namely the likelihood function, we assumed that it is a normal distribution with an expectation of c_0 and a variance of ε^2, where c_0 is the species concentration at a given time calculated from the model with parameter k, and ε is the experimental uncertainty of species concentration.

Due to the independent prior distribution, $g(k)$ for the entire kinetic model can be evaluated as $g(\boldsymbol{k}) = \prod g(k)$, since they are independent of each other. As for $f(d|\boldsymbol{k})$ with multiple experimental data, a common approach [5, 6] is to multiply $f(d|\boldsymbol{k})$ for each data point together, i.e., $f(\boldsymbol{d}|\boldsymbol{k}) = \prod f(d|\boldsymbol{k})$. To facilitate optimization, we constructed a loss function by taking the natural logarithm of $f(\boldsymbol{d}|\boldsymbol{k})g(\boldsymbol{k})$, cancelling the constant term, and taking the negative of the result:

$$Loss = \sum_S \left\{ \frac{1}{2\varepsilon_i^2} \sum_T \left[\frac{n}{N} \sum_N (c_{pred} - c_{measure})^2 \right] \right\} + \sum_j 2x_j^2 \quad (3)$$

Here, $N = 160$ is the number of data points in a curve, T is the number of temperatures, S is the number of species, and c_{pred} and $c_{measure}$ are the mole fractions predicted by the model and of shock tube measurement, respectively. To reduce the computational cost, c_{pred} is calculated by neural networks. ε_i is the experimental uncertainty for species i, and j is the number of pre-exponential factors to be tuned. x_j is defined by Eq. (2). n is the number of effective independent-data points, which should be the same as N, if we take the common approach to calculate $f(d|k)$ with multiple experimental data, i.e., multiplying them together. However, such a treatment relies on an assumption that the experimental uncertainty of each data point is mutually independent. This assumption is theoretically and practically problematic when time-histories of shock tubes are used as targets, and we will discuss this in detail in Sec. 3.2. Another approach is to assume the effective number of independent data points is 1 in each species profile, similar to the approach described in [4], which averages the discrepancies of all data points in one profile. In the current study, optimizations with both $n = 1$ and $n = 160$ are conducted and compared. The setting of effective independent-data number is adopted from [14]. In fact, moving n into the denominator part as N/n, and merging with ε_i, different choice of n can be interpreted as different experimental uncertainties.

For target selection, it is also a common practice to select one or several points along the time-histories of shock tube species measurements as optimization targets. To demonstrate this approach, the last points (at 1.6 ms) in every profile were selected as targets. The loss function is

$$Loss = \sum_S \left\{ \frac{1}{2\varepsilon_i^2} \sum_T (c_{pred,1.6} - c_{measure,1.6})^2 \right\} + \sum_j 2x_j^2 \quad (4)$$

Similarly, here we assumed that the likelihood functions for each temperature and species are mutually independent.

Due to the nonlinear response of the species profiles and the loss function to the kinetic parameters, convex optimization algorithms are usually inefficient and suffer from local minima traps. Instead, we adopted stochastic gradient descent (SGD) algorithm, developed to optimize large-scale deep neural networks, to find the minimum value of loss function and the optimal model parameters.

$$k^* = \arg\min_k Loss \quad (5)$$

Since we have replaced the physical model with neural networks, the gradients information used in the optimization process can be efficiently computed by backpropagation of neural networks, rather than solving the computationally expensive sensitivity equations. The iteration times of all three optimization strategies are 200, where the curve of the loss function enters a near-plateau region. The figures of loss function versus iteration times in three cases are shown in Sec. 1 of the Supplementary Materials.

2.2 Neural networks as response surfaces

Due to the high computational cost of Bayesian approaches in mechanism optimization, its implementation usually relies on response surfaces, or surrogate models, which can relate model parameters to the model prediction. Response surfaces that are commonly used in the combustion community include polynomial chaos expansion [3] and neural networks [6]. Here, we trained neural networks as response

surfaces. The difference between our approach and previous works on neural networks as surrogate models is that our neural networks can represent the mapping relationship between a group of pre-exponential factors and multiple output data, namely time-histories of eight species, instead of only one output data, such as the ignition delay times in [7]. In other words, our neural network model is a full-field surrogate model, rather than simply predicting a scalar output value. Another advantage of neural networks is that using gradient descent algorithm in MAP requires gradient information, and the backpropagation algorithm of neural networks can highly reduce the cost of gradient computation.

The network structure, as shown in Fig. 1, is inspired by the recently developed Deep Operator Neural Network [15] that has a parameter neural network and a coordinate neural network. The number of neurons is labeled under each layer. The number of neurons and layers is obtained by grid searching, i.e., increasing the number until its performance does not improve. For the parameter network, the input features are $\ln(k_i/k_{i,0})$ for 111 elementary reactions. For the coordinate (time) network, the input is time. The outputs of the two networks undergo reshaping and multiplication, and then form the prediction of species concentration at a given time, as shown in Fig. 1. It is worth noting that we trained different neural networks under different initial temperatures, so temperature is not considered as an input. The data used to train a certain network are all under the same initial temperature. The loss function for training is

$$Loss = \sum(c_{pred} - c_{train})^2, \qquad (6)$$

where the summation is over all data points at a certain temperature, including the profiles for eight species. Adam algorithm [16] is used to minimize the loss function till convergence. The initial point of optimization is the parameter values in the original mechanism, i.e., $\ln(k_i/k_{i,0})$=0.

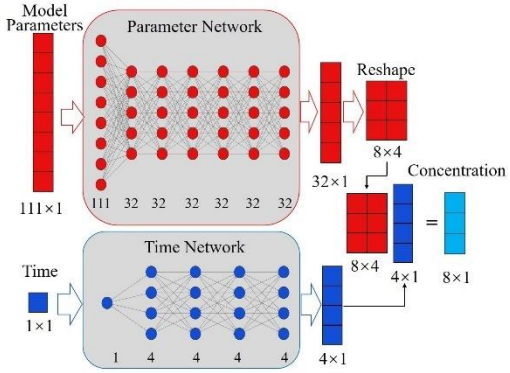

Fig.1. The structure of the neural network used as surrogate models. The number of neurons is labeled under each layer.

For each network, the training data are 5000 samples with different model parameters generated by Cantera [17] according to the experimental conditions using the C_3 sub-mechanism. The input parameters, $\ln(k_j/k_{j,0})$, are randomly sampled by Latin hypercube sampling [18] with bounds of $\pm\ln f_j$. The results show that neural networks can predict the species concentrations and the gradients of concentration to input parameters accurately. The validation results can be found in Sec. 2 of the Supplementary Materials, from which we can find that the prediction error is relatively small, compared with the experimental uncertainty shown in the shadow region in Sec. 4 of the Supplementary Materials. We also validate the gradient computation of the network, as shown in Sec. 7 of Supplementary Materials.

3 Results and discussion

3.1 Comparison of optimized mechanisms with the original mechanism

Three approaches were utilized to optimize the trial mechanism: the entire curve as target with averaged errors (n = 1) and point-wise errors (n = 160), and the last point in each profile as target. To compare the performance of the three optimized mechanisms against measurements, we defined a normalized error (NE) for a given species profile:

$$\mathrm{NE} = \frac{1}{N\varepsilon_i}\sum_N |c_{pred} - c_{measure}|, \qquad (7)$$

where N = 160 is the number of data points in each profile and ε_i is the experimental uncertainty for species i. For each mechanism, we have 40 profiles (eight species at five temperatures) to compare with, so we can calculate the average of 40 NEs as an overall measurement of the performance of this approach. In addition, to assess the closeness between the prediction and the measurement only at the last point (at 1.6 ms), we can also define the normalized error at the last point (NELP):

$$\mathrm{NELP} = \frac{1}{\varepsilon_i}|c_{pred,1.6} - c_{measure,1.6}|. \qquad (8)$$

We then averaged 40 NELPs together for each mechanism. The averaged NEs and NELPs of three optimized mechanisms and the initial mechanism are shown in Table 1. In Table 1 and hereafter, the mechanisms optimized against entire curves with n = 160 and n = 1 are labeled as C-160 and C-1, respectively, and the mechanism optimized against the last point is abbreviated as LastP. From Table 1, we can see significantly improved agreement with the measurements with all three optimization approaches.

Detailed comparisons for individual temperature and specific species measurements between the NEs of the original mechanism and those of the three optimized mechanism are included in Sec. 3 of the Supplementary Materials. Regardless of the number and choice of the target, the optimized mechanism almost always has reduced NEs except for C_2H_2 and pC_3H_4 under certain temperatures. A possible explanation is that the relative experimental uncertainties of these two species are very large. They are nearly ten times larger than the species concentrations themselves at 1250 K. Therefore, the contribution of these species to the loss function is very small, so in the optimization process, the mechanism is tuned to fit other species at the cost of worse predictions for these two species.

Table 1
Comparisons of the averaged normalized errors (NEs) and normalized errors at the last points (NELPs).

Mechanisms	Averaged NE	Averaged NELP
Original	0.754	0.876
C-160	0.168	0.210
C-1	0.210	0.296
LastP	0.194	0.227

3.2 The influence of effective independent-data number

The influence of effective independent-data number will be elucidated via comparison between the performance and optimization in model parameters of C-160 and C-1. From Table 1, we can see that C-160 has both smaller averaged NE and averaged NELP, implying that it performs better compared with C-1 in terms of the closeness to experimental measurements, although the difference is slight. In Table 2, the differences between the NEs of C-1 and those of C-160 for all 40 cases are shown, where the number indicates the ratio of the difference to the corresponding experimental uncertainty. In most cases, the differences between NEs of C-160 and C-1 are less than 10% of the experimental uncertainty, with C-160 slightly outperforming C-1. However, for pC_3H_4 at 1410 K, C-1 is closer to the measurements compared to C-160. As is mentioned above, the relative experimental uncertainty of pC_3H_4 is so large that the optimizer tends to prioritize fitting other species, leading to larger errors in pC_3H_4. The complete curves of measurements with shadows indicating the uncertainties and predictions with C-160 and C-1 are summarized in Sec. 4 of the Supplementary Materials for visualization of this discussion.

The reason why the overall performance of C-160 is better than C-1 is that $f(\boldsymbol{d}|\boldsymbol{k})$ in the loss function is given a higher weight, so the prior distribution only plays a minor role in informing the posterior distribution. The optimizer just needs to decrease the value of $f(\boldsymbol{d}|\boldsymbol{k})$ to make the prediction closer to experimental measurements. On the contrary, the C-1 approach needs to keep a balance between $f(\boldsymbol{d}|\boldsymbol{k})$ and $g(\boldsymbol{k})$, which will prevent rate parameters from deviating far away from the nominal values.

To compare the difference in rate constants of C-160 and C-1, 10 most tuned elementary reactions (measured by the absolute value of x_j) in each mechanism and the corresponding x_j are shown in Fig. 2. x_j can be interpreted as the normalized change of parameters, such that an increment of 0.5 in value corresponds to one standard deviation of the experimental uncertainty. Although the performances of C-1 and C-160 are similar according to Table 1, their optimized rate constants are quite different. Although a significant tuning in C-160 resulted in better performance compared to C-1, it is noted that more than 10 constants are tuned over 1σ, and one tuned even over 3σ. This is generally considered to be highly improbable in model development.

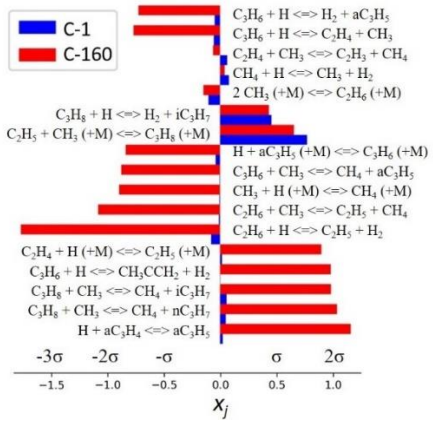

Fig. 2. The x_j of the most tuned reaction rates in C-1 and C-160.

Table 2
Differences between the NEs of C-1 and those of C-160. Differences larger than 10% of the corresponding experimental uncertainties are bolded.

	1250 K	1290 K	1330 K	1370 K	1410 K
H_2	-0.018	-0.003	0.026	0.072	0.086
C_2H_2	0.001	-8E-5	-0.002	-0.024	-0.067
CH_4	**0.164**	**0.221**	**0.170**	**0.156**	0.080
C_2H_4	0.077	0.089	0.025	0.012	**0.242**
C_2H_6	-0.021	0.007	0.072	**0.172**	**0.296**
pC_3H_4	-3E-4	-0.007	-0.031	-0.070	**-0.118**
C_3H_6	0.008	0.009	-0.057	-0.056	-0.021
C_3H_8	0.081	**0.143**	0.058	-0.071	-0.013

To further illustrate the influence of the number of effective independent data on the tuning of kinetic parameters, optimizations using a sweep of n between 1 and 160 were conducted, similar to the process that resulted in C-1 and C-160. The largest change and the magnitude of all changes to the parameters are presented as max $|x_j|$ and $||\boldsymbol{x}||$, respectively. As shown in Table 3, when n increases, max $|x_j|$ increases accordingly when n is beyond 10. Furthermore, the max $|x_j|$ reaches 2σ when n is near 30, while it reaches

3σ when *n* is near 80. This demonstrates that although including more targets could improve model performance, over-tuning might become an issue.

A physical interpretation of cause for the over-tuning problem is that the two terms in the loss function in Eqs. (4) and (5) measure the closeness of prediction to the experimental data and the closeness of tuned rate parameters to the nominal value, respectively. If we focus on tuning rate parameters to fit experimental data, the likelihood function is given too much weight. This approach is then similar to maximum likelihood estimation (MLE) [13], which totally ignores the prior distribution, instead of MAP. Theoretically, the justification for multiplying the likelihood function in each point together to construct an overall likelihood function is based on the assumption of independent likelihoods, which means that the difference between each pair of measurement and model prediction is independent. As mentioned in the Introduction, this is generally adopted by previous studies on Bayesian model calibration. However, this is probably not true for the species histories in shock tube measurements.

Table 3
Influence of the number of effective independent data on x_j. Maximum $|x_j|$ near 2σ and 3σ are in bold.

n	1	10	30	50	80	160
max $\|x_j\|$	0.77	0.76	**0.96**	1.24	**1.49**	1.77
$\|\|\boldsymbol{x}\|\|$	0.92	1.31	2.17	2.69	3.20	3.96

A potential explanation of this problem is that this approach folds systematic error and model error into i.i.d. Gaussian distributions, so parameters are forced to be tuned away from true value to offset the systematic error and model error. Following [19, 20], we can express the discrepancies between model predictions and measurements at data point *i* as:

$$y_i = f_i(\lambda) + \varepsilon_{mi} + \varepsilon_{di} = g_i + \varepsilon_{di} \qquad (9)$$

where y_i is an experimental measurement, $f_i(\lambda)$ is the model prediction given parameter λ, ε_{mi} is the model error, ε_{di} is the measurement error, and $g_i$is the true process. In fact, ε_{di} can be further decomposed into ε_{si} and ε_{ri}, which are systematic error and random noise, respectively. Model error is defined as the discrepancy between model predictions (with true parameters) and ground truth [20]. In traditional Bayesian inference, we assume that the model with true parameters can predict the true process, and measurement data is ground truth plus random noise. Generally, modeling the random noise as i.i.d. Gaussian distributions is a valid assumption, so in this scenario, as more data is used, the posterior distribution would shrink around the true parameter values [19]. In the current case, however, we also lump systematic error and model error into an i.i.d. Gaussian distribution, which is intrinsically a mis-specification. For model error, it is a deterministic variable, rather than a random variable. The common sources of model error in kinetic model include: (a) missing species/reactions; (b) deviation of untuned variable, like our case, where b-factors and activation energies are not calibrated; (c) lumping of several reactions, which is common in reduced mechanisms. For systematic error, it may be subject to a distribution, but not necessarily an i.i.d. Gaussian distribution. One example is that, if the real temperature behind reflected shock wave deviates from the theoretical value, the error would be present throughout the entire time domain, not independent for each point. In fact, as pointed by [14], even noise can be correlated. Consequently, in the calibration process, in order to fit the experimental data, the parameters would be overtuned to offset the systematic error and model error.

Another potential reason is the ill-posedness of this inverse problem [21], which means that the inference results are very sensitive to the data used for inference. In our case, the inferred parameters of C-160 and C-1 are pretty different, but the predictions of two groups of parameters are very similar. Thus, small changes of experimental data would lead to large changes of inference results. One way to solve ill-posed inverse problem is to introduce regularization of parameters [21], which is same as imposing prior information. Gaussian prior is equivalent to l_2 regularization [22], as shown in Eq. (3).

In Bayesian analysis against batch-wise data, such as laminar flame speed and ignition delay time [5, 6], the drawback of such a mis-specification is not obvious, because the experimental data are not too many. But in high-resolution data, such as spatial distribution and time-series, the draw back becomes obvious, and has great impact on the estimated parameters. Therefore, special attention needs to be paid to determine the number of effective independent data, *n*. The potential solution for determining a suitable *n* could be: (a) plotting a curve of prediction error versus *n*, and select *n* based on the acceptable error; (b) treating *n* as a parameter to be calibrated, like [5]; (c) using the model comparison method described in [5] to compute the evidence function with different *n*, where the one with largest evidence would be the optimal one. Here we only implement (a) for demonstration. In Fig. S16, it is obvious that when the averaged NE is smaller than 0.2, further increase of *n* only leads to slightly improvement of performance, so the optimal *n* would be around 1.

There are two noteworthy points. First, we only show the point estimation here (MAP), without sampling the posterior distribution. The C-1 approach, although can provide a more reasonable point estimation, would have a far larger posterior uncertainty than C-160, since $n = 1$ is equivalent to multiplying the experimental uncertainty by 160. In other words, this approach provides a reasonable point estimation but with larger uncertainty, while C-160 approach provides a point estimation with a concentrated posterior but deviated far away from our

prior knowledge on the parameters. Fig. 3 shows a schematic of the posterior distribution of two strategies, together with the ideal case of C-160 with perfect model and accurate likelihood. Second, the discussion of effective independent data points is based on a precondition that the prior distribution of parameters is not a uniform distribution. In some literature [5, 6], the prior distribution of rate constants is assumed as uniform distribution, in which case MAP can be considered as MLE with bounded parameters. In this case, the value of effective independent data points does not affect the optimization result. We can consider MLE as an asymptotic case with an effective independent-data number approaching infinity.

The analysis above can be a reference for both modelers and experimentalists. For modelers, they need to consider taking measures (e.g., use small n) to avoid overtuning, especially when reduced models are used, which can introduce large model error, and when high-resolution data are used. For experimentalists, providing a detailed probability density of experimental error, especially systematic error, would be very helpful for model calibration. If we can know the real distribution of total measurement error (i.e., accurate likelihood function), there is no need to introduce n (given that model error is not large), and then we can get a right point estimation with a concentrated posterior, as shown as the ideal case in Fig. 3.

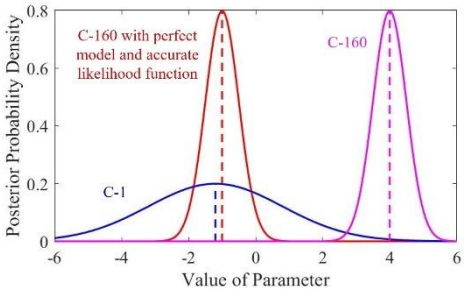

Fig. 3. The schematic of posterior distribution of C-1 and C-160. C-160 with perfect model and accurate likelihood function is used as an ideal case of Bayesian inference. The dashed vertical lines represent MAP point estimations.

3.3 The influence of target selection

In previous studies to develop kinetic models, two methods were adopted to select optimization targets in species time-histories: selecting several points in a curve [7, 8], and using all data points in the entire curve [9, 10]. A comparison of these two approaches will be discussed in this section. To avoid the influence of the effective independent-data number, C-1 was selected to represent optimization against the entire curve and compared with LastP.

From Table 1, we can see that LastP has both smaller averaged NELP and averaged NE than those of C-1, although the differences are minor. Similarly, shown in Table 4, the NEs of C-1 and LastP are similar for all 40 cases, with half positive and half negative results. This trend is further confirmed by species evolution histories included in Sec. 4 of the Supplementary Materials.

We also selected the most changed rate parameters (measured by the absolute value of x_j) in each mechanism, and compared the x_j in Fig. 4. The x_j of these reactions in the two mechanisms are very similar, except that they have opposite signs for the H-abstraction reaction generating nC_3H_7 (the third one from the top). In addition, there are only two parameters in each mechanism tuned over or near 1σ: hydrogen abstraction generating iC_3H_7 and propane synthesis by C_2H_5 and CH_3, with all others far smaller than 1σ. In a nutshell, the performance and updated kinetic parameters are similar for LastP and C-1.

Table 4
Differences between the NEs of LastP and those of C-1. Differences larger than 10% of the corresponding experimental uncertainties are in bold.

	1250 K	1290 K	1330 K	1370 K	1410 K
H_2	-0.042	0.004	-0.084	-0.068	-0.053
C_2H_2	0.001	-0.001	0.002	3E-5	-0.005
CH_4	-0.027	-0.036	-0.034	-0.019	0.011
C_2H_4	**-0.117**	**-0.156**	-0.024	**0.142**	**0.103**
C_2H_6	-0.043	-0.062	-0.071	-0.073	-0.070
pC_3H_4	2E-5	0.001	0.002	0.002	0.003
C_3H_6	-0.058	0.064	**0.146**	**0.145**	0.060
C_3H_8	**-0.213**	**-0.230**	-0.011	**0.111**	0.048

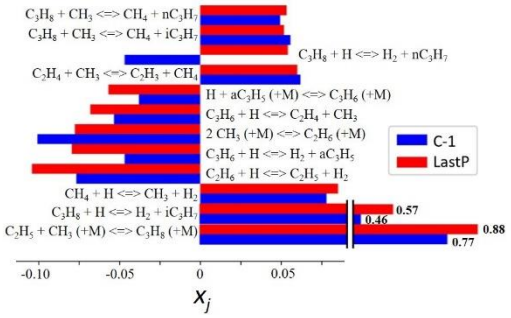

Fig. 4. The x_j of top 10 highly-tuned reaction rates in C-1 and LastP.

A potential explanation for the slightly smaller NE and NELP for LastP is that the biggest discrepancies between the curves of C-1/LastP and experimental measurements usually lie in the last points. Shown in Table 1, the averaged NELP are larger than the averaged NE for all three mechanisms. However, the experimental uncertainty for a certain curve is a constant throughout the entire time domain. Thus, the likelihood function in LastP is larger than C-1. Similar to the discussion in Section 3.2, the increase of likelihood function is equivalent to the increase of n, which will weaken the constraints of the prior distribution and make the model fit targets better.

There are two implications of the observation above. First, even keeping the effective independent-data number unchanged, different selection of targets can also change the relative weights of prior distributions and likelihood functions. Compared with using the data in the entire curve and averaging by the number of data point, using the last point can increase the weight of likelihood functions. We also can conjecture that compared with selecting another point in a species evolution curve, where the discrepancy is usually smaller than the last point, selecting the last point as target can increase the weight of the likelihood function. In addition, a modeler might need to consider the position of target points when determining the number of effective independent data. Second, from the perspective of mechanism optimization (i.e., point estimation), the information provided by a curve can be represented by several critical points. As we mentioned before, the critical point approach has been used by many previous papers, such as [7, 8]. However, from the perspective of uncertainty quantification, using representative points cannot be equivalent to using all points of the curve. For example, when n = 160, the later approach obviously gives a narrower posterior distribution, as discussed before, due to its amplified measurement uncertainty. If n = 1, the difference of posterior distributions between the critical point approach and entire curve approach can be known by sampling the posterior distribution, which is left for future studies. For experimentalists, data acquisition and measurements could focus on these critical points (although these points might be difficult to know a priori) for point estimation, but more data are still valuable to narrow the estimation uncertainty.

It seems that in the current case the last point contains almost all the information of a curve. A natural hypothesis is that all points in this curve would share similar sensitivity to a set of reactions. This is called a universal sensitivity direction [22], where the relative importance of the sensitive reactions remains the same for all data points although the magnitude of the sensitivity might change. If that holds, all points in the data series are correlated, so that the response of the curve can be inferred by the response of a single point during optimization.

To test this hypothesis, we calculated the inner product of the normalized sensitivity vectors at 1.6 ms and that at 0.4 ms/0.8 ms/1.2 ms during the pyrolysis. The results for the initial temperature of 1410 K case are listed in Table 5, while the results for other temperatures are attached in Sec. 5 of the Supplementary Materials. We can see that under most scenarios, the inner products of two normalized sensitivity vectors are close to unity. This universal sensitivity would allow representing the entire curve with the last point. For those that are not close to unity, we chose the one that deviated the most for further discussion, which is pC_3H_4. We plot the normalized sensitivity vector of pC_3H_4 at 0.4 ms and 1.6 ms in Fig. 5. Although the normalized sensitivities to most reactions are similar, there are still some reactions with much difference, such as $C_2H_5 + CH_3 (+M) \rightarrow C_3H_8 (+M)$ and $C_2H_2 + CH_3 \rightarrow H + pC_3H_4$. Thus, the universal sensitivity direction does not hold, at least for pC_3H_4 under 1410 K. This is similar to Fig. 8 of [11], where the ratio of the sensitivities to any two reactions is not a constant as time evolves, since different reactions play the dominate role at different stages. It may be due to the large experimental uncertainties of pC_3H_4, so that the lack of universal sensitivity direction does not affect the optimization results. Thus, it is still an open question that how much information of a species evolution curve can be represented by a single point.

Table 5
Inner product of the normalized sensitivity vectors at 1.6 ms and at 0.4 ms/0.8 ms/1.2 ms. Values below 0.9 are in bold.

	0.4 ms	0.8 ms	1.2 ms
H_2	0.921	0.970	1.000
C_2H_2	0.945	0.985	1.000
CH_4	0.936	0.985	1.000
C_2H_4	0.925	0.984	1.000
C_2H_6	**0.892**	0.975	1.000
pC_3H_4	**0.850**	0.960	1.000
C_3H_6	0.921	0.970	1.000
C_3H_8	0.945	0.985	1.000

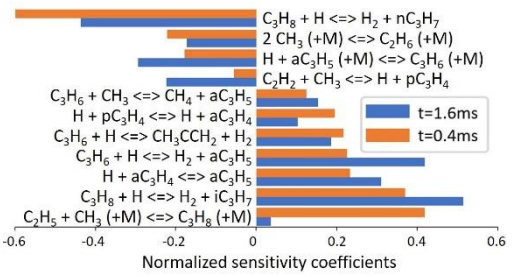

Fig. 5. Top 10 sensitivity coefficients of pC_3H_4 concentration at 1.6 ms and 0.4 ms.

It is noteworthy that in our pyrolysis case the shape of curve is monotonic or even near-linear for some temperatures and species, so maybe one point can represent the entire curve. In other cases, such as autoignition, the shape of curves could be more complex and the sensitivity direction changes with time significantly, such that multiple points are needed.

4 Conclusion

We investigated the influences of the effective independent-data number and selection of targets on the Bayesian optimization of chemical models based on species time-histories measured in shock tubes. Neural networks are trained as response surfaces and maximum a posterior estimation is used to obtain the optimal parameters. Three optimization strategies were used: using the entire species time-history curve with an effective independent-data number of 1 (C-1) and 160 (C-160), and using the last point of each curve

(LastP). It is shown that all three optimized models fit measurements better compared to the original one.

Comparing C-1 with C-160, increasing the number of targets results in slight improvement in predicting the measurements by increasing the weight of the likelihood function; however, the constraints of the prior distribution might be weakened and may lead to overtuning of parameters beyond common practice. Therefore, for modelers, the number of effective independent data should be examined carefully when utilizing Bayesian approaches for optimizing chemical models. For experimentalists, the detailed probability distribution of measurement error will be extremely valuable.

Comparing C-1 with LastP, LastP has similar and even slightly better agreement with the measurements. Using only the last point increases the weight of the likelihood functions, but no overtuning was observed. An implication is that several critical points could be enough for point estimation, suggesting that priority should be given to these critical points in handling data in modeling and acquiring samples in experiments. For the reduction of estimation uncertainty, more data are still useful.

In the future, we will combine neural networks with the MCMC algorithm to compute the full posterior uncertainty of the estimated kinetic parameters to further investigate the questions arise from this work.

Acknowledgements

HC, WJ and SD acknowledge the support from Weichai Holding Group Co., Ltd. HC acknowledges the helpful discussion with Prof. Peng Zhao at the University of Tennessee, Prof. Liming Cai at Tongji University, and Dr. Qiaofeng Li at MIT. The authors also express gratitude to reviewers of this paper for the improvement of the manuscript.

Supplementary material

SuppMat.doc and MechII_C3.yaml are attached.

References

[1] D.F. Davidson, R. K. Hanson, Recent advances in shock tube/laser diagnostic methods for improved chemical kinetics measurements, *Shock Waves* 19 (2009) 271-283.

[2] M. Frenklach, Systematic optimization of a detailed kinetic model using a methane ignition example, *Combust. Flame* 58 (1984) 69-72.

[3] D.A. Sheen, X. You, H. Wang, T. Løvås, Spectral uncertainty quantification, propagation and optimization of a detailed kinetic model for ethylene combustion, Proc. Combust. Inst. 32 (2009) 535-542.

[4] T. Turányi, T. Nagy, I.G. Zsély, M. Cserháti, T. Varga, B.T. Szabó, I. Sedyo, P.T. Kiss, A. Zempleni, H.J. Curran, Determination of rate parameters based on both direct and indirect measurements, *Int. J. Chem. Kinet.* 44 (2012) 284-302.

[5] K. Braman, T.A. Oliver, V. Raman, Bayesian analysis of syngas chemistry models, *Combust. Theory Modell.* 17(2013) 858-887.

[6] J. Wang, Z. Zhou, K. Lin, C.K. Law, B. Yang, Facilitating Bayesian analysis of combustion kinetic models with artificial neural network *Combust. Flame* 173 (2016) 468-482.

[7] S. Banerjee, R. Tangko, D.A. Sheen, H. Wang, C.T. Bowman, An experimental and kinetic modeling study of n-dodecane pyrolysis and oxidation, *Combust. Flame* 163 (2016) 12-30.

[8] D.A. Sheen, H. Wang, Combustion kinetic modeling using multispecies time-histories in shock-tube oxidation of heptane, *Combust. Flame* 158 (2011) 645-656.

[9] C. Olm, T. Varga, É. Valkó, H.J. Curran, T. Turányi, Uncertainty quantification of a newly optimized methanol and formaldehyde combustion mechanism, *Combust. Flame* 186 (2017) 45-64.

[10] T. Varga, I.G. Zsély, T. Turányi, T. Bentz, M. Olzmann, Kinetic analysis of ethyl iodide pyrolysis based on shock tube measurements, *Int. J. Chem. Kinet.* 46 (2014) 295-304.

[11] S.J. Cassady, R. Choudhary, V. Boddapati, N.H. Pinkowski, D.F. Davidson, R.K. Hanson, The pyrolysis of propane, *Int. J. Chem. Kinet*. 52 (2020) 725-738.

[12] H. Wang, X. You, A.V. Joshi, S.G. Davis, A. Laskin, F. Egolfopoulos, C.K. Law, USC Mech Version II. High-Temperature Combustion Reaction Model of H2/CO/C1-C4 Compounds. http://ignis.usc.edu/USC_Mech_II.htm

[13] D.P. Bertsekas, J.N. Tsitsiklis, Introduction to probability, Athena Scientinis, 2000.

[14] S. Wang, Y. Ding, R.K. Hanson, Information-Driven Design for Shock Tube/Laser Absorption Studies of Fundamental Rate Constants in Combustion, with Application to Methanol Pyrolysis, *arXiv preprint* arXiv:1911.02009 (2019).

[15] L. Lu, P. Jin, G. Pang, Z. Zhang, G.E. Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, *Nat. Mach. Intell*. 3 (2021) 218-229.

[16] D. Kingma, J. Ba, Adam: A method for stochastic optimization, Proceedings of the 3rd International Conference for Learning Representations (2015).

[17] D. G. Goodwin, R. L. Speth, H. K. Moffat, B. W. Weber, Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes, https://www.cantera.org, version 2.5.1 (2021).

[18] M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 42 (2000) 55-61.

[19] K. Sargsyan, H.N. Najm, R. Ghanem, On the statistical calibration of physical models, *Int. J. Chem. Kint.* 47 (2015) 246-276.

[20] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, *J. R. Statist. Soc. B* 63 (2001) 425-464

[21]J. Adler, O. Öktem, Solving ill-posed inverse problems using iterative deep neural networks, *Inverse Probl.* 33 (2017) 124007.
[22]K.P. Murphy, Machine learning: a probabilistic perspective, MIT press, 2012.
[23]W. Ji, T. Yang, Z. Ren, S. Deng, Dependence of kinetic sensitivity direction in premixed flames, *Combust. Flame* 220 (2020) 16-22.