Bayesian Inference and Experimental Design of Combustion Kinetic Models

by

Huaibo Chen

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN MECHANICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2024

Authored by: Huaibo Chen
Department of Mechanical Engineering
November 30, 2023

Certified by: Sili Deng
Assistant Professor of Mechanical Engineering, Thesis Supervisor

Accepted by: Nicolas Hadjiconstantinou
Professor of Mechanical Engineering
Graduate Officer, Department of Mechanical Engineering

Bayesian Inference and Experimental Design of Combustion Kinetic Models

by

Huaibo Chen

Submitted to the Department of Mechanical Engineering
on November 30, 2023 in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN MECHANICAL ENGINEERING

ABSTRACT

In combustion kinetic model calibration, researchers usually use experimental data to reduce the uncertainty of kinetic parameters, and Bayesian inference is the most common approach to do inverse calibration. This thesis explores two interconnected aspects of Bayesian approaches in the context of combustion kinetics models: how to utilize high-resolution species profiles in Bayesian inference and how to identify the most informative experimental conditions to collect data. In the first part, we investigated the impact of the effective independent-data number and target selection on Bayesian inference of kinetic parameters using species time-histories obtained from shock tube experiments. Neural networks serve as response surfaces. Maximum a posteriori estimation and Markov chain Monte Carlo sampling are employed to determine optimal parameters as well as their uncertainty. Three optimization strategies are employed: utilizing the entire species time-history curve with effective independent-data numbers of 1 (C-1) and 160 (C-160), and using only the last point of each curve (LastP). All three improved models fit experimental data better. Comparing C-1 with C-160 reveals that increasing the number of targets improves prediction accuracy but may lead to overtuning. Comparing C-1 with LastP, LastP exhibits comparable or slightly better agreement with measurements, suggesting that focusing on critical points is effective for point estimation. However, C-1 shows different posterior uncertainty from LastP in both parameters and predictions, despite their similarity in the point estimation.

Experimental data obtained at different experimental conditions (e.g., pressure, temperature, equivalence ratio, etc.) is not equally informative when it is used to calibrate kinetic parameters. Thus, experimental design becomes an important topic in combustion kinetics, where the most informative condition can be identified by algorithms. In the second part, we propose an efficient Bayesian experimental design algorithm that integrates Laplacian approximation-based experimental design with gradient-based design optimization, employing sophisticated neural network response surfaces for mapping kinetic parameters to target prediction at a wide range of thermodynamic conditions. The algorithm demonstrates efficiency and robustness against local maxima. Additionally, to meet various needs in kinetic experiments, we develop various experimental design targets based on the posterior covariance matrix, including model-oriented, parameter-oriented, target-oriented, and parallel experimental design. The proposed method, utilizing a full posterior covariance matrix without fixing any parameter of insensitive reactions, achieves significant acceleration compared to

previous methods, demonstrating effectiveness in reducing parameter and target uncertainty as well as designing multiple experiments simultaneously.

Thesis supervisor: Sili Deng
Title: Assistant Professor of Mechanical Engineering

Acknowledgments

I express my sincere gratitude to Professor Sili Deng, my advisor, for her unwavering guidance throughout this project and her continuous support throughout my academic journey at MIT. I learned a lot from her in the perspectives of academic knowledge, group management, networking etc., or in a nutshell, how to be an excellent PI in a top-tier university.

Special thanks to Dr. Weiqi Ji, whose invaluable assistance played a pivotal role in my transition from a combustion physics researcher to one integrating combustion with machine learning. I extend my appreciation to my undergraduate advisors, Professors Liguang Li, Peng Zhao, Fei Qi and Xi Xia, for introducing me to the field of combustion and providing invaluable guidance from academic, career, and life perspectives. Their support was instrumental in my admission to MIT.

I am grateful to the members of the machine learning subgroup in the Deng Energy and Nanotechnology group—Dr. Qiaofeng Li, Benjamin Koenig, Stefan Borjan, and Edoardo Ramalli—for their weekly discussions, advice, and support. I also acknowledge the broader DENG group, including Maanasa Bhat, Suyong Kim, Valerie Muldoon, Gwendolyn Tsai, Dr. Yuesen Wang, Chuwei Zhang, and Dr. Jianan Zhang, for fostering a collaborative and friendly work environment over the past two years.

The work included in this thesis received support from Weichai Holding Group Co., Ltd. and the National Science Foundation under Grant No. CBET-2143625. I extend my appreciation to Prof. Peng Zhao at the University of Tennessee, Prof. Liming Cai at Tongji University, Dr. Qiaofeng Li, Dr. Qinghui Meng at the University of Minnesota, and Mr. Wendi Dong at Stanford University for their insightful discussions.

Finally, heartfelt thanks to my parents Mr. Xiangyang Chen and Mrs. Jian Zhang, for their enduring support and love. Their efforts to create a warm family environment and guide me in decision-making, while respecting my opinions, have left an indelible mark on my journey. Every place I have been to bears the footprints of my parents. This Master's thesis is dedicated to them on the eve of their arrival in the United States next month.

Contents

List of Figures

List of Tables

Chapter 1

Background

With the development of computer hardware and computational fluid dynamics (CFD) techniques, simulation plays an increasingly important role in combustion science and engineering. In academia, CFD is generally used to unravel combustion physics and discover new phenomena [1], especially for scenarios where it is hard, or expensive, to conduct experiments [2]. In industry, CFD is widely used to guide new designs of engines and has been able to substitute a large amount of experiments [3].

However, combustion CFD still has large uncertainty even for most detailed simulations. First, due to the unsteady, chaotic, and multiscale nature of turbulence, direct numerical simulation (DNS) is computationally expensive and only can be conducted in limited scenarios. Either Reynolds averaged Navier-Stokes (RANS) or Large-Eddy Simulation (LES) cannot resolve Kolmogorov scales, and relies on closure models, which introduce fluid modelling uncertainty [4]. Second, the computation of chemistry requires a chemical kinetic model (or "mechanism" in some literature). In turbulent combustion simulations, chemical models are usually highly reduced, having lumped reactions or missing reactions and species; even if detailed models are used, the kinetic parameters usually have large uncertainty, which can greatly influence the accuracy and reliability of simulation results [5]. Third, in turbulent combustion simulation, fluids and chemistry are highly coupled. Due to the nonlinearity of the Arrhenius equation, the chemical source term cannot be directly averaged along with velocities. The coupling between fluid and chemistry also needs modelling, such as flamelet models, probability density function (PDF) approaches, and multiple mapping conditioning (MMC) [1]. This further introduces assumptions, and hence uncertainty, into simulations. Fourth, in order to reduce computational budgets, heat and mass transport are usually simplified, such as mixture-averaged diffusion models and unity Lewis number assumption, which introduces uncertainties in heat and mass transport modelling [6]. These uncertainties from different sources may never be rooted out. Thus, how to quantify the uncertainty of simulations, and how to reduce the uncertainty in models at specific scenarios, become important topics in the combustion community, and there are many recent works focusing on this problem [7]. In this thesis, we focus on the uncertainty quantification (UQ) from chemical kinetic models, and the uncertainty from other sources is beyond our scope.

UQ usually includes two tasks. First is propagating the uncertainty of parameters into the model's prediction (hence informing the downstream decision making), which is called *forward problems*. In order to reduce the uncertainty of models, a common practice is to

experimentally measure the quantity of interest (QoI) that can be predicted by models, and infer parameter uncertainty from experimental data. In this case, the uncertainty is propagated from experimental measurements to parameters, which is called *inverse problems* or *model calibration*. In this thesis, we deal with the inverse UQ, or to be specific, Bayesian inference of kinetic parameters given experimental measurements. In the first part, we investigate how to utilize time series data of shock tube species profiles. In the second part, we study how to efficiently design experimental conditions so that the gathered data can optimally inform parameters. In the first chapter, we briefly overview relevant background knowledge and recent advances.

1.1 Uncertainty Quantification of Combustion kinetics

A combustion kinetic model includes the species, reaction pathways, and parameters describing the reaction rate of each elementary reaction. Generally, there are three ways to determine the reaction rate constants: direct measurement by experiments, quantum chemistry calculation, and calibration against global experimental data. Unfortunately, all of them possess intrinsic uncertainty: experimental data can be affected by a systematic error in instruments and random signal noise [8]; the inputs of quantum chemistry calculation, such as potentials, electronic wavefunctions, and geometry always have uncertainties, and high-level theory is usually computationally prohibitive, especially for large molecules [9]. As pointed out by Wang and Sheen [5], perfect kinetic models will never exist. Even for hydrogen combustion kinetics, which includes only 20 elementary reactions and 8 species, the uncertainty space is too large to "pin" the reaction rate constants at a fixed point without considering uncertainties [5]. The uncertainty of kinetic parameters can have a large impact on model predictions. Fig. 1.1 illustrates the impact of the kinetic uncertainty on the species profile predictions of ethylene oxidation. The left panel is done by USC Mech II [10], which shows a large uncertainty range. Even if we decreased the uncertainty factors to 1.15 (it means the 1-σ multiplicative uncertainty bound is the nominal value times or divided by 1.15), as shown by the right panel, the uncertainty in prediction is still obvious. Note that even nowadays, 16 years after the development of USC Mech II, the uncertainty factors of most kinetic parameters are still larger than 1.15. Given that an perfect representation of actual physics does not exist, the focus of kinetic model development should shift to quantifying the uncertainty in the model from finding the exact values for each kinetic parameter. This measurement of imprecision relates the level of "trust" attributed to the simulation outcomes when employed for decision-making.

Another important application of UQ in chemical kinetics is solving the model proliferation problem. Before combustion engineers start to conduct UQ, there usually exist several models with different reaction rate constants. For example, before high-pressure laminar flame speeds are available, there are several models for hydrogen combustion with different reaction rate constants, and all of them can make good predictions against experimental data. However, they show quite different predictions of the laminar mass burning rate at elevated pressure [17], as shown by Fig. 1.2. In fact, after the assessment of uncertainty, denoted by the shadow region of Fig. 1.2, we can find that the predictions from different models are within the prediction uncertainty. Thus, different models can be considered as

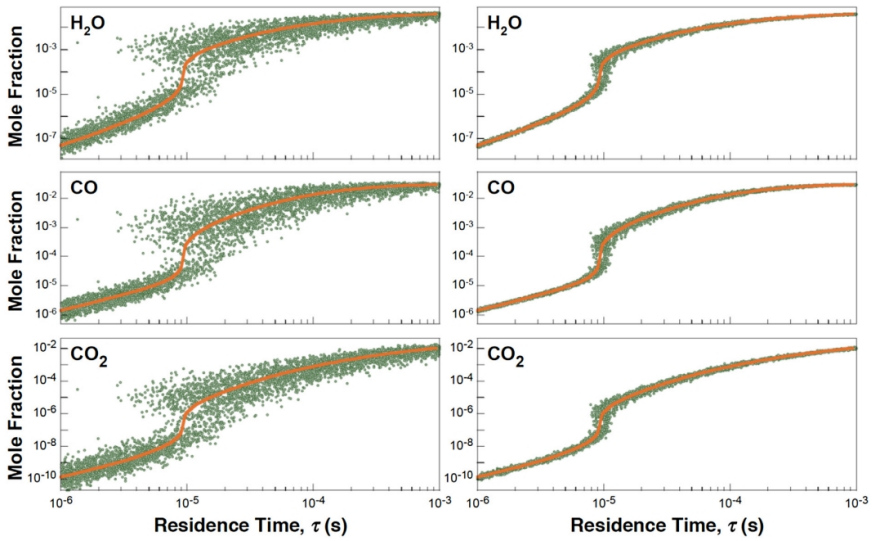


Figure 1.1: Monte Carlo sampling of species trace prediction for ethylene oxidation in air in a perfectly stirred reactor (phi=0.5, p=30 bar, and T=1200K) within the uncertainty of USC Mech II [10]. The left panel represents the original uncertainty, while the right panel shows the prediction after setting all the uncertainty factors into 1.15. Red lines represent the prediction of nominal values of parameters. This figure is adopted from [5].

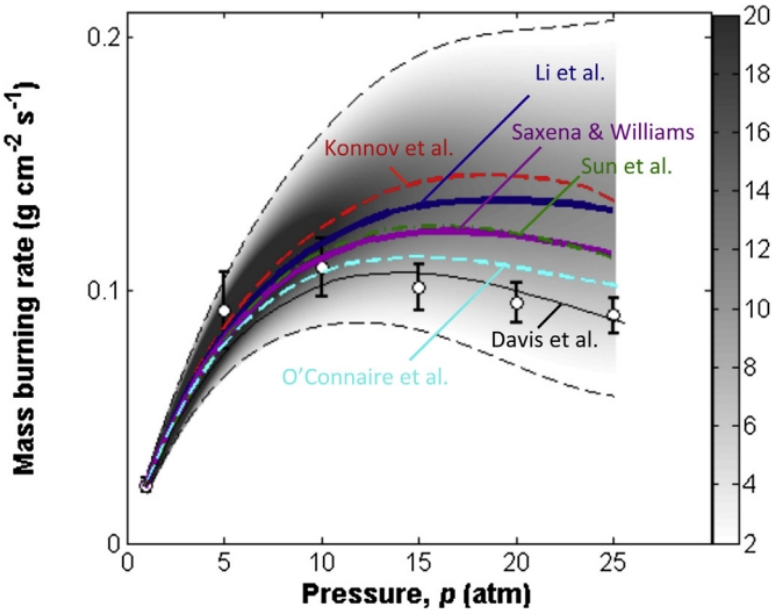


Figure 1.2: Uncertainty range of trial model of [11] (denoted by shadow area and dashed black lines for 2-σ bounds) overlapped by the predictions of different models [11]–[16], and experimental measurements (dots with 2-σ error bar). This figure is adopted from [5].

different statistical samples from the uncertainty space [5]. Since no high-pressure laminar flame data was available, the uncertainty at high pressure is far larger compared with that at low pressure. With a careful assessment of uncertainty, the controversy of models with dif-

ferent parameters does not exist anymore. Meanwhile, in the development of kinetic models, the uncertainty associated with the model is highly needed.

In UQ of combustion kinetics, we have a parametric model describing the reaction pathways with kinetic parameters assigned to each elementary reaction. For a forward elementary reaction

$$\sum_{i=1}^{N} v_i' \mathrm{M}_i \xrightarrow{k_f} \sum_{i=1}^{N} v_i'' \mathrm{M}_i, \tag{1.1}$$

where M_i is species i, N is the number of species, v_i' and v_i'' are the stoichiometric coefficients for reactants and products respectively (for species not involved in this reaction, the coefficients are just 0), k_f is the reaction rate constant. The reaction rate of each elementary reaction is described by the Law of Mass Action [6]:

$$\omega = k_f(T) \prod_{i=1}^{N} c_i^{v_i'}, \tag{1.2}$$

where T is temperature, c_i is the concentration of species i. Furthermore, the reaction rate constant can be expressed by Arrhenius Law [6]:

$$k(T) = A e^{-E_a/R^o T}, \tag{1.3}$$

where A is the pre-exponential factor, E_a is the activation energy, R^o is the universal gas constant. For some elementary reactions, A is found to be temperature-dependent, so a modified Arrhenius law is used [6]:

$$k(T) = A T^b e^{-E_a/R^o T}, \tag{1.4}$$

where b is the temperature exponent. Based on the reaction pathways and kinetics parameters, chemical kinetics can be mathematically formulated as a group of ordinary differential equations (ODEs). The chemical and energy source terms in species and energy conservation equations of CFD can be computed by integrating the ODE system. Due to the huge difference in the reaction rate for different reactions, the time scales of different species are quite different. Mathematically, this is manifested by the strong stiffness of this ODE system. Hence, the integration of such ODE systems requires a small step size and many iterations [18]. In UQ of combustion kinetics, the forward problem refers to assessing the impact of uncertainty of kinetics parameters E_a, A, and b on the accuracy of prediction, while the inverse problem refers to how to calibrate the uncertainty of these parameters based on some global experimental targets, such as laminar flame speeds, ignition delay time, and species profiles in shock tubes, etc. In order to eliminate the uncertainty from sources other than chemical kinetics, these experiments are usually done in well-isolated and homogeneous environments and are hardly influenced by fluid transport.

The general workflow for UQ of combustion kinetics is shown in Fig. 1.3. In UQ algorithms, we usually need to repeatedly evaluate the physical model; sometimes the derivatives of target with respect to parameters are needed. Thus, we usually build a surrogate model (i.e., response surface) mapping the kinetic parameters to model predictions. Such an approximate function is cheap to evaluate and make derivatives accessible. Different types of

surrogate models will be introduced in Section 1.3. After replace the physical model by a surrogate model, we can efficiently implement forward or inverse UQ algorithms. For forward problems, the uncertainty of parameters will be propagated into the uncertainty of predictions; for inverse problem, the uncertainty of measurements would be propagated into parameters (sometimes combined with prior distributions). For forward problems, as long as a surrogate model has been built, the uncertainty from parameters can be propagated into targets using Monte Carlo simulation (for polynomial chaos expansion, a special type of surrogate models, some statistical properties of targets can be directly computed by coefficients in this surrogate model). Thus, the main focus for forward problems is building surrogate models, which will be introduced in details in Section 1.3. In Section 1.2, we will introduce the algorithms for UQ of combustion kinetics, mainly for inverse problems.

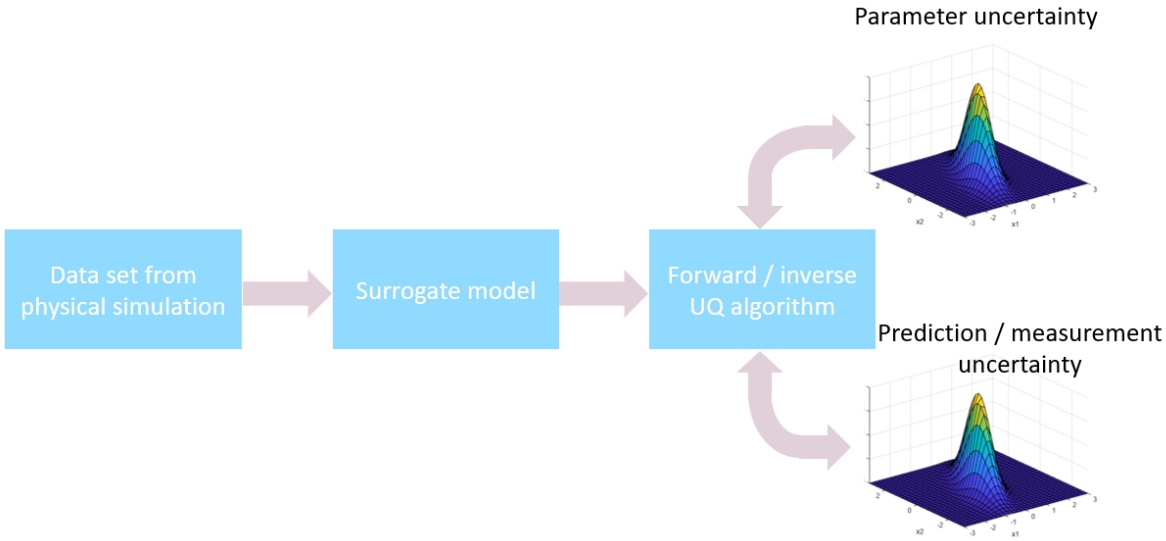


Figure 1.3: The flowchart of uncertainty quantification for combustion kinetics.

1.2 Algorithms for Kinetic Uncertainty Quantification

We define $\boldsymbol{k}$ as the vector of all kinetic parameters in a model. It is a common practice that only the pre-exponential factors are considered in UQ of combustion kinetics, while other parameters are frozen [19], [20]. We also adopt this convention in this thesis unless additionally indicated. We define $\boldsymbol{M}(\boldsymbol{k};\boldsymbol{d})$ as the prediction of the kinetic model at experimental condition $\boldsymbol{d}$ when the kinetic parameters are $\boldsymbol{k}$. Here, the experimental condition is the one used to initiate an experiment, such as pressures, initial temperatures, and the composition of initial mixtures. In many literature, the kinetic parameters are normalized to x_i by its corresponding uncertainty factor f_i, which reflects its prior uncertainty or bounds:

$$x_i = \frac{\ln(k_i/k_{i,0})}{\ln f_i}, \tag{1.5}$$

where $k_{i,0}$ is the nominal value of parameters, and f_i represents the multiplicative uncertainty of k_i, as mentioned in Fig. 1.1.

1.2.1 Solution Mapping Method

The history of kinetic UQ can date back to the pioneering works by Frenklach and coworkers [21], [22], where solution mapping method is used to calibrate the parameter uncertainty from experimental data. They first conducted sensitivity analysis for a specific target to determine the active parameters of a target, and then learned a surrogate model (here referred as "solution mapping") from these active parameters to the prediction of the model for a target to replace the expensive computation of ODE systems:

$$f_i(\boldsymbol{k}) \approx M(\boldsymbol{k}; d_i), \tag{1.6}$$

where i is the index for different experimental targets. Please notice that here only active parameters are selected for the construction of solution mapping, but we still use notation $\boldsymbol{k}$. According to the principle of effective sparsity [23], only a small amount of parameters will influence the model's prediction at a specific target, and their correlation can be at most second order. Thus, in [21] and [22] they use second-order polynomials to construct the solution mapping. As introduced in the end of Section 1.1, response surface techniques have became an essential procedure of UQ nowadays, and we usually separate the following UQ algorithms with surrogate models. However, since the authors of [21], [22] call their algorithms "solution mapping method", and their works are the first attempt to conduct UQ of combustion kinetics, "solution mapping" can refer to both the surrogate model and the entire procedure presented in [21], [22]. After having a cheap approximation of physical models, they optimize the kinetic parameters by minimizing sum-of-square errors between the model's prediction and experimental data:

$$\boldsymbol{k}^* = \underset{\boldsymbol{k}}{\operatorname{argmin}}\ \Phi(\boldsymbol{k}) = \underset{\boldsymbol{k}}{\operatorname{argmin}} \sum_{i=1}^{T} \left(\frac{f_i(\boldsymbol{k}) - y_i}{\sigma_i} \right)^2, \tag{1.7}$$

where T is the number of experimental data, y_i is the i-th experimental data point, σ_i is the experimental uncertainty of i-th data points, and $\phi(\boldsymbol{k})$ is the objective function. As for the uncertainty of parameters, the two papers used different methods. In [21], the 95% confidence region is obtained by F distribution:

$$\Phi\left(\boldsymbol{k}\right) = \Phi\left(\boldsymbol{k}^*\right) \left\{ 1 + \frac{p}{n-p} F_{5\%}(p, n-p) \right\}, \tag{1.8}$$

where n is the number of experimental data point, p is the number of parameters, $F_{5\%}$ is the probability density function (PDF) of F distribution corresponding to top 5% cumulative density. In [22], since they calibrate the kinetic parameters using data points less than number of parameters, the parameter hypersurface where the model can exactly predict the experimental measurements are given and plotted (see Fig. 6 of [22]) combined with the constraints of prior bound for each parameters.

1.2.2 Bound-to-Bound Data Collaboration

Based on solution mapping techniques (second-order polynomials), Frenklach and coworkers developed GRI-Mech [24], the first kinetic model using systematic approaches of optimization

and uncertainty quantification. They further developed a deterministic data collaboration framework called bound-to-bound data collaboration (B2BDC) [25], where parameters are required to lie in both the prior intervals and the ranges where the model's prediction are within uncertainty bounds of experiments:

$$\mathcal{F} = \{\boldsymbol{k} | \boldsymbol{k} \in \mathcal{H}, y_i - \sigma_i \leq f_i(\boldsymbol{k}) \leq y_i + \sigma_i, i = 1, 2, \ldots, n\}, \tag{1.9}$$

where $\mathcal{H}$ is the set of prior hypercube defined by the upper and lower bounds of each parameter

$$\mathcal{H} = \{\boldsymbol{k} | lb_i \leq k_i \leq ub_i, i = 1, 2, \ldots, p\}. \tag{1.10}$$

The posterior bounds of parameters and predictions are obtained by maximizing and minimizing prediction or parameters within the feasible set $\mathcal{F}$. For example, the posterior bounds of prediction at d_i are:

$$\left[\min_{k \in \mathcal{F}} f_i(\boldsymbol{k}), \quad \max_{k \in \mathcal{F}} f_i(\boldsymbol{k})\right]. \tag{1.11}$$

For the forward problem, one just needs to replace $\mathcal{F}$ with $\mathcal{H}$ in Eq. 1.11. After that, a series of papers have been published dealing with the data inconsistency [26], model discrimination [27] and model error [28] etc. In fact, B2BDC already shows the features of Bayesian statistics, where the parameters are estimated based on both prior understanding and observations [29], but as a deterministic framework, it is still different from the Bayesian analysis we will introduce in the following part.

1.2.3 Spectral Expansion Method

With the development of polynomial chaos expansion (PCE), i.e., spectral expansion, and its application in combustion science, Wang and coworkers developed a method for inversion of kinetic parameters based on PCE [30]. The details of PCE will be shown in Section 1.3. In [30], the parameters are first normalized by Eq. 1.5, and hence $x_i \sim N(0, \frac{1}{2})$, then distributions of experimental data and parameters are both expanded as polynomial chaos:

$$\boldsymbol{x} = \boldsymbol{x}^{(0)} + \sum_{i=1}^{p} \boldsymbol{\alpha_i} \xi_i + \sum_{i=1}^{p} \sum_{j \geq i}^{p} \boldsymbol{\beta}_{ij} \xi_i \xi_j + \cdots, \tag{1.12}$$

$$y_r(\boldsymbol{\xi}) = y_{r,0} + \sum_{i=1}^{n} \sigma_r \delta_{ir} \xi_i, \tag{1.13}$$

where $\boldsymbol{\alpha_i}$, $\boldsymbol{\beta_{ij}}$ and $\boldsymbol{x}$ are p dimension vectors, p is the number of active parameters, δ_{ir} is Kronecker product. Before model optimization against experimental data, $[\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \cdots, \boldsymbol{\alpha_p}] = \frac{1}{2}\boldsymbol{I}_p$, where $\boldsymbol{I}_p$ is an $p \times p$ identity matrix. $\boldsymbol{\beta_{ij}}$ is zero vector. Then they connect inputs and outputs by second-order polynomial response surfaces

$$f_r(\boldsymbol{x}) = f_{r.0}(\boldsymbol{x}) + \sum_{i=1}^{p} a_{r,i} x_i + \sum_{i=1}^{p} \sum_{j \geq i}^{p} b_{r,ij} x_i x_j. \tag{1.14}$$

Plugging Eq. 1.12 into Eq. 1.14, we can get the PCE of the model's prediction

$$f_r(\boldsymbol{\xi}) = f_r\left(\mathbf{x}^{(0)}\right) + \sum_{i=1}^{p} \hat{\alpha}_{r,i}\xi_i + \sum_{i=1}^{p}\sum_{j\geq i}^{p} \hat{\beta}_{r,ij}\xi_i\xi_j + \cdots, \tag{1.15}$$

where $\hat{\boldsymbol{\alpha}}_r = \frac{1}{2}\boldsymbol{I}_p\boldsymbol{a}_r$, $\hat{\boldsymbol{\beta}}_r = \frac{1}{4}\boldsymbol{I}_p^T\boldsymbol{b}_r\boldsymbol{I}_p$ before model optimization. Match the PCEs of experimental data and model's prediction by minimizing the sum-of-square loss, i.e., Eq. 1.7, and optimize the PCE coefficients of kinetic parameters:

$$\begin{aligned}\Phi\left(\boldsymbol{x}^{(0)*}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*\right) = \min_{\boldsymbol{x}^{(0)},\boldsymbol{\alpha},\boldsymbol{\beta}} \sum_{r=1}^{n} \Bigg\{ \left[y_{r,0} - f_r\left(\boldsymbol{x}^{(0)}\right)\right]^2 + \sum_{i=1}^{p}\left[\sigma_r\delta_{ir} - \hat{\alpha}_{r,i}\right]^2 \\ + \sum_{i=1}^{p}\sum_{j=i}^{p} \hat{\beta}_{r,ij} + \cdots \Bigg\} / {\sigma_r^{\text{obs}}}^2.\end{aligned} \tag{1.16}$$

Then substituting the optimized PCE coefficients Eq. 1.12 and Eq. 1.15, we can get the posterior parameters and prediction distributions, respectively.

1.2.4 Bayesian Approaches

Until now, all the methods except B2BDC try to optimize kinetic parameters in a least-square manner. This is the typical Frequentist school in statistics [29], where the parameters are tried as a deterministic constant. As a parallel school, Bayesian statistics treat parameters as random variables, and the observation would help update the belief of variables based on prior knowledge of them. The key difference between the two schools of statistical inference is whether prior information is utilized. Bayesian inference has been used in model calibration in the last century [31]–[33], but its application in the combustion community is relatively late. Kraft and coworkers [34] applied Bayesian inference to the hydrodynamic model of a rotating disc contactor. In 2009, Najm et al. [35] applied Bayesian inference in combustion chemistry for the first time, where the activation energies and pre-exponential factors of a single-step methane combustion scheme is calibrated. Braman, Oliver and Raman [19] quantified the uncertainty of a hydrogen combustion model involving 20 elementary reactions, while Miki et al. [36], [37] used Bayesian inference to calibrate A, b, E_a of $\mathrm{H} + \mathrm{O}_2 \rightarrow \mathrm{OH} + \mathrm{O}$ reaction. In the recent ten years, Bayesian analysis has become more and more popular in the combustion community [5].

The continuous Bayes Theorem argues that

$$p(\boldsymbol{k}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{k})p(\boldsymbol{k})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{k})p(\boldsymbol{k})}{\int_{\boldsymbol{K}} p(\boldsymbol{y}|\boldsymbol{k})p(\boldsymbol{k})d\boldsymbol{k}}, \tag{1.17}$$

where $p(\boldsymbol{k})$ is the prior distribution of kinetic parameter, $p(\boldsymbol{y}|\boldsymbol{k})$ is the likelihood function, $p(\boldsymbol{y})$ is the the probability density of obtaining measurement $\boldsymbol{y}$, also called evidence function, $\boldsymbol{K}$ is the support of $\boldsymbol{k}$, and $p(\boldsymbol{k}|\boldsymbol{y})$ is posterior distribution. By introducing measurement y, one can update the distribution from prior to posterior. Thus, Bayesian inference is a natural way of continuous learning [19]. In Bayesian analysis, prior distribution reflects

one's previous understanding or belief of the parameters. In combustion kinetics, prior distributions are usually obtained from direct measurements or theoretical calculations of kinetic parameters. Likelihood function $p(\boldsymbol{y}|\boldsymbol{k})$ provides the probability density of obtaining measurement $\boldsymbol{y}$ given kinetic parameter $\boldsymbol{k}$. In model calibration, likelihood usually represents the noise, or uncertainty of a measurement. We usually assume that the measurement is generated from model prediction at "ground truth" parameters plus measurement noise:

$$\boldsymbol{y} = M(\boldsymbol{k_0}; \boldsymbol{d}) + \boldsymbol{\epsilon}, \tag{1.18}$$

where $\boldsymbol{k_0}$ is the ground truth of physical parameters. Although we may never know the ground truth, we usually assume that it exists. $\boldsymbol{\epsilon}$ is the measurement noise following a certain probabilistic distribution. For example, for a Gaussian noise $N(0, \Sigma)$, the likelihood function can be written as

$$p(\boldsymbol{y}|\boldsymbol{k}) = \frac{1}{(2\pi)^{n/2}\det(\Sigma)^{1/2}} \exp\left\{-\frac{1}{2}[\boldsymbol{y} - \boldsymbol{M}(\boldsymbol{k}, \boldsymbol{d})]^\top \Sigma^{-1} [\boldsymbol{y} - \boldsymbol{M}(\boldsymbol{k}, \boldsymbol{d})]\right\}, \tag{1.19}$$

where n is the number of measurements, also the dimension of this multivariate Gaussian distribution. The likelihood also measures the distance between the model's prediction and measurements. The likelihood achieves its maximum value when model's prediction can perfectly meet with experimental data. A special case is that the noise of different measurements is independent of each other, which means Σ is a diagonal matrix. In this case, the likelihood can be written as

$$p(\boldsymbol{y}|\boldsymbol{k}) = \frac{1}{\prod_{i=1}^{n}\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - M(\boldsymbol{k}, d_i)}{\sigma_i}\right)^2\right], \tag{1.20}$$

where σ_i is the standard deviation of each measurement. Eq. 1.20 is for additive Gaussian noise. For multiplicative noise, just replace y_i, $M(\boldsymbol{k}, d_i)$ and σ_i by $\ln y_i$, $\ln M(\boldsymbol{k}, d_i)$ and $\ln \sigma_i'$, respectively, where $\ln \sigma_i'$ is the multiplicative noise level similar to f_i in Eq. 1.5. Note that likelihood measurement the distance between model prediction and measurement weighted by experimental uncertainty, and maximizing the likelihood function is exactly the same as minimizing least-square error in Eq. 1.7 and Eq. 1.16. In fact, maximum likelihood estimation (MLE) is the same as least-square regression [38].

Although we can express posterior distribution by Bayes Theorem, in most of the time we cannot obtain the posterior distribution analytically, and drawing samples from the posterior is not an easy task. The reason is that the evidence function $p(\boldsymbol{y})$ in Eq. 1.17 is unknown. Due to the high-dimensional nature of many parametric physical models, $p(\boldsymbol{y})$ cannot be accurately computed by numerical integration. Thus, different methods have been developed to draw samples from or approximate posterior distributions.

Markov Chain Monte Carlo

For sampling the posterior distribution, Markov chain Monte Carlo (MCMC) is the most widely used algorithm. The basic idea of MCMC is to construct a Markov chain so that the steady distribution of the chain is exactly the posterior distribution. It can be theoretically proven that MCMC asymptotically converges to the target posterior distribution [38]. The

original version for MCMC is called Metropolis-Hastings (M-H) algorithm developed in the 1950s [39] and extended in 1970s[40]. Currently, in the combustion community M-H algorithms are the most generally used MCMC algorithm [19], [20], [35], [41]. Duane et al. [42] combined the Metropolis algorithm with Hamiltonian dynamics to mitigate random walks and hence accelerate the convergence. Li et al. [43] utilized this hybrid Monte Carlo (HMC) algorithm to sample the uncertainty of chemical reaction neural networks (CRNN) [44].

The Methods of Uncertainty Minimization Using Polynomial Chaos Expansion

However, MCMC is computational prohibitive. Another widely used Bayesian framework in kinetic community called the method of uncertainty minimization using polynomial chaos expansion (MUMPCE) [45] linearizes polynomial response surfaces around maximum posterior point, so that the posterior distributions are Gaussian and can be analytically derived. Specifically, the parameters are normalized by Eq. 1.5, where $\ln f_i$ represents double of the standard deviation for Gaussian prior distribution, or half length of uniform prior distribution. For parameters with uniform prior distribution, the optimal parameters can be obtained by

$$\Phi\left(\boldsymbol{x}^*\right) = \min_{-1<x_i<1} \sum_{r=1}^{n} \left[\frac{f_r(\boldsymbol{x}) - y_r}{\sigma_r}\right]^2, \tag{1.21}$$

and for Gaussian prior distributions, the optimal parameters are

$$\Phi\left(\boldsymbol{x}^*\right) = \min_{x} \left\{\sum_{r=1}^{n} \left[\frac{f_r(\boldsymbol{x}) - y_r}{\sigma_r}\right]^2 + \sum_{j=1}^{k} 4x_j^2\right\}. \tag{1.22}$$

Here, obviously, they assume Gaussian, independent measurement noise. The authors derived these objective function by plugging polynomial chaos into second order polynomial response surfaces. In fact, it can be directly obtained by taking the logarithm of numerator part of Eq. 1.17 and maximize it. Eq. 1.21 is pretty similar to least-square regression in Eq. 1.7, but with bounds for parameters, indicating the impact the uniform prior distribution. Eq. 1.22 is same as least-square regression with l_2 regularization [38]. In statistical inference, Eq. 1.21 and Eq. 1.22 are called maximum a posteriori (MAP) estimation. Then linearizing the response surface, the covariance matrix for uniform prior is

$$\Sigma^* = \left[\sum_{r=1}^{n} \frac{J_r(\boldsymbol{x}^*) J_r(\boldsymbol{x}^*)^T}{\left(\sigma_r\right)^2}\right]^{-1}, \tag{1.23}$$

and for Gaussian prior distribution the covariance matrix is

$$\Sigma^* = \left[\sum_{r=1}^{n} \frac{J_r(\boldsymbol{x}^*) J_r(\boldsymbol{x}^*)^T}{\left(\sigma_r\right)^2} + 4\mathbf{I}\right]^{-1}. \tag{1.24}$$

Here $J_r(\boldsymbol{x}^*)$ is the $k \times n$ Jacobian matrix of model response evaluated at $\boldsymbol{x}^*$, where k is the number of parameters, n is the number of experimental data points. In Bayesian statistics, the technique of approximating posterior distribution by linearizing models around

MAP point is called Laplacian approximation, which provides a Gaussian distribution whose logarithm of PDF has the same curvature with real posterior distributions. Fig. 1.4 adopted from [46] compares the results of Laplacian approximation, MCMC and the unnormalized density. It is clear that Laplacian approximation can capture the peak mode of posterior distribution and has similar "width" around MAP point (please note that the the peak values are not necessary the same, although looks same in this figure). MCMC can capture all the modes while Laplacian approximation only can capture the MAP mode.

Variational Inference

Another popular way of Bayesian computation in the computer science community is variational inference [47], where a parametric distribution is specified and parameters are optimized to decrease the K-L divergence of parametric distribution and posterior distribution. This is used in [43] and [48] and integrated into a chemical kinetic inference toolkit [49].

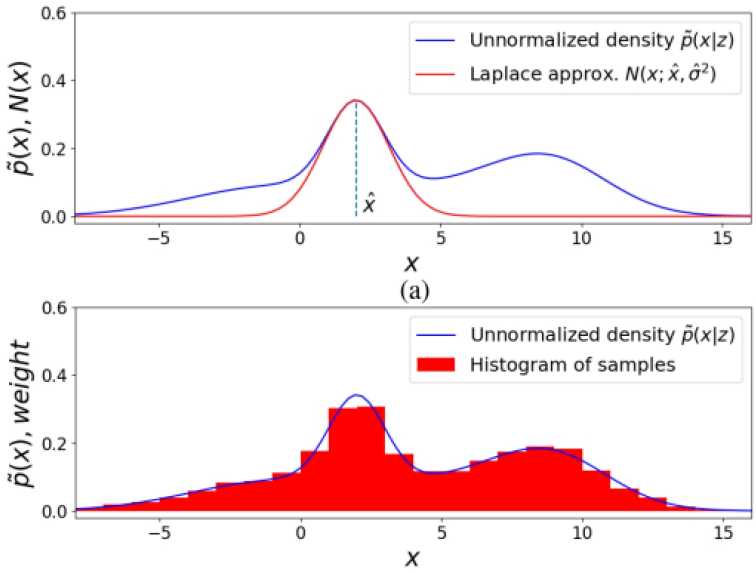


Figure 1.4: The comparison of MCMC, Laplacian approximation and unnormalization density. This figure is adopted from [46].

For other methods of UQ for combustion kinetics, please refer to [5].

1.3 Response Surfaces

As mentioned in Section 1.1, the model needs to be evaluated repeatedly and gradient information is usually needed for optimization. Thus, it is a common practice to build a surrogate model which can be evaluated very fast. In this section we briefly introduce different types of response surfaces that have been used in the combustion community.

1.3.1 Polynomials

Solution mapping method utilizes second order polynomials as response surfaces [22]. The data is generated based on factorial design, which means all the combinations of $x_i = 1$ or -1 are used. For p active parameters, the physical model needs to be evaluated for 2^p times. Although effective sparsity shows that only a small portion of parameters are active for a specific target, for large model factorial design is still very time-consuming. Davis et al. [50] proposed a sensitivity-based method of constructing polynomials. The first-order and second-order derivatives at nominal values are computed by a central difference scheme. In this approach, the number of model evaluation scales with the number of active parameters in a polynomial relation, instead of exponentially increasing in factorial design. In addition, they find that using polynomials up to the degree of two is accurate enough.

1.3.2 Polynomial Chaos Expansion

At the same time, in the engineering mechanics community, polynomial chaos expansions (PCE) method was developed as an efficient way to represent stochastic processes [51]. Then, it was applied in the UQ of combustion chemistry [35], [52], [53]. PCE also utilizes polynomials to approximate input-output relations of parametric physical models. Different from the solution mapping method, however, physical models are evaluated at quadrature points, while solution mapping uses data generated using factorial design [22], sensitivity-based design [50] or randomly generated data [54] to conduct polynomial regression. Thus, PCE enjoys high accuracy with low computational cost. Generally speaking, as long as a response surface has been built, the forward propagation can be done by Monte Carlo sampling within parameter uncertainty space and passing them through the response surface [54]. For PCE, however, the statistics of prediction can be directly computed from polynomial coefficients and hence no Monte Carlo simulation is needed. Suppose we have an infinite set of zero means, unit variance random variables $\xi = \{\xi_i\}_{i=1}^{\infty}$, it can be proven that any finite variance random variable X can be expanded as [51]:

$$X = \sum_{k=0}^{\infty} \alpha_k \Psi_k \left(\xi_1, \xi_2, \ldots\right), \tag{1.25}$$

where α_k are deterministic coefficients, Ψ_k are a series of orthogonal polynomials with respect to the PDF of ξ, i.e.:

$$\langle \Psi_i \Psi_j \rangle \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Psi_i(\xi)\Psi_j(\xi)p(\xi)d\xi = \delta_{ij} \left\langle \Psi_i^2 \right\rangle. \tag{1.26}$$

If ξ is Gaussian, then Hermite polynomials should be used; if ξ is a uniform distribution, then Legendre polynomials should be used [55]. In practice, the number of random variables and degree of polynomials should be truncated within a finite number. The coefficients α_k can be determined by Galerkin projection:

$$\alpha_k = \frac{\langle X \Psi_k \rangle}{\left\langle \Psi_k^2 \right\rangle}. \tag{1.27}$$

According to the different methods of doing Galerkin projection, PCE can be divided into intrusive and non-intrusive PCE. The former first represents both model inputs and model outputs as polynomial chaos, and then substitutes both inputs and outputs into the governing equations, where α_k for inputs are known while α_k for outputs are unknown. Due to the linearity of inner product operation in Eq. 1.26, eventually the governing equation would be transformed to the inner product operation between ξ and polynomial coefficients α_k. The inner product operation between ξ can be computed in advance, and then by substituting the inner product and solving the deterministic governing equation, we can obtain α_k of outputs. Then the statistics of outputs can be directly computed from α_k. The details of this process are shown in [56]. For non-intrusive PCE, Eq. 1.27 is computed by numerical quadrature, where the forward physical model needs to be evaluated at different values of ξ [52]. In order to reduce the number of physical model evaluations, Gaussian quadrature points are usually adopted.

1.3.3 High Dimensional Model Representation

Another response surface technique steaming from the chemistry community is high dimensional model representation (HDMR) [57], [58]. Here we only give a brief overview. For details, readers should refer to [57] and [58]. The mapping function of parameters to model outputs can be decomposed into hierarchical correlation functions:

$$\begin{aligned} f(x) =& f_0 + \sum_{i=1}^{n} f_i\left(x_i\right) + \sum_{1\le i<j\le n} f_{ij}\left(x_i, x_j\right) + \sum_{1\le i<j<k\le n} f_{ijk}\left(x_i, x_j, x_k\right) + \dots \\ &+ \sum_{1\le i_1<\dots<i_l\le n} f_{i_1 i_2 \cdots i_l}\left(x_{i_1}, x_{i_2}, \dots, x_{i_l}\right) + \dots + f_{12\dots n}\left(x_1, x_2, \dots, x_n\right), \end{aligned} \tag{1.28}$$

where x is parameters, f_0 is the mean value of $f(x)$, f_i describes the contribution of each individual parameters to the variation of $f(x)$, f_{ij} describes the correlated contribution of two parameters to the variation of $f(x)$, and so on. Due to the principle of effective sparsity [23], in practice correlation functions with order higher than two are usually truncated. HDMR, from the perspective of function forms, is a generalized version of PCE, since each correlation function is not necessarily orthogonal polynomials of the given form. Different from non-intrusive PCE where physical models are evaluated at quadrature points, HDMR relies on evaluation at Monte Carlo samples and then conducts analysis of variance (ANOVA).

1.3.4 Neural Networks

In Recent years, deep learning has achieved great success in image recognition [59], natural language processing [60] and control in complex tasks [61]. Indeed, for these complex tasks without physical models, they by nature are high-dimension problems. Even for the data set of small pictures, CIFAR-10 [62], whose resolution is 32×32, the dimension is 1024. Deep learning can recognize even larger figures with good accuracy, while other machine learning algorithms cannot handle such a high-dimensional regression or classification problem. Inspired by the success of deep learning in complex tasks, many studies have been done to use neural networks (NN) as response surfaces in UQ [63], [64]. Many deep learning frameworks

for scientific computing are also developed [65], [66]. In the combustion community, Wang et al. [20] is the first work to use NN as response surfaces. They showe that multi-layer perceptron (MLP) response surfaces need fewer training samples to achieve the same accuracy compared with HDMR and PCE. Zhang et al. [67] utilize MLP as response surfaces to conduct UQ for FFCM-2 models with more than 1000 reactions. They indicate that the NN response surface can include all the parameters as inputs without doing dimensionality reduction first. They also demonstrate generalized response surfaces where experimental conditions, such as temperatures, pressures, and equivalence ratios, are also included as the inputs of response surfaces. This can avoid repeated procedures of constructing response surfaces for different conditions. Besides MLP, other types of NN are also used for combustion kinetic UQ. In our group, DeepONet-type NN is generally used for response surfaces [68], [69]. Deep operator neural network(DeepONet) has two separate nets and merge on the last layer to give the final prediction. In [68], parameters are inputted into the first net, while the mixture fraction is inputted into the second net to provide a prediction of temperature at a given mixture fraction under different kinetic parameters. In [69], non-linear independent dual systems (NIDS) [70], a structure based on DeepONet is used. Time is inputted into the second net to give a prediction of 8 species time series parameterized by kinetic parameters.

1.4 Experimental Design for Efficient Uncertainty Quantification

As is shown in Section 1.2, many algorithms have been developed to reduce the uncertainty of kinetic parameters based on experimental measurements of global combustion properties, such as laminar flame speeds, ignition delay times, species profiles, etc. Obviously, not all measurements are equally informative: some would be very helpful to reduce the posterior uncertainty, while others are not. Then an important question is how to pick these more informative experimental conditions. Empirically, we can have two simple principles to assess the amount of information provided by an experiment. First is always picking experimental conditions where the model prediction for this target is very sensitive to the parameters we want to infer. The second is picking one with small measurement uncertainty. In fact, these two principles are used by experimentalists in the kinetic community all the time. The first principle is manifested by the fact that direct measurements of rate constants are usually conducted at conditions where model prediction is very sensitive and only sensitive to one or two reaction rate constants [71]. The second principle is manifested by the persistent efforts of experimentalists to reduce the uncertainty of instruments [72] and to eliminate the interference of non-chemical factors, or non-ideal effects in measurements [73]. However, these two principles are more like empirical guidelines and cannot provide rigorous decisions when experimentalists need to pick conditions from a design space. At the same time, as an important topic of statistics, many experimental design algorithms have been developed by statisticians and computer scientists [74], [75]. An observation from these algorithms is that a rigorous, algorithmic way for design of experiments (DoE) *must* involve posterior distributions. Thus, different DoE algorithms are usually developed based on different methods of UQ. In recent years, as UQ draws more and more attention in the combustion community,

many DoE algorithms have been developed or implemented by combustion engineers [7].

Generally, experimental design includes two stages. The first is the computation of information gain for a given experimental condition, and the second is the maximization of information gain over the design space. In the following subsections, we will overview the Bayesian and non-Bayesian methods for computing / estimating information gains in Sections 1.4.1 and 1.4.2, respectively, with emphasis on the Bayesian side.

1.4.1 Non-Bayesian Approaches

Earliest attempts for DoE in the combustion community focus on sensitivity analysis, where different methods, both global and local, have been developed to give a less biased assessment of the significance of different parameters [76], [77]. Just as stated in the last paragraph, however, sensitivity analysis still lies in the category of empirical methods, since no posterior distribution is evaluated. To the author's knowledge, the earliest rigorous algorithm for combustion DoE was done by Frenklach and coworkers [78] based on B2BDC framework. In that paper, inference is done first based on a collection of data. Then the measurement uncertainty of each target is slightly perturbed to observe the sensitivity of the parameter posterior bound length to the measurement uncertainty of a certain target. A larger sensitivity indicates a larger information gain for this experiment. After that paper, as Bayesian analysis became popular in the combustion community, many researchers applied Bayesian DoE in combustion problems, which will be introduced in Section 1.4.2.

In recent years, many non-Bayesian algorithms have been developed by combustion engineers. Sheen and Manion [79] developed a method called experimental design by differential entropy (EDDI). Specifically, we have a set $\mathcal{A}$ of experimental targets where we can do experiments, and another set $\mathcal{B}$ of targets where we care about the prediction uncertainty but cannot do experiments. We start from a prior model, and generate synthetic data for these targets at set $\mathcal{A}$. Then inference is done based on these synthetic data. In [79], MUMPCE is used as the inference framework. After we get the posterior model, we can compute the differential entropy

$$S_{ij} = \frac{\partial \ln \sigma_j^*}{\partial \ln \sigma_i^{\text{obs}}}. \tag{1.29}$$

where i is the index only in set $\mathcal{A}$, j is the index in both set $\mathcal{A}$ and $\mathcal{B}$. Then the information flux for each target in $\mathcal{A}$ can be computed:

$$\Phi_r = \sum_j (S_{rj})^2 - \sum_i (S_{ir})^2, \tag{1.30}$$

where r is the index in set $\mathcal{A}$. If the smallest Φ_r in set $\mathcal{A}$ is negative, remove it from $\mathcal{A}$ and repeat this procedure to compute new Φ_r; otherwise, the current set $\mathcal{A}$ is the optimal set of experimental conditions. Turányi and coworkers [80] extended this DoE framework to the UQ methods developed by Turányi's group, and derived the analytical form of differential entropy. Recently, Yang and coworkers [81] developed an experimental platform integrating sensitivity entropy-based design to pick experimental conditions informative to kinetic UQ [82], and similarity-based design to substitute conditions where experiments are hard to

conduct by these easy conditions [83], [84]. Inspired by the entropy of a discrete random variable, sensitivity entropy is defined by

$$\mathrm{HS} = -\sum S_{i_1,\ldots,i_s} \log_e \left(S_{i_1\ldots i_s}\right), \quad 1 \le i_1 < \cdots < i_s \le n \tag{1.31}$$

where $S_{i_1\ldots i_s}$ is the Sobol index in global sensitivity analysis. Similar to random variables, the summation of all Sobol index should be 1, and hence entropy can be generalized to sensitivity. Low sensitivity entropy means the target is only sensitive to a few reactions. As for similarity-based design, cosine similarity of Sobol index [83] or active subspace [84] at two designs is computed. Cosine similarity near 1 means that the two targets constrain the same set of kinetic parameters, and hence one condition where experiments are hard or expensive to conduct can be replaced by the other one. These methods, although are more empirical and lack the guidance of rigorous theory, show good performance in numerical experiments, and meet our intuition.

1.4.2 Bayesian Approaches

Since the second part of this thesis will focus on Bayesian DoE, in this section, we provide a more comprehensive introduction of Bayesian DoE following Huan and Marzouk [85], and then introduce some applications of Bayesian DoE in combustion kinetic problems. In [85], expected information gain is estimated using Monte Carlo sampling without any assumption, while the Bayesian design in the combustion community usually uses a simplified version to reduce computational cost. Since a rigorous DoE algorithm involves the evaluation of posterior uncertainty, different DoE algorithms can be developed based on different methods of Bayesian inference.

Overview

The Bayes Theorem with design $\boldsymbol{d}$ can be written as

$$p(\boldsymbol{k}|\boldsymbol{y},\boldsymbol{d}) = \frac{p(\boldsymbol{y}|\boldsymbol{k},\boldsymbol{d})p(\boldsymbol{k})}{p(\boldsymbol{y}|\boldsymbol{d})} = \frac{p(\boldsymbol{y}|\boldsymbol{k},\boldsymbol{d})p(\boldsymbol{k})}{\int_{\boldsymbol{K}} p(\boldsymbol{y}|\boldsymbol{k},\boldsymbol{d})p(\boldsymbol{k})d\boldsymbol{k}}, \tag{1.32}$$

Here we assume that the prior distribution is irrelevant to $\boldsymbol{d}$. Kullback–Leibler (K-L) divergence is generally used to measure the difference between two distributions from information-theoretic perspective. In Bayesian DoE, K-L divergence of prior and posterior distributions is used to measure the information gain:

$$D_{\mathrm{KL}}\left[p(\boldsymbol{k}|\boldsymbol{y},\boldsymbol{d})\|p(\boldsymbol{k})\right] = \int_{\boldsymbol{K}} p(\boldsymbol{k}|\boldsymbol{y},\boldsymbol{d}) \ln\left[\frac{p(\boldsymbol{k}|\boldsymbol{y},\boldsymbol{d})}{p(\boldsymbol{k})}\right] d\boldsymbol{k}, \tag{1.33}$$

where $\boldsymbol{K}$ is the support of kinetic parameters $\boldsymbol{k}$. K-L divergence is always non-negative, and it equals to zero if and only if two distributions are identical. The larger the K-L divergence is, the larger the difference between two distributions. Since we cannot know the experimental data $\boldsymbol{y}$ *a priori*, we can weight the K-L divergence by the PDF of $\boldsymbol{y}$, leading to *expected information gain* (EIG):

$$EIG(\boldsymbol{d}) = \int_{\boldsymbol{Y}}\int_{\boldsymbol{K}} p(\boldsymbol{y}|\boldsymbol{d})p(\boldsymbol{k}|\boldsymbol{y},\boldsymbol{d}) \ln\left[\frac{p(\boldsymbol{k}|\boldsymbol{y},\boldsymbol{d})}{p(\boldsymbol{k})}\right] d\boldsymbol{k}d\boldsymbol{y}. \tag{1.34}$$

By rearranging different terms in Eq. 1.32 and substitute into 1.34, we can transform both the part within the logarithm and out of the logarithm, obtaining

$$EIG(\boldsymbol{d}) = \int_{\boldsymbol{Y}} \int_{\boldsymbol{K}} \ln \left[\frac{p(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{d})}{p(\boldsymbol{y}|\boldsymbol{d})} \right] p(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{d}) p(\boldsymbol{k}) d\boldsymbol{k} d\boldsymbol{y}. \tag{1.35}$$

The goal of Bayesian DoE is to find the optimal experimental condition:

$$\boldsymbol{d}^* = \arg\max_{\boldsymbol{d} \in \mathcal{D}} EIG(\boldsymbol{d}), \tag{1.36}$$

where $\mathcal{D}$ is the design space.

Full Bayesian Design

This method is based on the Monte Carlo estimation of evidence function (i.e., $p(\boldsymbol{y})$). The integral in Eq. 1.35 can be estimated by a double-loop Monte Carlo simulation

$$EIG(\boldsymbol{d}) \approx \frac{1}{n_{\text{out}}} \sum_{i=1}^{n_{\text{out}}} \left\{ \ln \frac{p\left(\boldsymbol{y}^{(i)}|\boldsymbol{k}^{(i)}, \boldsymbol{d}\right)}{\frac{1}{n_{\text{in}}} \sum_{j=1}^{n_{\text{in}}} p\left(\mathbf{y}^{(i)}|\boldsymbol{k}^{(i,j)}, \boldsymbol{d}\right)} \right\}, \tag{1.37}$$

where the n_{in} and n_{out} are the numbers of samples for the inner loop and outer loop, respectively. The inner loop is used to compute $p(\boldsymbol{y}|\boldsymbol{d})$. $(\boldsymbol{k}^{(i)}, \boldsymbol{y}^{(i)})$ are drawn from $p(\boldsymbol{k})p(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{d})$, while $\boldsymbol{k}^{(i,j)}$ for a given i are drawn from $p(\boldsymbol{k})$. This method can be considered as associated with MCMC algorithm for Bayesian inference. Huan and Marzouk [85] tested this algorithm in the inference of 2 kinetic parameters of a hydrogen combustion model, and targets are ignition delay time, H atom peak time and peak concentration. Numerical results showed that the posterior after experiments in the high EIG area is obviously narrower than that in the low EIG area.

However, the computational cost of double-loop Monte Carlo is very large. Rainforth et al. [86] have proved that this nested Monte Carlo estimator has error scales with computational cost with power $\frac{-1}{3}$. In the combustion community, the simplified version of full Bayesian DoE is implemented.

Bayesian Design Using Laplacian Approximation

As we introduced in Section 1.2, Laplacian approximation is commonly used to approximately compute the posterior distribution. There are two algorithms of DoE based on Laplacian approximation. A relatively sophisticated version was developed by Long et al. [87]. Specifically, the posterior is approximated by linearizing physical models, so that the posterior distribution is Gaussian and the K-L divergence in Eq. 1.34 can be derived analytically. Then the integral weighted by $p(\boldsymbol{y})$ is computed by Gaussian quadrature. In the combustion community, a further reduced version is applied.

Kraft and coworkers [88] utilize the Fisher information matrix to design optimal operation conditions of a compression ignition engine so that kinetic parameters can be efficiently inferred. The Fisher information matrix is defined as:

$$F(\boldsymbol{k}, \boldsymbol{d}) = E_{\boldsymbol{y}} \left[\left(\nabla_{\boldsymbol{k}} \log p(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{d}) \right) \left(\nabla_{\boldsymbol{k}} \log p(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{d}) \right)^{\top} \right]. \tag{1.38}$$

When Gaussian likelihood is used, we can get that (for details, see Eq. 14 of [88]):

$$F(\boldsymbol{k}, \boldsymbol{d}) = J(\boldsymbol{k}, \boldsymbol{d})^T \Sigma^{-1} J(\boldsymbol{k}, \boldsymbol{d}). \tag{1.39}$$

where $J(\boldsymbol{k}, \boldsymbol{d})$ is the Jacobian matrix of the multi-input, multi-output model, Σ is the covariance matrix of measurement noise. Note that this is pretty similar to the covariance matrix in Eq. 1.23, where the only difference is that Eq. 1.23 assumes independent noise so the covariance matrix can be written as the denominator. They used D-optimality to define optimal design, where the optimal design gives maximum determinant of the Fisher information matrix. This is equivalent to minimizing the determinant of the posterior matrix under a uniform prior and linearized response surface. vom Lehn et al. [89] also utilized a D-optimality approach based on linearized response surfaces but based on Gaussian priors. The idea is that for each design point, nominal prediction is directly used as $x^{(0)*}$ and Eq. 1.24 is used to estimate potential posterior covariance matrix. The goal is to find the design with the minimum determinant of the potential posterior covariance matrix. Similarly, Wang et al. [8] assumed that both kinetic uncertainty and systematic errors would linearly impact the measurement, so that the information gain can be analytically derived using the Jacobian matrix.

The difference between the reduced version described in the last paragraph and the sophisticated version [87] is that the reduced version uses the posterior when $\boldsymbol{y}$ is exactly the model prediction at nominal parameter values to compute the information gain, while the complex version integrates the information gain with respect to all the possible experimental measurement. In fact, it can be proved that for a Bayesian linear model with Gaussian prior Σ_0 and Gaussian noise $\epsilon \sim N(0, \Sigma)$, where measurement is given as $\boldsymbol{y} = X\boldsymbol{k} + \epsilon$, $\ln[det\left(\Sigma_0^{-1}\right) / det(X\Sigma^{-1}X^T + \Sigma_0^{-1})]$ is same as full Bayesian DoE given by Eq. 1.35 [90]. Thus, as long as the physical model is linear enough around the nominal value, linear approaches of DoE should be close to full Bayesian design.

Chapter 2

Utilization of Species Profiles in Bayesian Inference

2.1 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 utilize state-of-the-art laser absorption techniques to make multi-species time-histories data available [91], which is shown to provide rich kinetic information. Wang and Sheen [92] shows that the multi-species time-histories can well constrain the prior model and lead to notable improvement in prediction precision in both species time-histories and global combustion data, such as laminar flame speeds. Since then, many species profile data is reported by experimentalists [8], [93], [94] and utilized in combustion kinetic UQ [92], [95]–[98].

However, there are still two problems that have not been answered in the utilization of species time-histories. The first problem lies in the target selection: some works only pick representative points to conduct UQ [92], [95], while others use all the data points in a curve [8], [96]–[98], but there is no comparison of the approaches in the literature on their influences on kinetic parameter estimation. The UQ results shows that both two approaches can make the posterior predictions cover the experimental data, or at least closer to experimental data compared with the prior model. Sometimes even only picking several points can make the posterior prediction of the entire curve fit the data well [92]. A potential argument for this observation is that, the physical model imposes constraints on the shape of species profiles so that by perturbing kinetic parameters, the curve usually will be scaled up / down along horizontal or vertical directions, and only several points would be enough for determining the entire curve. In this work, we will compare the results of Bayesian inference using both representative points and all points in species profiles. The argument above is also tested by the assessment of sensitivity directions.

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. In some works, the data is normalized by the number of data points to avoid overfitting [8], [97], [98]. For example, in [97], an objective function is minimized to obtain the optimal kinetic parameters:

$$E(\boldsymbol{k}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\frac{\boldsymbol{y_{ij}} - \boldsymbol{M_{ij}}(\boldsymbol{k};\boldsymbol{d_i})}{\sigma\left(\boldsymbol{y_{ij}}\right)}\right)^2, \tag{2.1}$$

where i is the index for data batches, and j is the index of data point, N is the number of data batches, and N_i is the number of data points in batch i. In the case of shock tube species time-histories, the "batch" refers to each curve, while the data points refer to all the points in a single curve. This is the typical way of normalizing all data points by the number of points in each curve. In [98], however, 10 points are artificially assigned to each OH profile are used. They claimed that 10 points should be enough to capture a curve, and a further increase in the number of data points does not lead to the improvement of inference. This can be considered as utilizing all data points in a curve, although they assume the number of points. In this work, we also compare the Bayesian approaches with and without normalization with respect to the number of data points.

In this chapter, we conducted Bayesian inference facilitated by adopting neural networks as response surfaces. We obtain the optimal parameters, i.e., maximum a posterior (MAP) estimation, by maximizing the numerator part of Bayes' law. Then MCMC sampling is done to draw samples from the posterior distributions. The elucidated influences of the target selection and the number of independent effective data will guide experimentalists and modelers in generating and adopting data for model development. This chapter is a modified version of [69].

2.2 Methods

To investigate the influence of target selection on kinetic model optimization, we combine deep neural networks and Bayesian inference to optimize a propane pyrolysis mechanism against species histories in shock tube measurements [94]. The trial mechanism is a C_3 sub-mechanism extracted from USC-Mech II [10] with 111 elementary reactions and 27 species. In the shock tube experiments previously reported in [94], 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 three steps: first, the neural network response surfaces model representing the response of species evolution to the perturbation of pre-exponential factors is trained; then, the numerator part of Bayes formula is maximized with the previous neural networks to obtain the MAP estimation; finally, Hamiltonian Monte Carlo algorithm [42] (HMC), a variation of the original MCMC algorithm in Pyro package [99] is used to draw samples of posterior distributions.

2.2.1 Bayesian Inference

As introduced in Section 1.2, Bayesian analysis has been generally used in the UQ of combustion kinetics. In this chapter, we first obtain the MAP estimation by an optimization approach. Notice that in Eq. 1.17, the denominator part is independent of $\boldsymbol{k}$. We only need to maximize $p(\boldsymbol{y}|\boldsymbol{k})p(\boldsymbol{k})$ and the corresponding $\boldsymbol{k}$ is the MAP estimation. 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 $p(\boldsymbol{k})$ are parameterized as Eq. 1.5, where $x_i \sim N(0, 0.25)$. The prior distribution $p(\boldsymbol{k})$ is mutually independent, so $p(\boldsymbol{k}) = \prod p(k)$. It is generally assumed that the measurement noise is independent of each other [19], [20], so $p(\boldsymbol{y}|\boldsymbol{k}) = \prod p(y|\boldsymbol{k})$. We can construct a loss function by taking the natural logarithm of $p(\boldsymbol{y}|\boldsymbol{k})p(\boldsymbol{k})$, canceling the constant term, and taking the negative of the result:

$$\text{Loss} = \sum_S \left\{ \frac{1}{2\sigma_i^2} \sum_T \left[\frac{n}{N} \sum_N \left(M(\boldsymbol{x}, \boldsymbol{d}) - y \right)^2 \right] \right\} + \sum_{j=1}^{111} 2x_j^2. \tag{2.2}$$

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 $M(\boldsymbol{x}, \boldsymbol{d})$ and y are the mole fractions predicted by the model and of shock tube measurement, respectively. To reduce the computational cost, $M(\boldsymbol{x}, \boldsymbol{d})$ is calculated by neural networks. σ_i is the experimental uncertainty for species i, and j is the index of pre-exponential factors to be tuned. x_j is defined by Eq. 1.5. n is the number of effective independent-data points, which should be the same as N, if we take the common approach to calculate $p(\boldsymbol{d}|\boldsymbol{k})$ with multiple experimental data, i.e., multiplying them together. The loss function is minimized to obtain the MAP estimation of kinetic parameters. 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 Section 2.3.2. Another approach is to assume the effective number of independent data points is 1 in each species profile, similar to the approach described in [97], 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 [8]. In fact, moving n into the denominator part as N/n, and merging with σ_i, different choices 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

$$\text{Loss} = \sum_S \left\{ \frac{1}{2\sigma_i^2} \sum_T \left(M(\boldsymbol{x}, \boldsymbol{d})_{1.6} - y_{1.6} \right)^2 \right\} + \sum_{j=1}^{111} 2x_j^2. \tag{2.3}$$

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

We adopt Adam optimizer [100] in PyTorch environment [101], developed to optimize large-scale deep neural networks, to find the minimum value of loss function and the optimal

model parameters, due to its adaptive step size. Here the step size "lr" is fixed at 0.01. The initial point of optimization is the parameter values in the original mechanism, i.e. $\ln(k_i/k_{i,0}) = 0$. Since we have replaced the physical model with neural networks, the gradient 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 Fig. B.6 of Additional Figures.

Then we draw samples from posterior distributions using Hamiltonian Monte Carlo algorithm [42], which is designed to reduce the random walks of samples in traditional Metropolis–Hastings algorithm. The HMC package in Pyro [99] is used combined with neural network response surfaces introduced in Section 2.2.2. The posterior distribution of the case with $n = 160$ is too narrow to sample from, so we only compare the posterior distributions of $n = 1$ case and the case utilizing only the last point of each curve. We obtained three chains for each case, and each chain contains 15,000 samples after 5,000 burn-in samples. We do Gelman-Rubin test [102] and check the auto-correlation of each chain for the diagnostics of convergence. The results of auto-correlation are shown in Figs. 2.1 and 2.2. The auto-correlation is computed by

$$C_i(\tau) = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left(\theta_{i,t} - \overline{\theta_i}\right) \left(\theta_{i,t+\tau} - \overline{\theta_i}\right), \tag{2.4}$$

where τ is the lag, i is the index for different coordinates (i.e., parameters), $\theta_{i,t}$ is the value of the i-th coordinate and the t-th sample, and $\overline{\theta_i}$ is the mean value of coordinate i over the entire chain. The autocorrelation package of pyro is used to compute the auto-correlation here. This quantity measures the correlation of two samples with a certain lag. As is shown in Fig. 2.1, for the case of utilizing the entire curve with $n = 1$, the auto-correlation decays to zero at around 50 lag, which means that the number of effective independent samples is $3\times15000/50 = 900$. From Fig. 2.1, it is shown that for the case of utilizing only the last point, the auto-correlation decays to zero at around 5 lag, and the number of effective independent samples is $3 \times 15000/5 = 9000$. A reference of effective independent samples is 500 in [103], so we have got enough number of samples to represent the posterior distributions. As for Gelman-Rubin test [102], we compute R using the package in Pyro. The maximum R among all coordinates for the entire curve case is 1.0049, while for the last point case it is 1.0005. The criterion of convergence is usually $R < 1.1$, so our samples already achieve convergence.

2.2.2 Neural Network Response Surfaces

Since model optimization and uncertainty quantification usually repeatedly evaluate the physical model, it is usually combined with response surfaces, which can relate model parameters to the model prediction. Response surfaces that are commonly used in the combustion community include polynomial chaos expansion [35] and neural networks [20], [67]. Here, we train deep 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

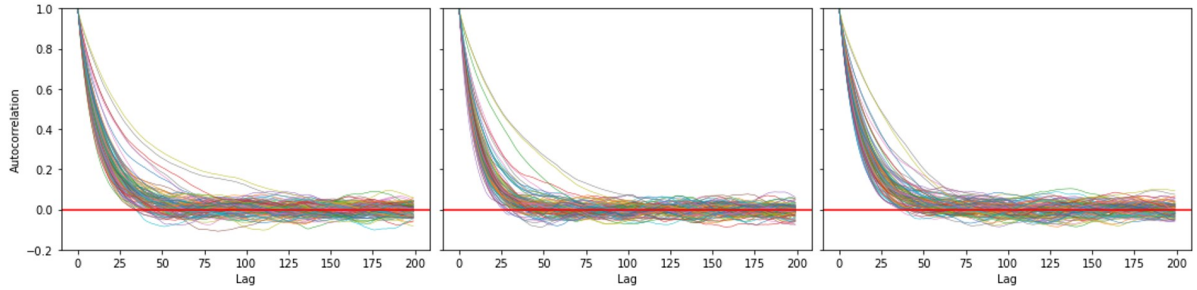


Figure 2.1: The auto-correlation of three chains for the case of utilizing the entire curve with $n = 1$. The red horizontal line indicates zero auto-correlation. Different colors represent different coordinates.

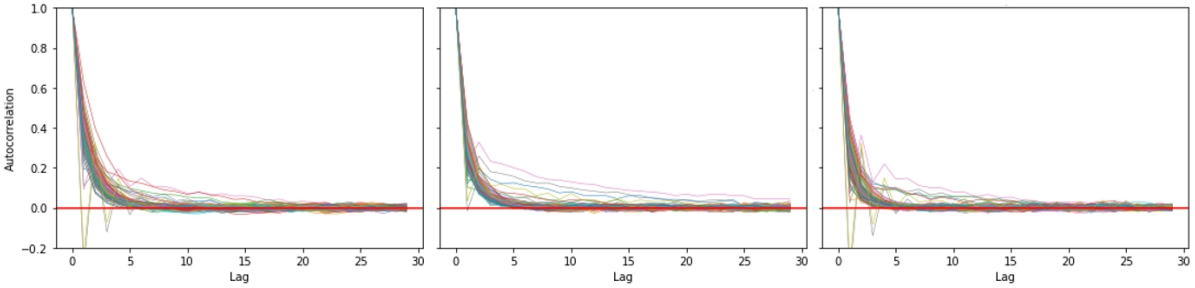


Figure 2.2: The auto-correlation of three chains for the case of only utilizing the last point. The red horizontal line indicates zero auto-correlation.

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 [20]. 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 obtaining 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. 2.3, is inspired by the recently developed Deep Operator Neural Network [65] 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. 2.3. It is worth noting that we train 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. Adam optimizer [100] is used to train the neural network till convergence, and the learning rate is 0.01.

For each network, the training data are 5000 samples with different model parameters generated by Cantera [104] according to the experimental conditions using the C_3 sub-

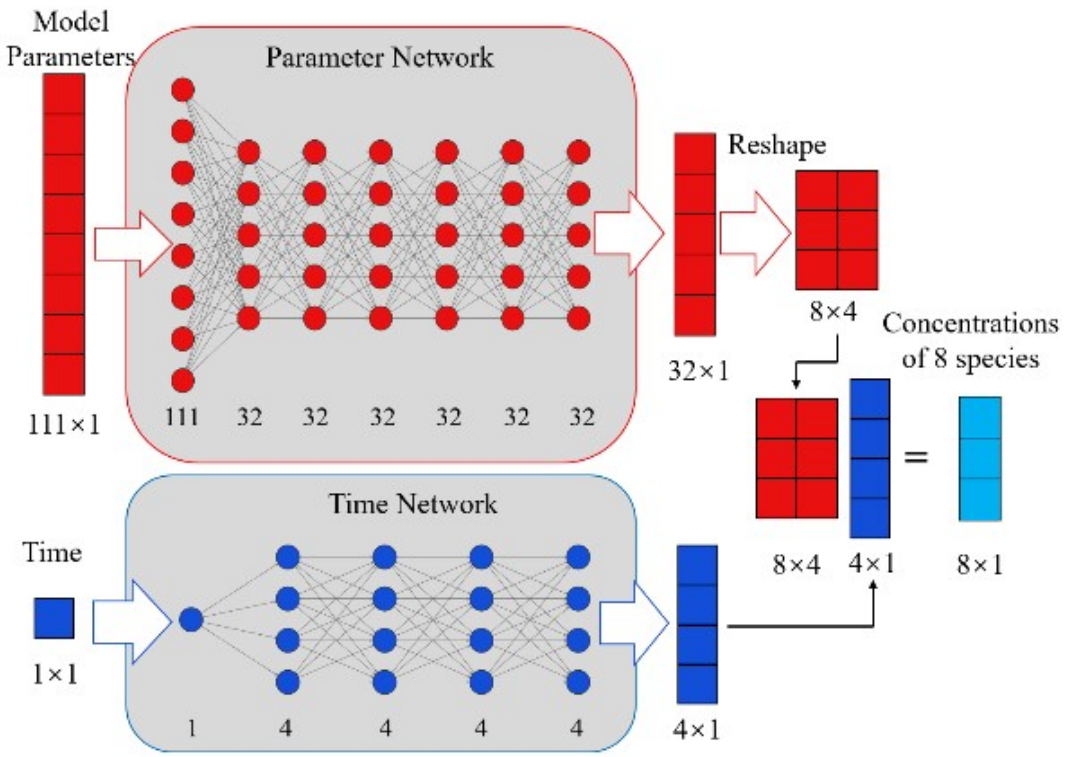


Figure 2.3: The structure of the neural network used as surrogate models. The number of neurons is labeled under each layer.

mechanism. The input parameters, $\ln(k_i/k_{i,0})$, are randomly sampled by Latin hypercube sampling [105] with bounds of $\pm \ln f_i$. Our validation shows that neural networks can predict the species concentrations and the gradients of concentration to input parameters accurately. The validation results can be found in Figs. B.1, B.2, B.3, B.4 and B.5 in Additional Figures, from which we can find that the prediction error is relatively small, compared with the experimental uncertainty shown in the shadow region. We also validate the gradient computation of the network, as shown in Tables A.1 and A.2 in Additional Tables.

2.3 Results and Discussion

2.3.1 Comparison of Optimized Mechanisms with the Original Mechanism

Three approaches are 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 define a normalized error (NE) for a given species profile:

$$\mathrm{NE} = \frac{1}{N\sigma_i} \sum_{N} |M(\boldsymbol{x'}, \boldsymbol{d}) - y| \,, \tag{2.5}$$

where $\boldsymbol{x'}$ is the optimized parameters, $N = 160$ is the number of data points in each profile and σ_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):

$$\mathrm{NE} = \frac{1}{\sigma_i} \left| M(\boldsymbol{x'}, \boldsymbol{d})_{1.6} - y_{1.6} \right| . \tag{2.6}$$

We then average 40 NELPs together for each mechanism. The averaged NEs and NELPs of three optimized mechanisms and the initial mechanism are shown in Table 2.1. In Table 2.1 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 2.1, we can see significantly improved agreement with the measurements with all three optimization approaches.

Table 2.1: 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

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 Tables A.3, A.4 and A.5. 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.

2.3.2 The Influence of Effective Independent-Data Numbers

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 2.1, 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 2.2, 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 predictions of the original model, C-1 and C-160 at 1250 K are compared against experimental data in Fig. 2.4. Similar to the results indicated by Table 2.2, both C-1 and C-160 are closer to experimental measurements, with C-160 slightly better than C-1, especially for CH_4. For cases with differences larger than 10% of experimental uncertainty (in bold in Table 2.2), the curves of measurements with shadows indicating the uncertainties and predictions with C-160 and C-1 are summarized in Table B.7 for visualization of this discussion.

Table 2.2: Differences between the NEs of C-1 and those of C-160. Differences larger than 10% of the corresponding experimental uncertainties are in bold

Species/Temperature	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	-8×10^{-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	-3×10^{-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

The reason why the overall performance of C-160 is better than C-1 is that $p(\boldsymbol{d}|\boldsymbol{k})$ 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 $p(\boldsymbol{d}|\boldsymbol{k})$ to make the prediction closer to experimental measurements. On the contrary, the C-1 approach needs to keep a balance between $p(\boldsymbol{d}|\boldsymbol{k})$ and $p(\boldsymbol{k})$, 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_i) in each mechanism and the corresponding x_i are shown in Fig. 2.5. x_i 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 2.1, 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.

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 $|x_i|$ and $||\boldsymbol{x}||_2$, respectively.

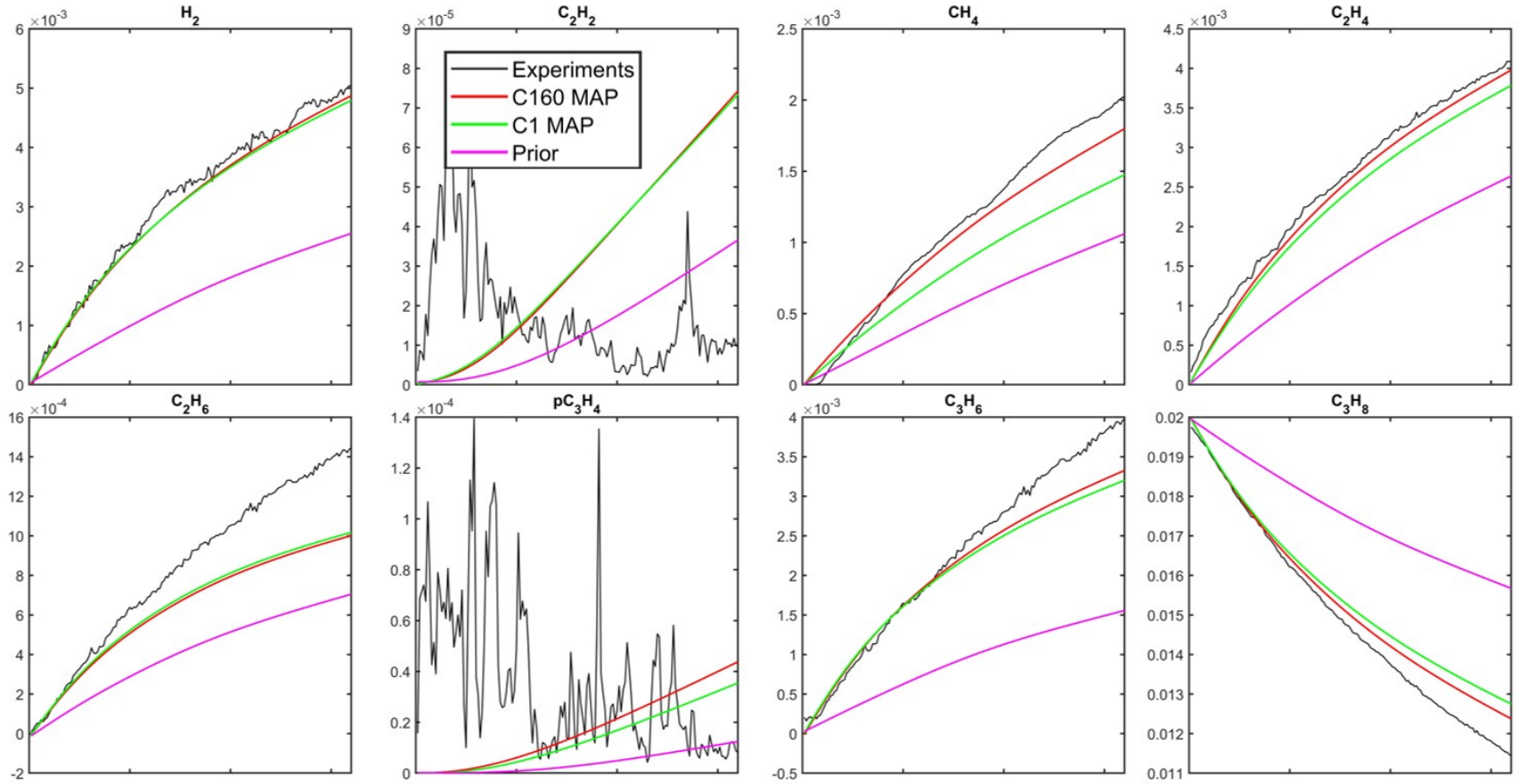


Figure 2.4: The comparison of the MAP predictions of C-1, C-160, original model and experimental data.

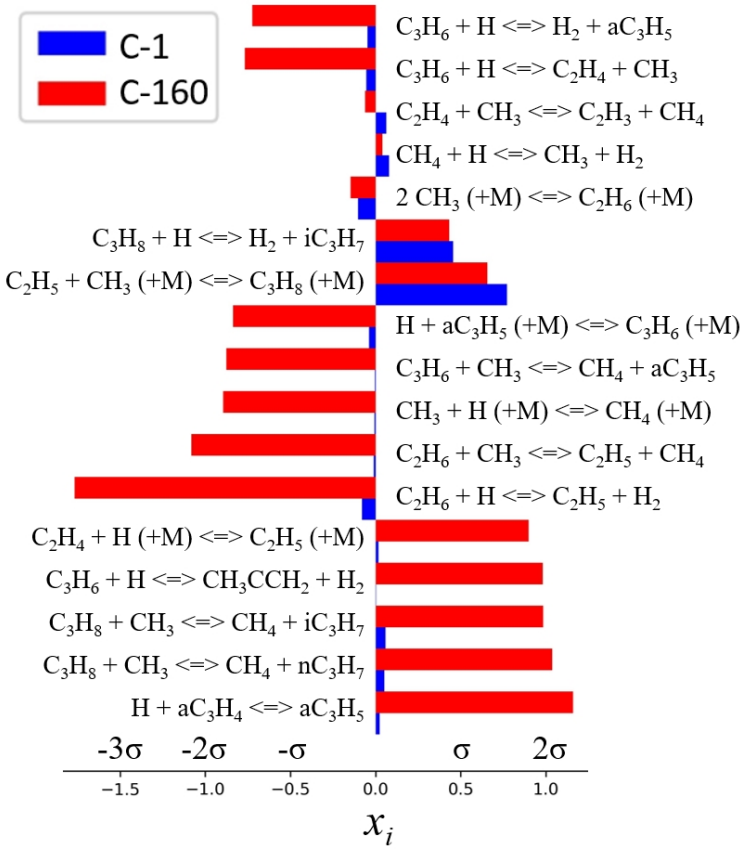


Figure 2.5: The x_i of the most tuned reaction rates in C-1 and C-160.

Table 2.3: Influence of the number of effective independent data on $|x_i|$. Maximum $|x_i|$ near 2σ and 3σ are in bold.

n	1	10	30	50	80	160
max $\|x_i\|$	0.77	0.76	**0.96**	1.24	**1.49**	1.77
$\|\|\boldsymbol{x}\|\|_2$	0.92	1.31	2.17	2.69	3.20	3.96

As shown in Table 2.3, when n increases, max $|x_i|$ increases accordingly when n is beyond 10. Furthermore, the max $|x_i|$ 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. 2.2 and 2.3 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) [29], 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 Section 2.1, 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.

The essence 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 [106], [107], we can express the discrepancies between model predictions and measurements at data point i as:

$$y_i = f_i(\boldsymbol{k}) + \varepsilon_{mi} + \varepsilon_{di} = g_i + \varepsilon_{di} \tag{2.7}$$

where y_i is an experimental measurement, $f_i(\boldsymbol{k})$ is the model prediction given parameter $\boldsymbol{k}$, 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 [107]. 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 [106]. 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 [8], 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 [108], 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 prediction corresponds to large changes of parameters. One way to solve ill-posed inverse problem is to introduce regularization of parameters [108], which is generally same as imposing prior information. Gaussian prior is equivalent to l_2 regularization [38], as clearly shown in Eqs. 2.2 and 2.3. We verify the ill-posedness of this problem by generating five sets of synthetic data using nominal parameters plus Gaussian noise. Note that by generating data in such a way, there is no model error and systematic error. The MAP of these five sets of data is shown in Fig 2.6. It is obvious that the inferred parameters are different when different noisy data is used, although these data sets share the same "ground truth". For example, the last fifth parameter inferred from the first and the fifth data set have a difference of up to one σ, where σ is the standard deviation of the prior.

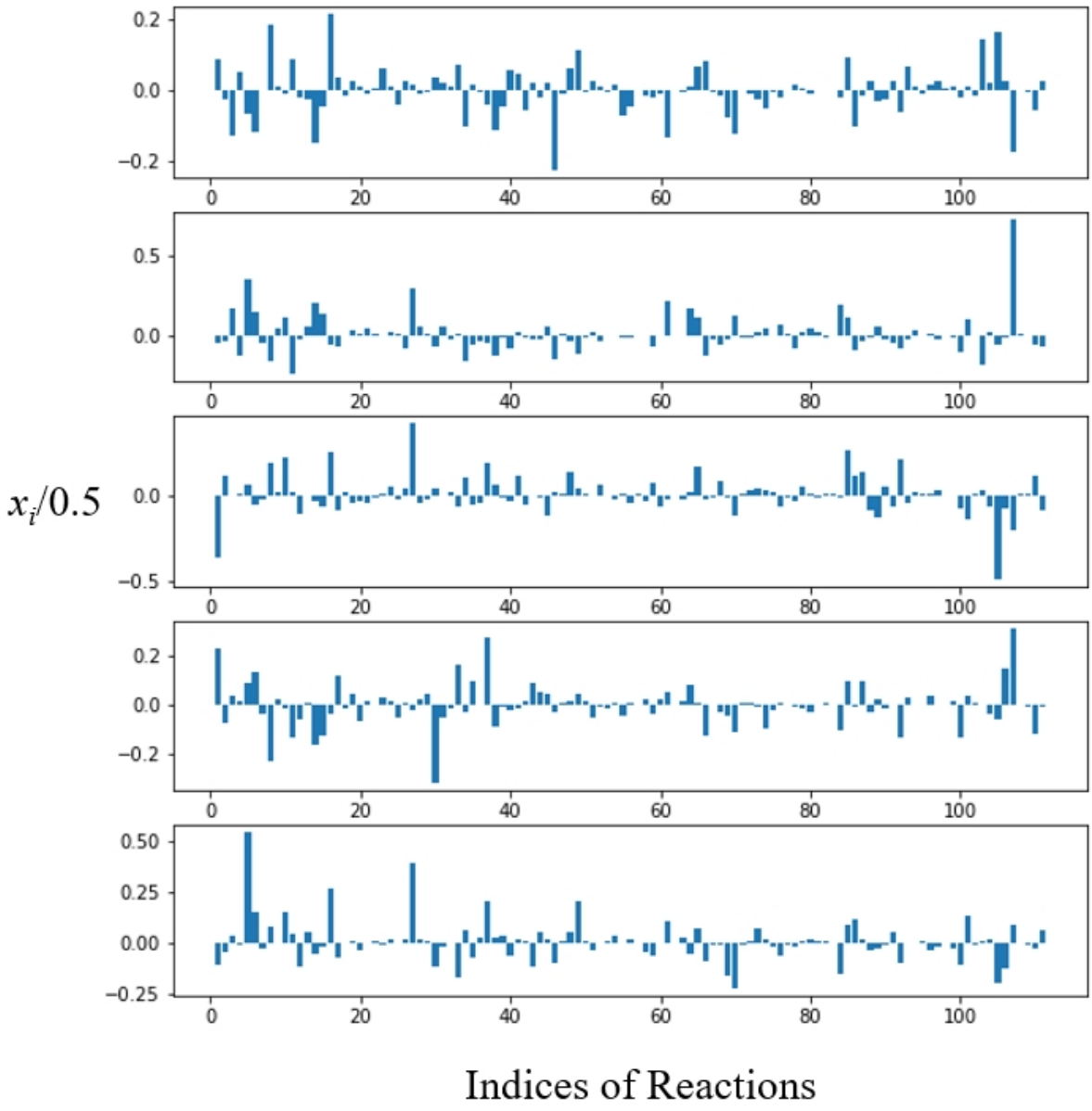


Figure 2.6: The inferred parameters of five synthetic data sets sharing the same ground truth but with different Gaussian noise.

In Bayesian analysis against batch-wise data, such as laminar flame speed and ignition delay time, the drawback of such a mis-specification is not obvious, because the experimental data are not too many and correlation between systematic errors is not very strong. 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. A theoretical, rigorous treatment of this problem would involve calculating the likelihood function using a non-diagonal covariance matrix to describe the correlation of data points [109]. For example, [8] assumed Gaussian uncertainty in initial temperature, pressure and composition of shock

tube pyrolysis experiments, and derive the covariance matrix of all points in a measurement curve by linear approximation. However, [8] also pointed out that even for random noise they can correlate to each other. In addition, there are some model and systematic errors that cannot be identified. Thus, choosing a suitable n is still needed. 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 [19]; (c) using the model comparison method described in [19] 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. 2.7, 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. Theoretically the optimal n should be the one with maximum curvature. This method is called L-Curve method [110].

There are two noteworthy points. First, we only show the point estimation here (MAP), without sampling the posterior distribution since the posterior is too narrow to be sampled from. 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. 2.8 shows a schematic of the posterior distribution of two strategies, together with the ideal case of C-160 with perfect well-posed 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 [19], [20], 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 from the perspective of point estimation. In this case, the ill-posedness may be very several and need careful assessment.

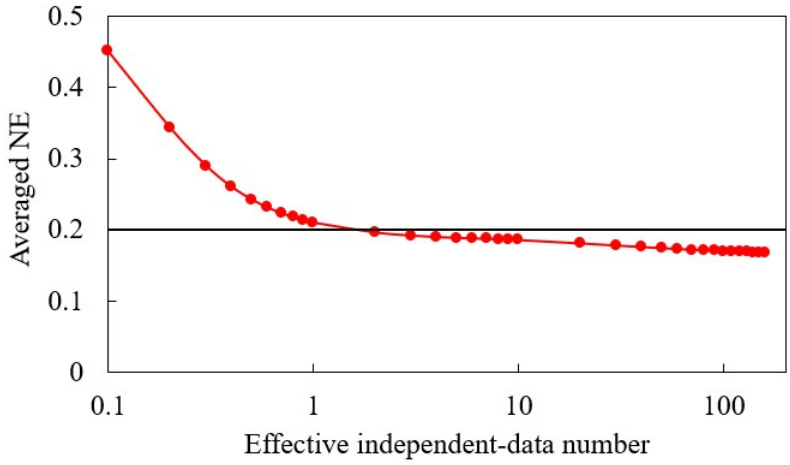


Figure 2.7: The value of averaged NE for different effective independent-data number n.

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; ill-posedness of the model-data system needs careful assessment. For experimentalists, providing a detailed probability density of experimental error, especially systematic error, would be very helpful for model calibration. If the system is well-posed, and we can know the accurate likelihood function without model error, 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 in Fig. 2.8.

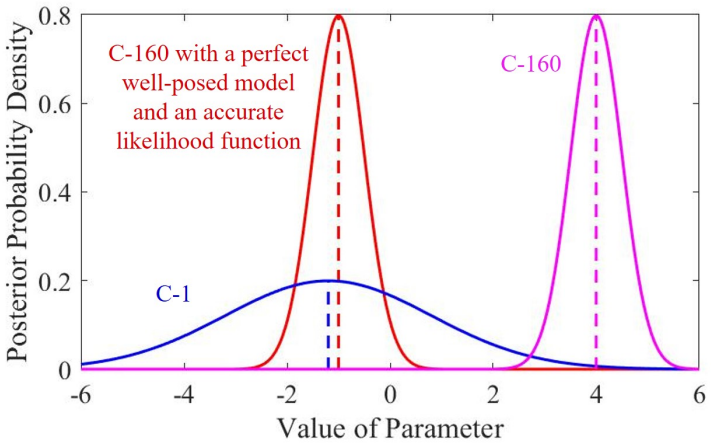


Figure 2.8: The schematic of posterior distribution of C-1 and C-160. C-160 with a perfect well-posed model and an accurate likelihood function is used as an ideal case of Bayesian inference. The dashed vertical lines represent MAP point estimations.

2.3.3 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 [92], [95], and using all data points in the entire curve [96]–[98]. A comparison of these two approaches will be discussed in this subsection. 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 2.1, 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 2.4, 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 B.8 of Additional Figures.

We also selected the most changed rate parameters (measured by the absolute value of x_i) in each mechanism, and compared the x_i in Fig. 2.9. The x_i 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 1 σ. In a nutshell, the performance and updated kinetic parameters are similar for LastP and C-1.

Table 2.4: Differences between the NEs of LastP and those of C-1. Differences larger than 10% of the corresponding experimental uncertainties are in bold.

Species/Temperature	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	3×10^{-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	$2E\times10^{-5}$	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

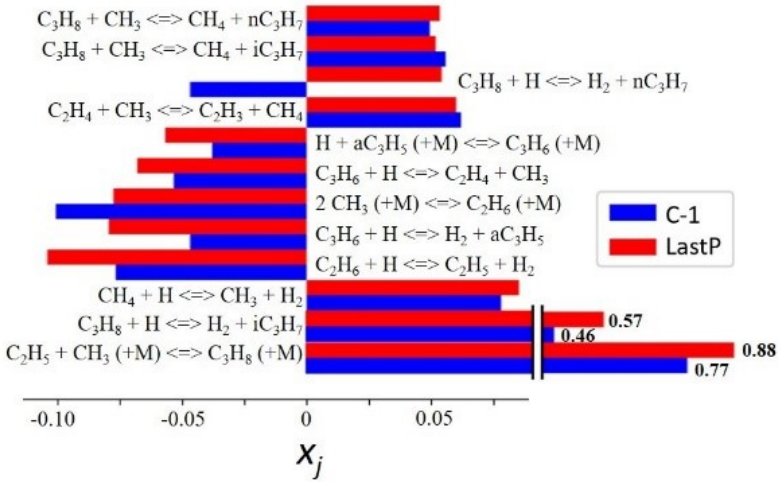


Figure 2.9: The x_i of top 10 highly-tuned reaction rates in C-1 and LastP.

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 2.1, 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 2.3.2, 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. 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.

However, from the perspective of uncertainty quantification, using representative points cannot be equivalent to using all points of the curve. Fig. 2.10 shows the kernal density estimation of posterior distribution for picked parameters obtained by MCMC samples. It is clearly shown that, although the predictions and parameter MAP are pretty similar for C-1 and LastP, the posterior uncertainties of parameters are quite different. For LastP, except two parameters, all other parameters almost follow the same distributions as priors. For C-1, however, the posteriors of many parameters are deviated from their priors. The posteriors of C-1 is generally narrower than LastP. Another phenomenon is that for C-1, the posterior mean values (the center of contours) are usually quite different from MAP values (shown by red dots). The MAP values of C-1 are usually similar to these of LastP and priors, while the posterior mean values are deviated from priors. A comparison of the predictions of MAP and posterior mean parameters are shown in Fig. B.9, which indicates that the predictions for MAP and posterior mean parameters only have slight difference. For a high-dimensional joint distribution, it is possible that the MAP estimation is different from the posterior mean. The posterior predictions are also shown in Fig. 2.11. Despite the similarity of predictions by posterior mean parameters of C-1 and LastP, their prediction uncertainties are very different. The prediction uncertainty of C-1 is greatly smaller than that of LastP, while both C-1 and LastP have smaller prediction uncertainties than the prior model. A potential explanation is that C-1 imposes more constraints for parameters since it includes the entire curve, while LastP only constrain the predictions at the last point. For these parameters that can give close prediction to experimental data on the last point, but cannot on other part of the curve, they are "admitted" in the LastP approach but "rejected" in the C-1 approach. However, one phenomenon that cannot be explained is that the uncertainty of LastP in the last point is still larger than that of C-1, which needs further analysis in the future. The uncertainty analysis shows that, although from the perspective of point estimation entire curve may be able to represented by several points, using entire curves of species concentration measurements is still helpful for the reduction of 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 [111], 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

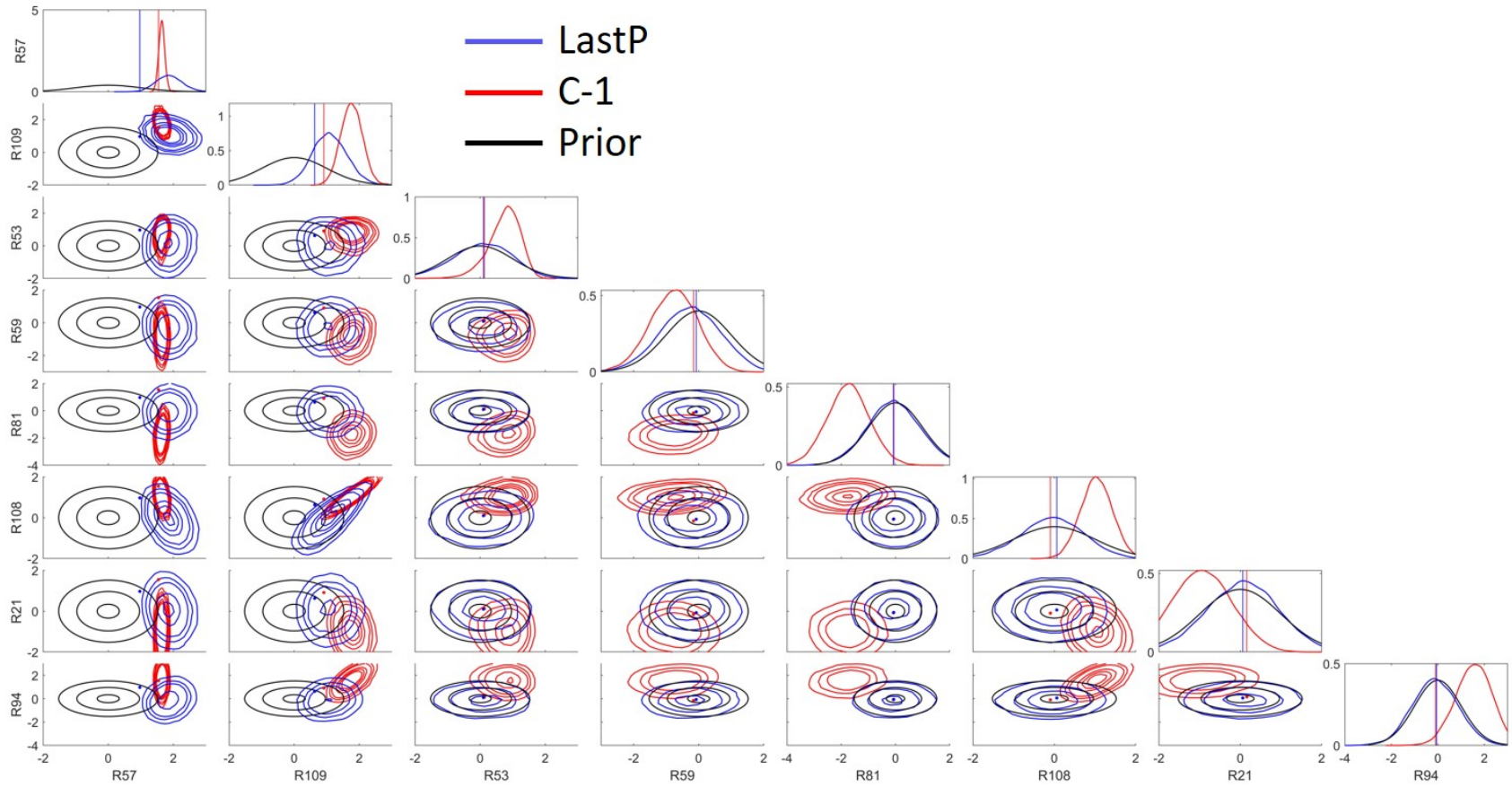


Figure 2.10: The comparison of posterior distributions of eight selected parameters in C-1, LastP and prior models. The contours of distributions are plotted at the levels of 0.05, 0.1, 0.15, 0.3, 0.4 and 0.5 for 2-D joint probability density functions. Blue and red dots/vertical lines show the MAP estimation for C-1 and LastP obtained by optimization.

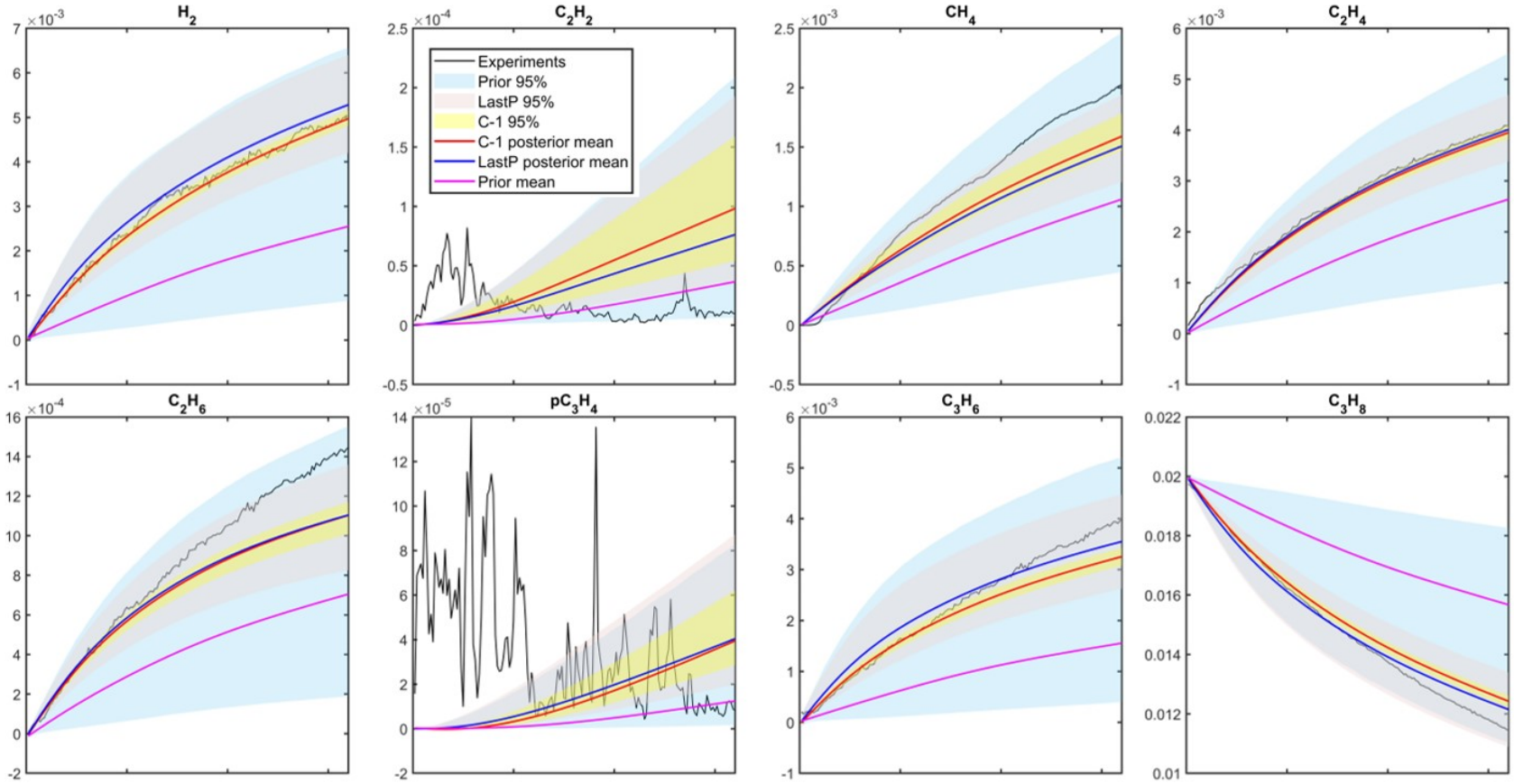


Figure 2.11: The comparison of posterior predictions of eight species at 1250K for C-1, LastP, and prior models.

Table 2.5: 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.

Species/Time	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

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 2.5, while the results for other temperatures are attached in Tables A.6, A.7, A.8, A.9 in Additional Tables. 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. 2.12. 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 [94], 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 the lack of universal sensitivity direction does not affect the optimization results. Thus, it is still an open question of how much information of a species evolution curve can be represented by a single point.

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 from the perspective of point estsimation. 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.

2.4 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. Maximum a posterior estimation and MCMC sampling are used to obtain the optimal parameters with uncertainty. 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.

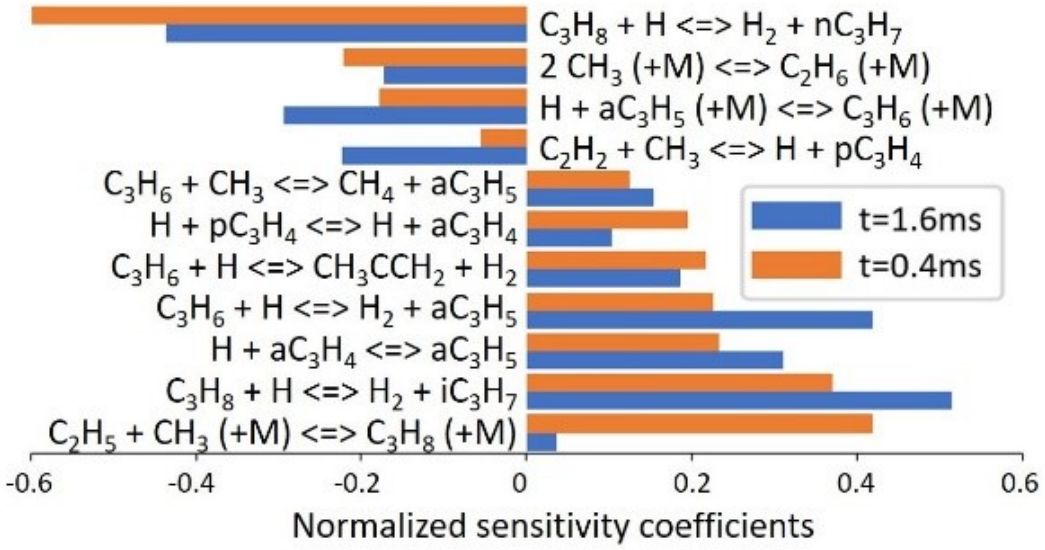


Figure 2.12: Top 10 sensitivity coefficients of pC_3H_4 concentration at 1.6 ms and 0.4 ms.

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 ill-posedness of data-model system should be assessed, and 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, however, C-1 has smaller parameter uncertainty and prediction uncertainty than LastP, and the mismatch of MAP and posterior mean values for parameters of C-1 is observed, indicating that from the perspective of UQ, using the entire curve is helpful to reducing the posterior uncertainty. The reasons for the difference between C-1 and LastP still need further studies.

Chapter 3

Gradient-Based Experimental Design

3.1 Introduction

In Section 1.4, we briefly overview the algorithms used by the combustion community for experimental design in order to efficiently reduce the uncertainty of combustion kinetic models. We notice that all algorithms rely on a set of candidates of experimental conditions. In most cases, the design space is continuous, such as a range of pressure, temperature, or equivalence ratios. We need to discretize the design space into a group of grid points in each dimension, just as what was done in [79], [81], [89]. Then the quantity measuring information gain is evaluated in a point-wise manner, and the point with the highest information would be picked as the optimal design point. This approach, however, is extremely inefficient when the design space is in high dimension. For example, for a 3-D design space of pressure, temperature and equivalence ratios, if we want to discretize each dimension into 20 grid points, then the entire space leads to 8000 grid points. The algorithms need to run 8000 times to find the optimal points. This would be worse in sequential experimental design, since in each iteration we need to repeat 8000 times of evaluation. However, if gradient information with respect to design is available, the gradient-based optimization algorithm can go from the initial point directly to the (local) maximum. In computer science and statistics community, some gradient-based experimental design algorithms have been developed. Huan and Marouzk [112] combined a double-loop Monte Carlo expected information gain (EIG) estimator with gradient descent. Foster et al. [113] combined gradient descent with variational experimental design. Goda et al. [114] combine gradient descent with multi-level Monte Carlo (MLMC) algorithms. Carlon et al. [115] combined Nesterov-accelerated stochastic gradient descent with Bayesian experimental design based on Laplacian approximation. However, these algorithms still cannot run on the fly with high throughput experiments. In this work, we will combine gradient descent with a very efficient D-optimality design based on a linearized response surface leveraging PyTorch auto-differentiation environment. As Section 1.4 states, if the model response is near linear around the nominal value, the D-optimality information gain computed by linearized response surface is close to EIG computed by sampling-based full Bayesian design.

As reviewed by Section 1.3, in order to evaluate the physical model with low computational cost, the response surface is built to replace the expensive physical model. Different

from UQ and point-wise DoE, gradient-based DoE has stricter requirements for response surfaces. First, the inputs of the response surface must include design variables covering the entire design space, otherwise, we cannot compute the derivative with respect to design variables. The influence of design variable inputs on the outputs is far larger than that of parameter inputs (for example, ignition delay time at different thermodynamic conditions can change across more than six orders of magnitude). In addition, the relation between outputs and design variable inputs can be highly nonlinear, and even non-monotonic (e.g., the impact of equivalence ratio to laminar flame speeds). These two characteristics pose great challenges to the expressive power of response surfaces. Second, it is hard to do dimensionality reduction, such as sensitivity analysis or active subspace, and build a response surface with fewer inputs. Since the response surface is required to predict model output across different experimental conditions, the sensitivity and active subspace will also change with experimental conditions. Third, different from doing inference for one time, we cannot use active learning algorithms such as [116], to make training samples focus on the posterior region, since in sequential design procedure we do not know the data a priori. Zhang et al. [67] also use response surfaces with thermodynamic conditions as inputs, but they only generated training samples at conditions of experimental data used to do inference. In this work, we need to build a response surface that can work across the entire design space and wide parameter uncertainty space (instead of only posterior region like [116]). We will use neural networks as response surfaces due to their great expressive power and previous experience in the literature that NN performs better in high-dimensional cases [20], [67].

In this chapter, we build a gradient-based D-optimality design framework assisted by *single* neural network response surface. Our first contribution is combining D-optimality design based on linearized response surface with gradient descent leveraging PyTorch auto-differentiation environment to efficiently searching for optimal experimental conditions. Compared with enumeration in a grid of experimental conditions, our method shows acceleration around five times. Numerical experiments show that the experimental data at designed conditions are more informative than random experiments. Our second contribution is to build a neural network response surface that can give accurate prediction across different thermodynamic conditions (i.e., design conditions). Performance in test sets shows that the accuracy can satisfy the requirement of kinetic UQ according to the criterion proposed by [67]. Our third contribution is to demonstrate the target-oriented experimental design inspired by [117]. Numerical experiments show that data from designed conditions can provide smaller posterior prediction uncertainty compared with random experiments. This method can also be used to find alternative conditions when the targets, whose prediction uncertainties are desired to be reduced, are not reachable in lab facilities, the task same as [83], [84], but in a more rigorous and information-theoretic manner.

3.2 Algorithms and Setup

3.2.1 Gradient-Based D-Optimality Experimental Design

As we reviewed in 1.4, vom Lehn, Cai and Pitsch [89] developed an algorithm based on MUMPCE framework, where they assumed that the MAP point is the same as the prior

nominal value. Then posterior covariance matrix is computed at nominal value as an estimation of the posterior if the experiment is done:

$$\Sigma(\boldsymbol{d}) = \left[\sum_{r=1}^{n} \frac{J_r(\boldsymbol{x_0}, \boldsymbol{d}) J_r(\boldsymbol{x_0}, \boldsymbol{d})^T}{(\sigma_r)^2} + 4\mathbf{I}\right]^{-1}, \tag{3.1}$$

where $\boldsymbol{x_0}$ is the nominal value, i.e., mean value of prior distribution, of kinetic parameters, $J_r(\boldsymbol{x_0})$ is the Jacobian matrix of all measurements with respect to kinetic parameters evaluated at $\boldsymbol{x_0}$ and d, and n is the number of experiments that we want to plan. Then the EIG is computed by the determinant of the covariance matrix, and the optimal experimental condition is the one maximizing EIG:

$$\boldsymbol{d}^* = \underset{\boldsymbol{d}\in D}{\operatorname{argmin}}\ det\left[\sum_{r=1}^{n} \frac{J_r(\boldsymbol{x_0}, \boldsymbol{d}) J_r(\boldsymbol{x_0}, \boldsymbol{d})^T}{(\sigma_r)^2} + 4\mathbf{I}\right]^{-1}, \tag{3.2}$$

where D is the entire design space. In fact, this algorithm is equivalent to EIG computed by Eq. 1.35 when the physical model is a linear model $\boldsymbol{y} = X\boldsymbol{k} + \epsilon$, and both prior $p(\boldsymbol{k})$ and zero-mean noise ϵ are Gaussian. It can be proved that [90]

$$\ln\left\{det\left(\Sigma_0^{-1}\right)/det(X\Sigma_\epsilon^{-1}X^T + \Sigma_0^{-1})\right\} = EIG(X), \tag{3.3}$$

where Σ_ϵ is the covariance of noise, and Σ_0 is the covariance of prior distribution, $EIG(X)$ is the EIG computed by Eq. 1.35. Thus, if the physical model is linear enough around the nominal value, then the result of D-optimality design based on linearized response surface should be pretty close to the full Bayesian experimental design.

However, in [89] the design is done in an inefficient way: constructing a set of candidates of experimental conditions by grid enumeration in a continuous design space. As we discussed in the Section 3.1, if we can utilize gradient information, the process of searching for optimal design would be more efficient. Thanks to the PyTorch auto-differentiation environment [101], either Jacobian matrix of neural network response surface or matrix computation can be differentiable with respect to design. Thus, the optimization problem in Eq. 3.3 can be computed by gradient descent. In addition, many efficient optimization algorithms are available in PyTorch library, which has achieved great success in deep learning.

To be specific, in our task, the optimization problem is

$$\min_{\boldsymbol{d}\in\mathcal{D}} -log_2\left\{det(\Sigma_0)\ /\ det\left[\sum_{r=1}^{n} \frac{J_r(\boldsymbol{z_0}, \boldsymbol{d}) J_r(\boldsymbol{z_0}, \boldsymbol{d})^T}{(\ log_{10}\sigma_r)^2} + \Sigma_0^{-1}\right]^{-1}\right\}. \tag{3.4}$$

Our neural network response surfaces take $z_i = \ln(k_i/k_{i,0})$ as inputs, so the prior mean value $\boldsymbol{z_0}$, and prior covariance matrix Σ_0 in Eq. 3.4 should correspond to the new inputs defined above, which is different from the x_i in [45]. The corresponding prior covariance matrix is a diagonal matrix with $\frac{1}{2}\ln f_i$ in Eq. 1.5 as diagonal elements. The n in Eq. 3.4 is the number of experiments we want to design. Our design targets are ignition delay times (IDT) or laminar flame speeds (LFS) of methanol/air mixture. According to our observation of the data set used by FFCM-1 [118], the 1-σ additive noise of LFS or IDT is around 10%

or 20 % of measured values. Thus, for simplicity, we assign a multiplicative noise $\sigma_r = 1.1$, which should be only slightly different from additive noise of 10% of the measured value. The Jacobian matrix J_r should be evaluated with log_{10} of LFS and IDT as outputs. Here in order to match with Shannon information entropy, we use 2 as the base number in the logarithm, but it has no difference from the Euler number in the perspective of finding optimal experiments.

Adam optimizer [100] is chosen due to its great success in deep learning and its adaptive learning rate for different coordinates. In our experience of implementing the most original version of gradient descent, i.e., $\theta_t = \theta_{t-1} - \nabla_\theta \phi(\theta_{t-1})$, the impacts of difference parameters on EIG are quite different, so the learning rate should be different for each coordinate. If the learning rate is too large, then strong oscillation is observed and the optimizer cannot converge to stationary points; if the learning rate is too small, the optimization process would be time-consuming. Thus, picking a suitable learning rate for each coordinate of design is tedious. After Adam is used, the learning rate can be fixed at 0.01. Weight decay is set as 0. Other parameters are kept as default values. Adam is designed for unconstrained optimization, while in this problem we need to constrain our decision variable within the design space. We combine projected gradient decent [119] with Adam optimizer. Specifically, after each time of updating the design, we project the design updated by Adam onto the design space. In combustion kinetic design, the design space is usually defined by a certain range of pressure, temperature, equivalence ratio, etc. Thus, the space is a hypercube, and the projection on it is easy to obtain. For each coordinate of design, if its value is larger than the range of this coordinate, the projection is the upper bound of it; if its value is smaller than the range, the projection is the lower bound; otherwise, the current design is within the design space and the projection is the design itself. In order to avoid being trapped in the local minimum, we initialize the optimizer from many random starting points. After the optimization of each random initialization is finished, the figures of each coordinate of design versus a number of iterations are plotted for convergence diagnostics. If all coordinates of all initialization arrive plateau region, it is considered all runs have converged. Among all initialization, the one with the largest information gain would be picked as the final optimal design.

In this work, we usually do experimental design in a sequential manner. After each time of design, we generate synthetic data at the design point according to Eq. 1.18. Then this synthetic data is used to do inference and obtain the posterior kinetic parameters. The method of inference is MUMPCE [45], [67] (i.e., Laplacian approximation [38]). Specifically, we obtain the posterior mean value of parameters by Eq. 1.22, and then obtain the posterior covariance matrix by Eq. 1.24. Two differences between our approach and the method in [67] (see Section 2.6) are that, first, we do not eliminate inconsistent data, since we generate synthetic data using perfect model plus Gaussian noise; second, we do not freeze any parameters. In the next iteration of design inference, the posterior model is used as the prior model to do experimental design and inference. Such a design-inference iteration is repeated for a given number of times.

In Eq. 3.4, the optimizer would minimize the volume of a hyper-ellipse described by the covariance matrix of all kinetic parameters. We call it *model-oriented* experimental design. In practice, the uncertainty of all kinetic parameters is usually not the primary goal considered in model calibration using experimental measurements. Thus, some modified objective

function can be employed depending on the different goals. First is the scenario where we only want to minimize the uncertainty of one or several kinetic parameters. We call it *parameter-oriented* experimental design. In this case, we can only pick the rows and columns corresponding to the parameters of our interests before we compute the determinants:

$$\min_{\boldsymbol{d}\in\mathcal{D}} - log_2 \left\{ det\left[(\Sigma_0)_{I,I}\right] \ / \ det\left[\left(\sum_{r=1}^{n} \frac{J_r(\boldsymbol{z_0},\boldsymbol{d}) J_r(\boldsymbol{z_0},\boldsymbol{d})^T}{\left(log_{10}\sigma_r\right)^2} + \Sigma_0^{-1} \right)_{I,I} \right]^{-1} \right\}, \tag{3.5}$$

where I is the index of parameters of interests. For example, if we only want to minimize the uncertainty of parameters 2 and 3, then $I = \{2, 3\}$, and $(\cdot)_{I,I}$ means picking the second and third rows and second and third columns of the matrix, so that we would obtain a new 2×2 matrix. The second scenario is that we want to minimize the prediction uncertainty of one or a series of targets [117]. We call it *target-oriented* experimental design. Given the covariance matrix of parameters, we can obtain the covariance of (a series of) targets analytically by linearizing the response surface. In real applications, we usually consider minimizing the absolute value of uncertainty of each target, instead of the volume of hyper-ellipse of targets' covariance matrix. Thus, here we minimize the sum of squares of the prediction standard deviation of all targets:

$$\min_{\boldsymbol{d}\in\mathcal{D}} \ - \text{sum\_diag}(J_t^T \Sigma_0 J_t) / \text{sum\_diag} \left[J_t^T \left(\sum_{r=1}^{n} \frac{J_r(\boldsymbol{z_0},\boldsymbol{d}) J_r(\boldsymbol{z_0},\boldsymbol{d})^T}{\left(log_{10}\sigma_r\right)^2} + \Sigma_0^{-1} \right)^{-1} J_t \right], \tag{3.6}$$

where sum_ diag$(\cdot)$ means the sum of all diagonal entries, and J_t means the Jacobian matrix of the targets whose uncertainty is desired to be reduced. For the covariance matrix, sum_ diag$(\cdot)$ exactly means the sum of squares of all standard deviations. The third scenario is designing multiple experiments simultaneously, or designing *parallel experiments*. Different from the sequential experimental design, where we conduct experiments and do inference after each design, parallel experimental design finishes the design of several experiments before conducting experiments. When the experimental design program is not accessible during experiments, parallel experimental design should be used before the beginning of experiments. We can simply include more experiments in the summation of Eqs. 3.4, 3.5 and 3.6. This is equivalent to augmenting the dimension of experimental data shown in the toy problem of [85], since extending the second dimension of the Jacobian matrix is equivalent to the sum of the outer product of the original-dimension Jacobian matrix.

3.2.2 Task Setup and Data Generation

In this work, we aim to design experiments of LFS and IDT measurements to minimize the uncertainty of a methanol combustion kinetic model developed by Zhang et al. [120], which is also used in [81]. Only pre-exponential factors are considered as uncertain parameters, while other kinetic parameters are fixed. The uncertainty factors f_i is adopted from the supplementary material of [81]. This mechanism includes 32 species and 197 elementary reactions. Among these 196 elementary reactions, 16 are fall-off reactions, which includes two set of kinetic parameters for high and low pressure limits. Thus, the dimension of

parameters is 213. In addition, this mechanism includes two duplicate reactions R95 and R96, and R96 is not included in the uncertainty factor of [81]. We assign the same uncertainty factor of R95 to R96.

Table 3.1: The design space for IDT and LFS.

Experimental Targets	Pressure [atm]	Temperature [K]	Equivalence Ratio
Ignition delay times	1-30	800-1600	0.5-5
Laminar flame speeds	0.5-10	300-500	0.8-1.5

In order to reduce computational costs, we use neural network response surfaces to replace expensive physical simulation. The training data is generated by physical simulation first, and then this data is used to train a response surface. We use 300,000 samples to build the response surface of IDT, among which 243,000 samples are for the training set, 30,000 for the test set, and 27,000 for the validation set. 106,920 samples are used to build the response surface of LFS, among which 87,480 samples are for the training set, 9,720 for the test set, and 9,720 for the validation set. The prior distributions of parameters $z_i = \ln(k_i/k_{i,0})$ are zero-mean Gaussian distributions with standard deviation $\frac{1}{2}\ln f_i$ for each i. The posterior distribution would not deviate from prior distributions too far. In order to emphasize the accuracy in the near-zero region of parameters, for each response surface, we generate three hierarchical data subsets with different widths of range for parameters, similar to [67]. Specifically, samples are generated by randomly sampling the uniform distribution defined on parameter uncertainty space and design space (except temperature in LFS). The design space is defined in Table 3.1. For pressure, the uniform distribution is defined on the logarithm of pressure. The parameter space is defined by a hypercube $[-\frac{k}{2}\ln f_i, \frac{k}{2}\ln f_i]$, where $k = 1, 2, 3$ for three subsets, respectively. In other words, the half-widths of the hypercubes for three subsets are σ_p, $2\sigma_p$, and $3\sigma_p$, respectively, where σ_p is the standard deviation of the prior distribution of z_i. Different from [67], where the sample of z_i of three subsets are generated from a Gaussian distribution with standard deviation $\frac{1}{10}\ln f_i$, $\frac{3}{10}\ln f_i$, and $\frac{1}{2}\ln f_i$ (i.e., $0.2\sigma_p$,$0.6\sigma_p$, and σ_p), respectively, our samples are generated from a uniform distribution, so that the response surfaces can still have good accuracy near the tail of the prior distribution. Then each subset is split into training, validation, and test sets using the same ratio. The number described at the beginning of this paragraph is the total number of samples from all three subsets. Except for the temperature coordinates of LFS in the training sets of subsets 1 and 2, Latin hypercube sampling [105] is used to make samples uniformly distributed, avoiding clusters and holes in sample space. For the temperature of LFS in the training sets of subset 1 and 2, however, in order to use the initial guess of profiles to accelerate the solution of LFS (a case cannot use an initial guess from different unburnt temperature in Cantera), we discretize the temperature range 800-1600K into 81 grid points (with incremental of 10K) and assign 360 samples for each temperature. Other coordinates, including kinetic parameters, pressure, and equivalence ratio are from random sampling. For the temperature of LFS in the training set of subset 3, since the range of uniform distribution is pretty large, the differences in kinetic parameters between cases can be pretty large. Our experience shows that restarting from an initial guess does not accelerate the solution. Thus,

we still use random sampling for all coordinates of samples and do not use initial guesses in LFS computation. For the temperature of LFS in validation and test sets, we also use random sampling for all coordinates in order to obtain a fair assessment for generalization performance.

After getting the samples of inputs, i.e., kinetic parameters and experimental conditions, we compute the outputs (LFS or IDT) by an open-source code Cantera [104]. The initial composition for both IDT and LFS is methanol/air mixture. For IDT, adiabatic, ideal gas, constant pressure reactor is employed. IDT is defined as the time of maximum temperature gradient. For LFS, Free Flame solver is employed. We use multi-component transport model and enable Soret diffusion. The mesh refine criterion is set as ratio=3, slope=0.06, curve=0.12, and prune=0.02. Domain width is set as 0.1, and loglevel=1. For subsets 1 and 2 of the training set, the solution array of a LFS case is stored and used as the initial guess of the next case to accelerate the computation. We need to order these samples so that the flame structure of two adjacent cases are similar, otherwise restarting from an initial guess cannot accelerate the computation. Compared with thermodynamic conditions (i.e., temperature, pressure, and equivalence ratios), the effect of different kinetic parameters on flame structures is negligible for subsets 1 and 2, from our experience. Thus, we only need to order thermodynamic conditions. For temperature, we already discretize the range 800K-1600K into 81 points. Samples should be first ranked by their temperature. Then, we divide the range of log pressure into 20 intervals. For samples with the same temperature, they are then ranked by their pressure interval. For samples with the same temperature and same pressure interval, we then rank them by their equivalence ratio. By such an ordering method, two adjacent samples in this sequence would have similar thermodynamic conditions. Restarting from the solution of the previous case can generally reduce the solution time by 5-10 times. The restarting case would start from the mesh of the previous case. In order to avoid the accumulation of the mesh, the solver starts from scratch (i.e., without the initial guess of the previous case) every 100 cases.

3.2.3 Neural Network Response Surfaces

As we discussed in Section 1.3, the response surface is usually built to replace the physical model in uncertainty quantification. In this work, we use neural networks as response surfaces due to their great expressive power, especially in high-dimensional cases. Gradient-based experimental design requires the gradient of information gain with respect to design variables, so design variables must be included as inputs. The network structure is shown in Fig. 3.1, including three hidden layers with 1024 neurons in each hidden layer. The input layer includes 216 neurons, corresponding to 213 kinetic parameters plus three experimental conditions. The output layer includes just one neuron, corresponding to LFS or IDT after taking log_{10}. We do not systematically test the dependence of prediction performance on the number of layers and number of neurons. Instead, we simply make the model in an over-parameterization regime. The generalization error of neural networks shows the "double-descent" phenomenon as indicated by [121]. When the parameters are less than samples, the generalization error first decreases, and then increases with the increase of the number of parameters; when parameters are more than samples, the generalization error always decreases with the increase of the model size. Thus, we simply choose numbers

of layers and neurons making the model in the over-parameterization regime, and then do trial-and-error several times to determine the final structure if necessary. Residual neural network (ResNet) [59] structure is adopted for its good property of mitigating gradient explosion and gradient vanishing during back-propagation. Although gradient vanishing and explosion should not be obvious in a three-hidden-layer network, we still adopt ResNet in order to get optimal performance. Two short circuits are set in the second hidden layer and the third hidden layer. Since we use the Jacobian matrix to estimate information gain, the gradient of information gain involves a second-order derivative of outputs with respect to inputs. Thus, the ReLU activation function, which is generally used in deep learning due to its good performance [122], cannot be used, since it is piece-wise linear and hence the second-order derivative is zero. Instead, the SiLU [123] activation function is employed.

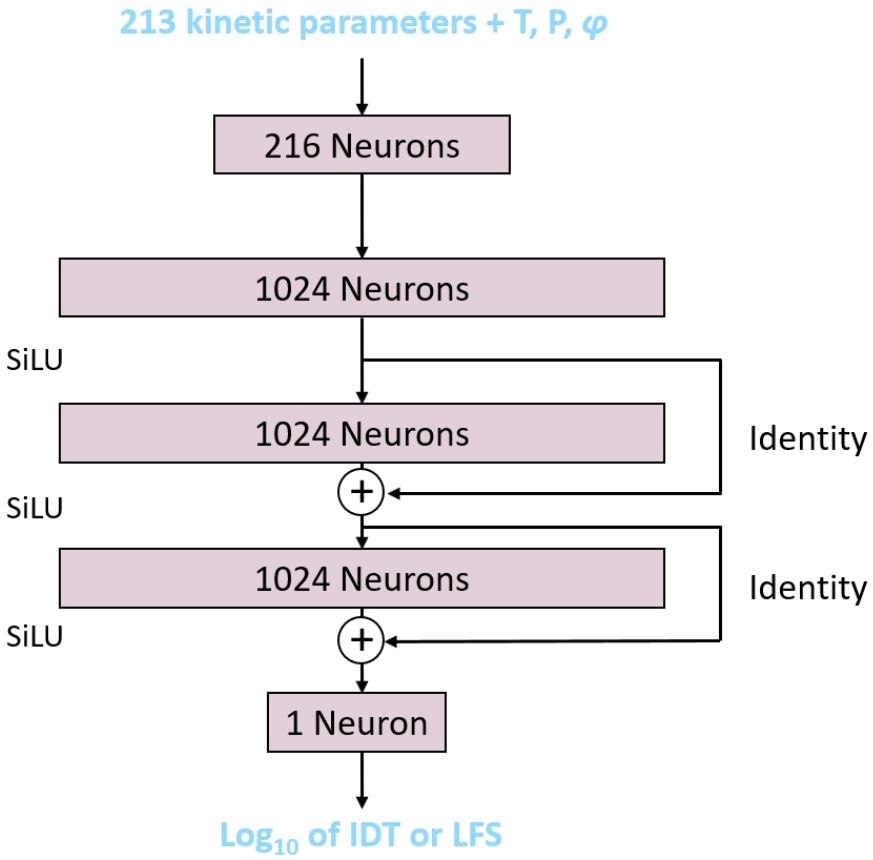


Figure 3.1: The structure schematic of 3-hidden-layer ResNet.

In this work, each data set for building a response surface is split into training, validation, and test sets. The validation set is used to choose hyper-parameters, and the test set is used to assess the performance of generalization. In order to fit data across multiple orders of magnitude (six orders for IDT and one order for LFS) and minimize the multiplicative error during training, we take the logarithm of IDT and LFS as the outputs of response surfaces. Before training, each dimension of inputs and outputs of the data is standardized so that the mean value is zero and the standard deviation is one. PyTorch library [101] is employed

in this work for neural network training. Adam optimizer is used with weight decay of 10^{-4}. The initial learning rate is 0.02 for IDT and 0.08 for LFS, and the scheduler of the learning rate is set as "reduce lr on plateau" with patience 400, factor 0.35, and threshold 0.01. Specifically, the learning rate will decrease to 0.35 times of the current value if in the past 400 epochs, the minimum value of loss does not decrease by 0.01. The batch size is set as 2048. Mean absolute error (MAE) is used as the loss function. In our experience, MAE loss can lead to lower training, validation and test loss values than mean square error (MSE) loss. The training process will stop when the learning rate decreases to 4×10^{-6}. The loss curves of IDT and LFS are shown in Fig 3.2.

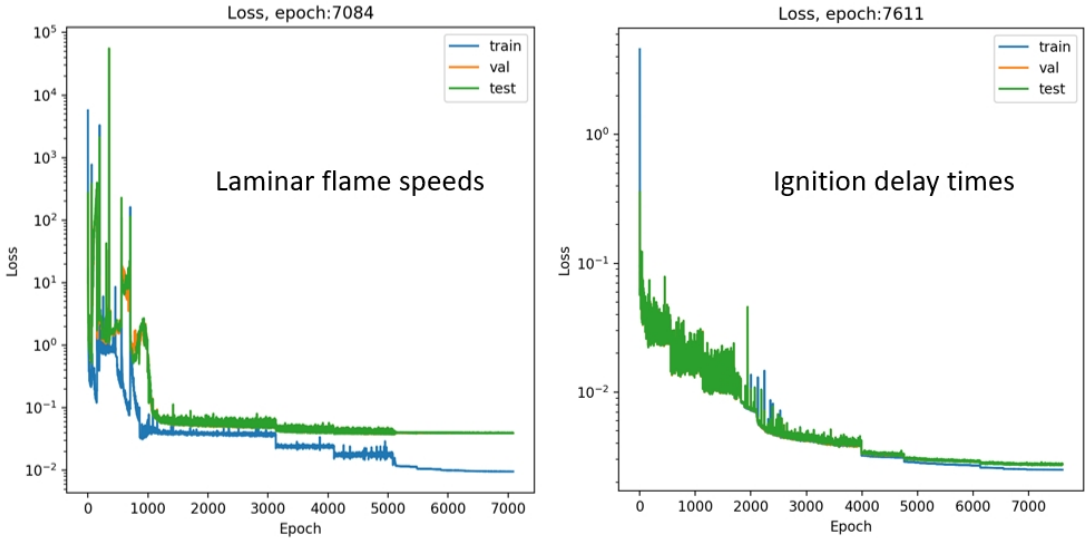


Figure 3.2: The loss values versus training epochs for LFS and IDT.

Table 3.2: Requirement of the accuracy of response surfaces proposed by [67].

Subsets	Mean error	95-percentile error
1 $[N(0, 0.2\sigma_p)]$	1 %	2 %
2 $[N(0, 0.6\sigma_p)]$	2 %	5 %
3 $[N(0, \sigma_p)]$	3 %	10 %

Table 3.3: The accuracy of our response surface for IDT.

Subsets	Mean error	95-percentile error	Maximum error
1 $[U(-\sigma_p, \sigma_p)]$	0.4 %	1.5 %	16 %
2 $[U(-2\sigma_p, 2\sigma_p)]$	0.7 %	2.5 %	40 %
3 $[U(-3\sigma_p, 3\sigma_p)]$	1.5 %	5.3 %	77 %

Table 3.4: The accuracy of our response surface for LFS.

Subsets	Mean error	95-percentile error	Maximum error
1 $[U(-\sigma_p, \sigma_p)]$	0.8 %	1.7 %	3.1 %
2 $[U(-2\sigma_p, 2\sigma_p)]$	1.4 %	3.6 %	8.5 %
3 $[U(-3\sigma_p, 3\sigma_p)]$	2.7 %	7.0 %	36.6 %

The assessment for accuracy of response surfaces follows the method in [67], where the mean and 95-percentile relative error of *test set* are compared with a given criterion, as shown in Table 3.2. Relative error is defined as

$$\epsilon = \frac{|y_{NN} - y_G|}{y_G}, \tag{3.7}$$

where y_{NN} is the prediction of LFS or IDT by neural networks, y_G is the corresponding ground truth. Note that the outputs of neural networks are the logarithms of LFS and IDT, while the relative error uses the original value of LFS and IDT. The mean, 95-percentile, and maximum relative error are shown in Tables 3.3 and 3.4. Note that the parameters of our data are sampled from uniform distributions covering a wider range than the Gaussian distributions in [67]. In addition, the experimental conditions are sampled from uniform distributions covering the range of design space, while in [67], only thermodynamic conditions appearing in experimental data are included. Thus, building such a response surface is far more challenging than the case of [67]. However, the performance of our neural networks still can satisfy the requirement in Table 3.2.

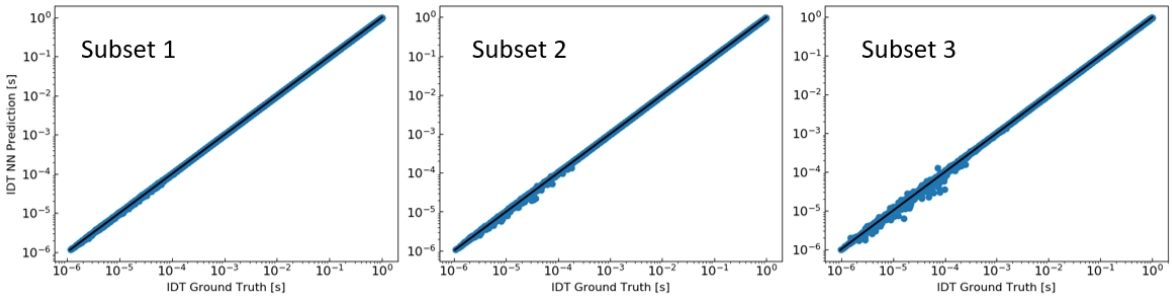


Figure 3.3: The scatter plot of prediction versus ground truth for the response surface of IDT. Samples are from test sets.

The scatter plots of prediction versus ground truth for samples in test sets are shown in Figs. 3.3 and 3.4. We can clearly see that for subsets 1 and 2, almost all samples lie on the diagonal line. For subset 3, a small portion of samples deviate from the diagonal line, but the degree of deviation is very small. In fact, the possibility that posterior mean values are out of $2\sigma_p$ is extremely small. Thus, we can conclude that the response surfaces are accurate enough for experimental design and UQ tasks.

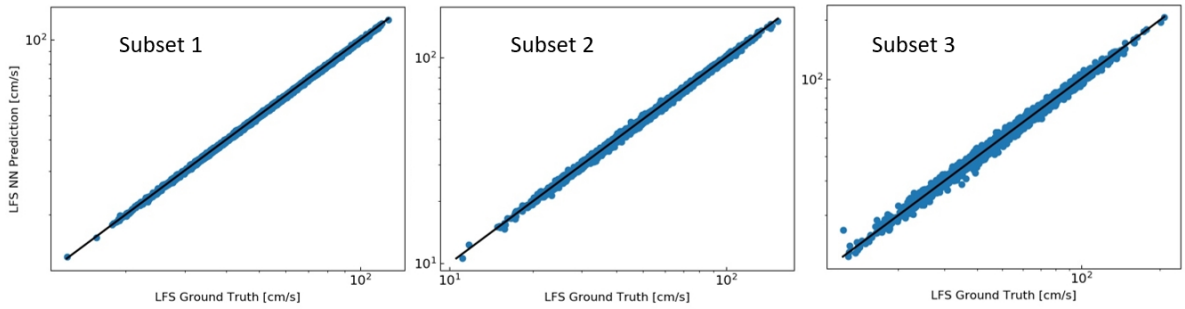


Figure 3.4: The scatter plot of prediction versus ground truth for the response surface of LFS. Samples are from test sets.

3.3 Numerical Experiments

In this section, we show the results of numerical experiments using the algorithms presented in 3.2. It it worth noting except 3.3.4, in all other subsections, the number of experiments n is set as 1, and sequential experimental design is carried for 5 iterations. In each iteration, we do experimental design first and then use the response surface to generate synthetic experimental data on the experimental condition provided by the design algorithms. After that, Bayesian inference based on Laplacian approximation is performed to get the posterior distribution based on the synthetic experimental data. The synthetic data is generated using the prediction of response surface plus Gaussian noise, so that there is no model error. The range of each experimental condition is shown in Table 3.1. All the experiments are ran on the NVIDIA GeForce RTX 2080 Ti Graphic Processing Units (GPU). The running time of experimental design by point-wise enumeration is around 70 iteration per second, while for gradient-based design (the methods of evaluating EIG at a given point keeps unchanged) developed in this thesis, the time of designing a experiment with 30 random initialization takes around 150 seconds. In a 3-D design space, we discretize each dimension into, for example, 50 grid points, and the totaly running time of enumeration is 1785 seconds. Our method can achieve an acceleration of around 10 times. The acceleration would be more obvious in a design space with higher dimensions.

3.3.1 Model-Oriented Experimental Design

In Bayesian experimental design, a common setting is to maximize the EIG with respect to all parameters. In this thesis, we call it *model-oriented design*. In the combustion community, this method has been used in [89]. We demonstrate this approach in our gradient-based experimental design framework for IDT and LFS, respectively, facilitated by neural network response surfaces. In fact, the focus of our work is parameter-oriented design, prediction-oriented design, and parallel design. In this subsection, we just want to demonstrate that our method can finish the same task as previous works.

Table 3.5: The selected experimental conditions and EIG for each iteration of the model-oriented design of LFS.

Iteration	T [K]	P [atm]	Equivalence Ratio	Expected Information Gain
1	304	2.61	0.8	1.20
2	300	6.13	1.49	0.95
3	300	2.71	1.5	0.40
4	338	10	0.8	0.38
5	300	0.5	1.5	0.28

Laminar Flame Speeds

In the model-oriented experimental design for LFS, the resultant experimental conditions of five iterations are shown in Table 3.5. We notice that all temperature lie in low-temperature regions, while pressure and equivalent ratios are pretty dispersed within the design space. The EIG gradually decreases as the number of iterations increases. For the first iteration, EIG of 1.2 means that the volume of the hyper-ellipse is reduced by a factor of 2.2. The trajectories of gradient ascent for 30 initialization is shown in three sub-figures in Fig. 3.5, corresponding to three experimental conditions. It is obvious that all initialization arrive at plateau in the end, which means that all trajectories get converged. Different trajectories do not converge to the same experimental condition, showing that the EIG is not convex with respect to experimental conditions. This demonstrates the necessity of using multiple random initializations in gradient ascent. In order to visualize the landscape of EIG, we plot the contour of EIG of the first iteration in pressure-temperature planes at four different equivalent ratios, as shown in Fig. 3.6. $\phi = 0.8, 1.48$ and 1.5 corresponds to the three values that trajectories converge to. Note that the scale of color bars is different for different sub-figures. It is vividly shows that the landscape is nonconvex, especially in the sub-figure of $\phi = 0.8$, where a double-peak structure is present. We also can verify that T=304K, $\log_{10}$P=0.42 lies in the global maximum of $\phi = 0.8$ sub-figure, and is not trapped in the local maximum (another peak).

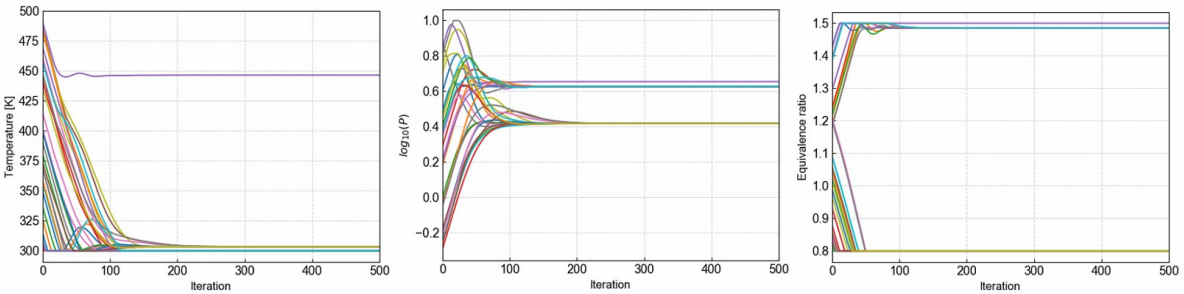


Figure 3.5: The change of temperature, $\log_{10}$ of pressure, and equivalence ratios during gradient ascent iterations for 30 random initialization of model-oriented experimental design of LFS. All coordinates of all 30 cases arrive in plateau regions.

The triangular plot of five most-shrunk parameters is shown in Fig. 3.7, where two circles

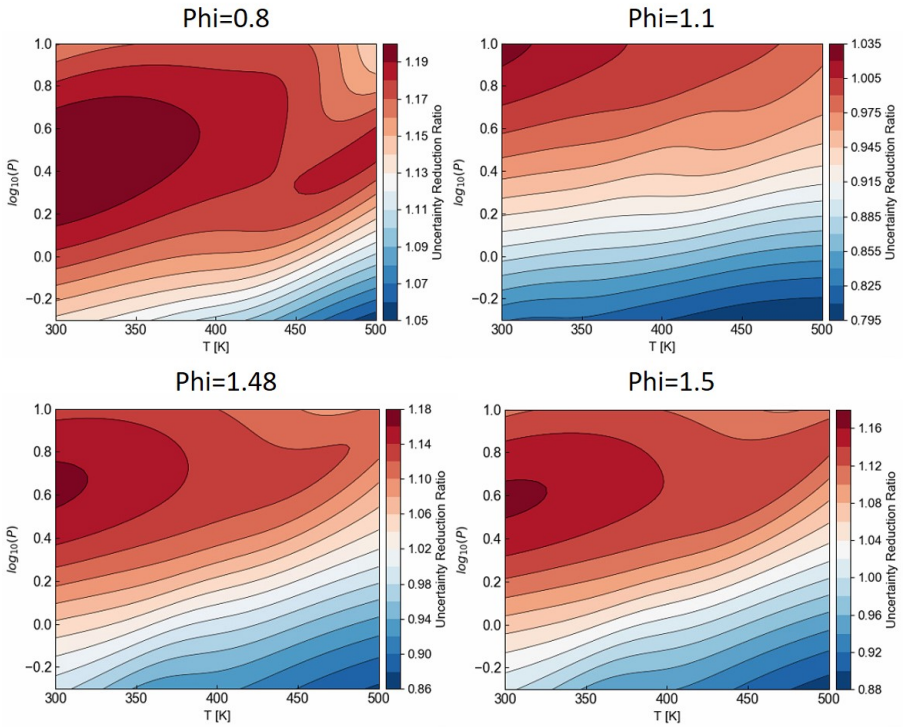


Figure 3.6: The iso-contours of EIG at ϕ=0.8, 1.1, 1.48, and 1.5 in the first iteration for the model-oriented design of LFS. Note that the range of color bars is different for four figures.

in each sub-figure represent 1-σ and 2-σ contours. We can observe that the posterior uncertainty after five designed experiments is very similar to that after five random experiments. A potential reason is that the model-oriented design tries to minimize the hyper-ellipse of all parameters, without focusing on a certain group of parameters, so a triangular plot cannot reflect the uncertainty of the entire model. In fact, in this section, we want to demonstrate that the typical needs in chemical kinetic experiments usually meet with target-oriented or parameter-oriented design, instead of model-oriented design.

In this part, we present the results of the model-oriented experimental design of IDT measurements. The selected experimental condition is shown in Table 3.6. Similar to the case of LFS, the EIG decreases as iteration goes on. For pressure and temperature, in each iteration, they lie (or near) in either the left end or the right end. The trajectories of gradient ascent are shown in Fig. 3.8. It is clearly shown that all initialization has converged. The contour of EIG of the first iteration is shown in Fig. 3.9, where the sub-figure of $\phi = 1.8$ and $\phi = 5$ correspond to two local maxima of equivalence ratios. We notice that in the sub-figure of $\phi = 1.8$, the right upper corner is a local maximum, which exactly corresponds to the orange curve in 3.9. We also can notice that as the equivalence ratio increases, the maximum EIG of this equivalence ratio also increases. Thus, the global maximum should lie in $\phi = 5$, and from this sub-figure, we can find that it is in T=800 K and P=30 atm, which verifies the result of gradient ascent.

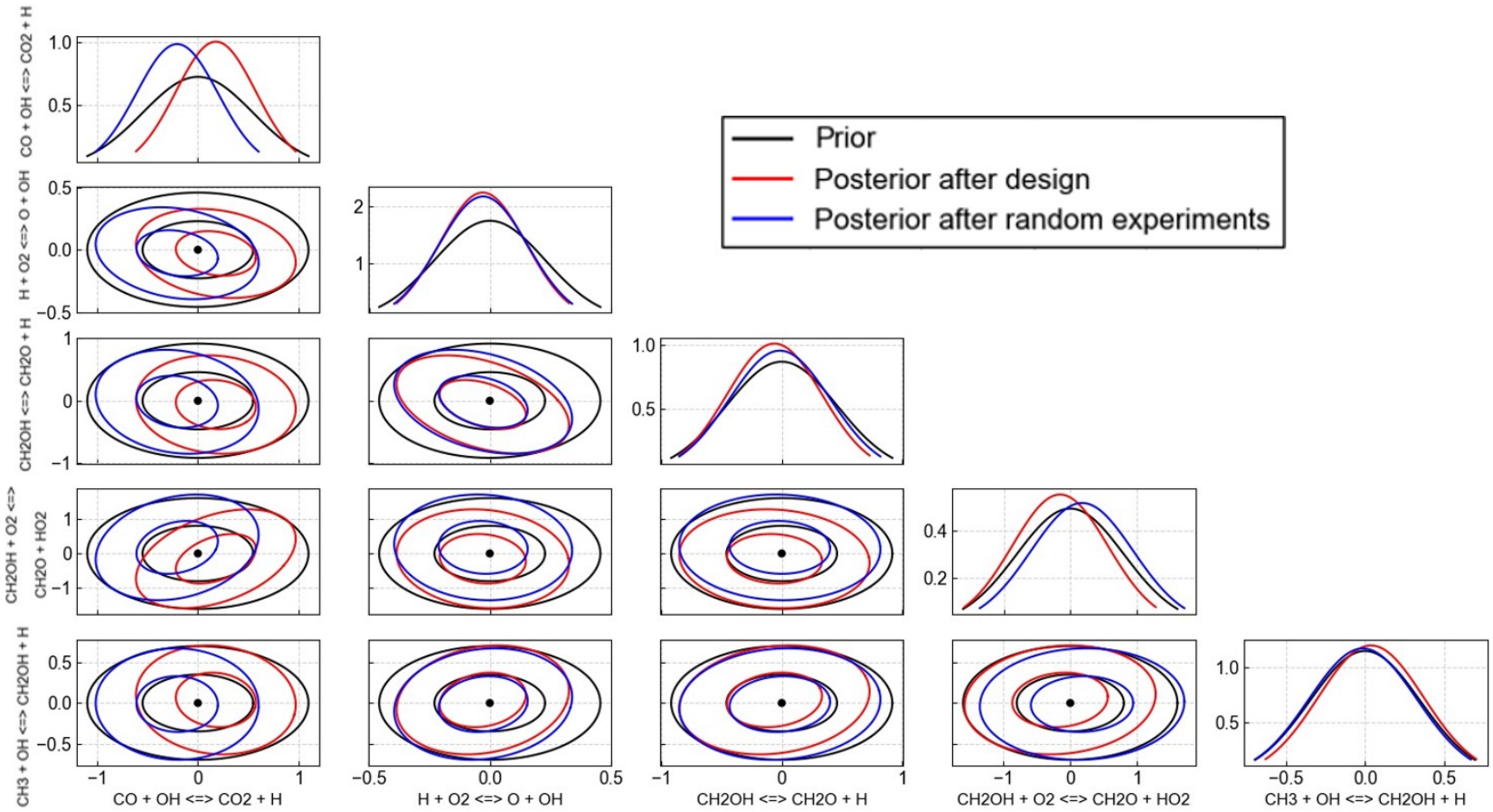


Figure 3.7: The prior and posterior uncertainty of the five parameters with the most shrunk variance after five random experiments and five model-oriented designed experiments for LFS. 1-σ and 2-σ contours are shown for each covariance figure. The black lines represent the prior model, and the black dots represent the nominal values of the prior model, which is also the ground truth used for general measurement data. The red and blue lines represent designed experiments and random experiments, respectively.

Ignition Delay Times

Table 3.6: The selected experimental condition and EIG for each iteration of model-oriented design of IDT.

Iteration	T [K]	P [atm]	Equivalence Ratio	Expected Information Gain
1	800	30	5	1.56
2	1600	27.9	2.3	0.93
3	1600	1	5	0.74
4	800	30	5	0.42
5	1600	1	1.8	0.31

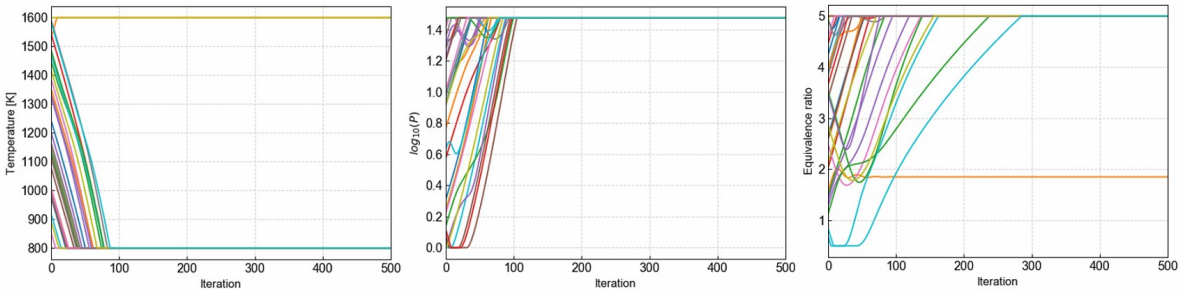


Figure 3.8: The change of temperature, $\log_{10}$ of pressure, and equivalence ratios during gradient ascent iterations for 30 random initialization of the first iteration of model-oriented design of IDT.

Fig. 3.10 shows the five most shrunk parameters after five designed experiments by the proposed method, compared with that after five random experiments. Different from the case of LFS, four parameters show smaller uncertainty after designed experiments, while only one shows smaller uncertainty by random experiments. The model after five designed experiments has a smaller hyper-ellipse volume of the total five parameters.

3.3.2 Parameter-Oriented Experimental Design

In practice, the quantity of interest (QoI) is usually several key parameters, instead of all parameters. In this thesis, we define the experimental design aiming at reducing the uncertainty of several specified parameters as *parameter-oriented* design. In the previous works, parameter-oriented design is usually done by fixing the disregarded parameters and only setting the parameters of our interest as adjustable parameters [85]. However, this can introduce extraordinary model error, as discussed in Chapter 2. In this thesis, we present a new algorithm for parameter-oriented design. We obtain the posterior distribution without fixing any parameter and only minimizing the volume of the hyper-ellipse composed by the parameters of interest. In this subsection, we demonstrate the feasibility of this method by numerical experiments of IDT and LFS.

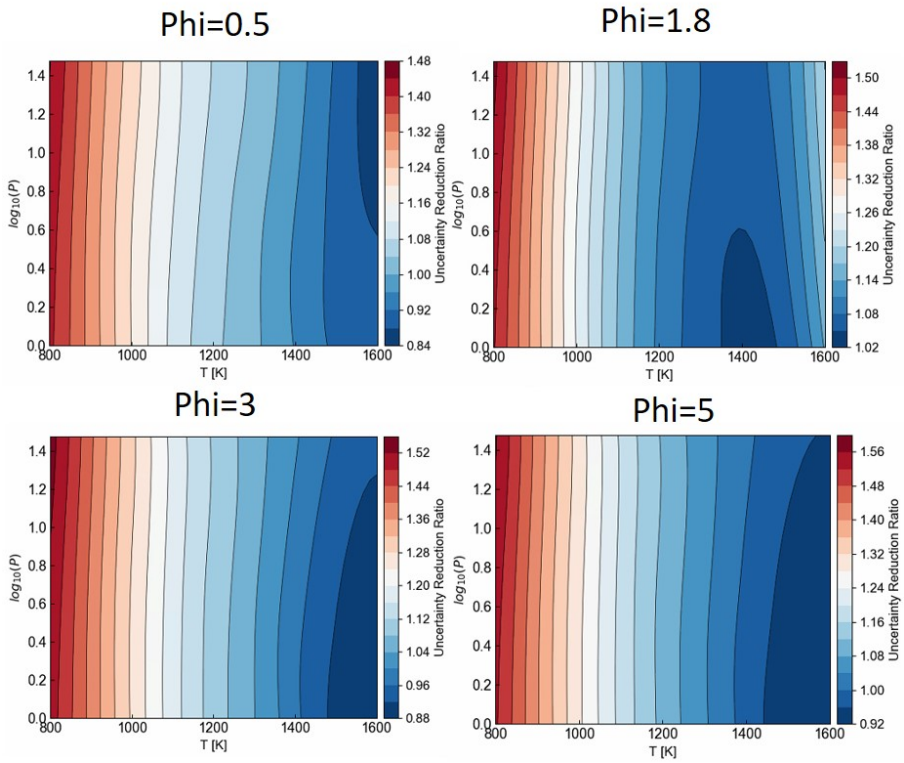


Figure 3.9: The iso-contours of EIG at ϕ=0.5, 1.8, 3, and 5 in the first iteration of the model-oriented design of IDT. Note that the range of color bars is different for the three figures.

Laminar Flame Speeds

Table 3.7: The selected experimental condition and EIG for each iteration of the parameter-oriented design of LFS.

Iteration	T [K]	P [atm]	Equivalence Ratio	Expected Information Gain
1	300	10	0.8	0.53
2	300	10	1.37	0.22
3	350	2	0.8	0.18
4	347	2	0.8	0.09
5	300	10	1.33	0.06

In this section, we present the outcomes of parameter-oriented experimental design for laminar flame speed experiments. The target parameters are parameter 153 and parameter 188. Table 3.7 illustrates the chosen experimental conditions of five design-inference iterations. Similar to the previous results, the EIG decays with each iteration. The equivalence ratio lies in the upper or lower bounds for each iteration. In terms of pressure and temperature, they distribute among the entire range of design space. The trajectories of gradient ascent are depicted in Fig. 3.11. It is evident that all initialization has converged

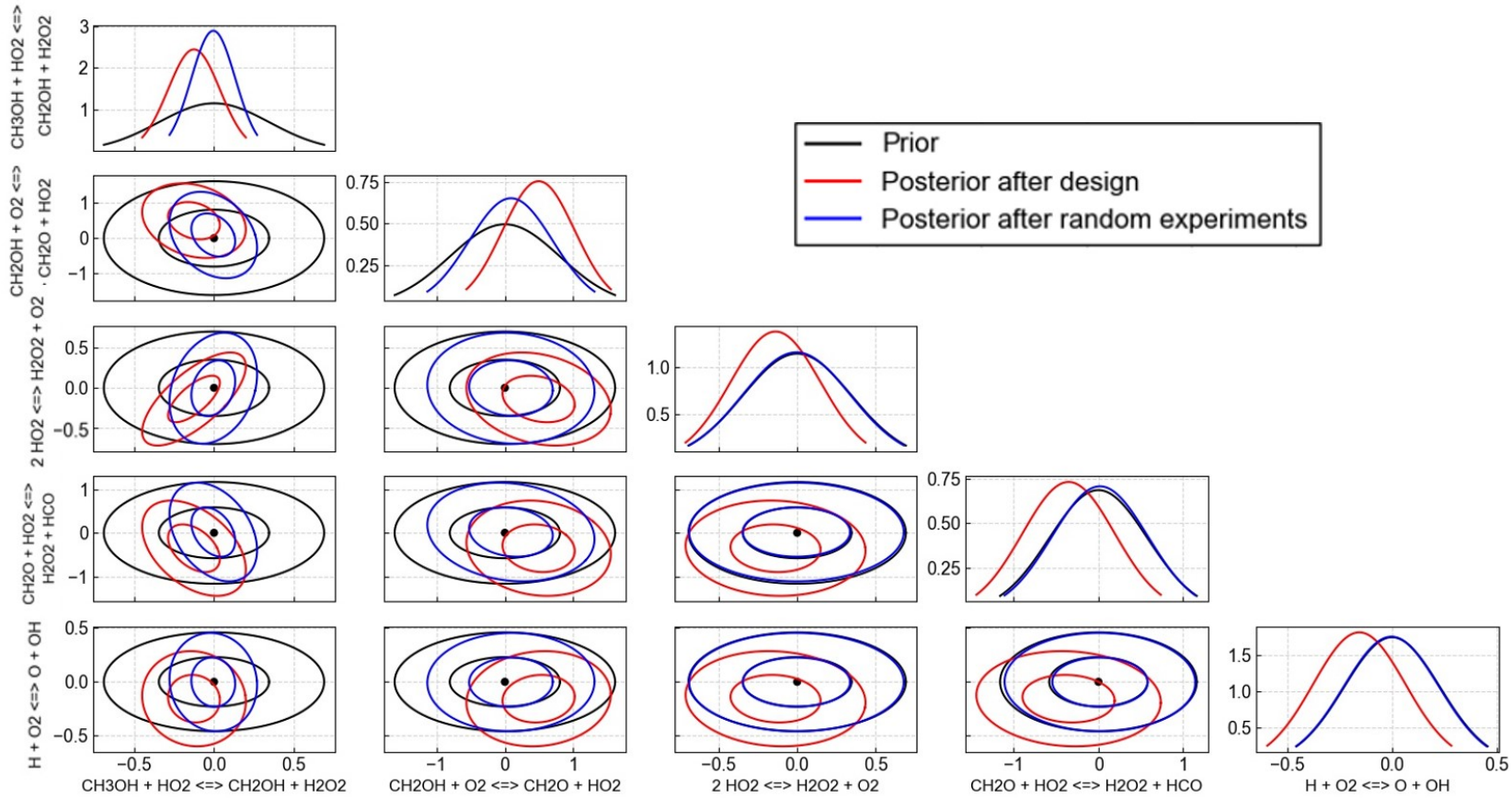


Figure 3.10: The prior and posterior uncertainty of five parameters with the most shrunk variance after five random experiments and five model-oriented experimental designs of IDT. 1-σ and 2-σ contours are shown for each covariance figure. Black lines represent the prior model, and black dots represent the nominal values of the prior model, which is also the ground truth used for general measurement data. The red and blue lines represent designed experiments and random experiments, respectively.

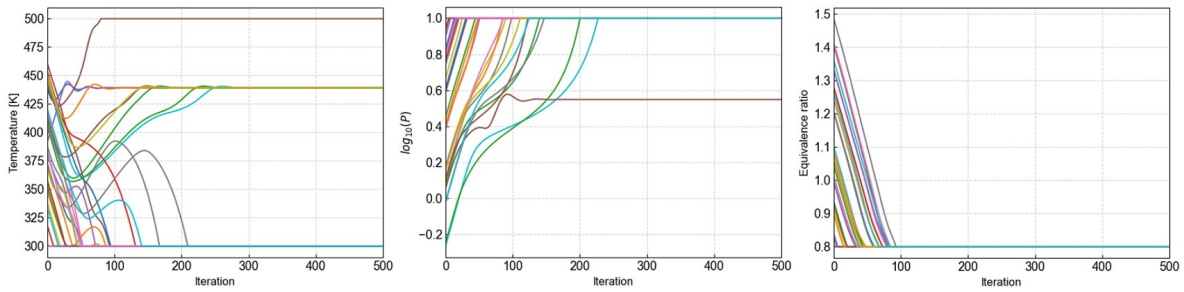


Figure 3.11: The 30 trajectories of temperature, $\log_{10}$ of pressure, and equivalence ratios during gradient ascent iterations.

to a plateau. The contour of the EIG from the first iteration is displayed in Fig. 3.12. The sub-figures corresponding to $\phi = 0.8$ represent only one local maximum (and also the global maximum) of equivalence ratio. Same as in previous cases, the contours are well consistent with the trajectories. The sub-figure of $\phi = 0.8$ clearly shows that there are multiple local maxima in terms of pressure and temperature, which indicates that initializing the optimizer from random starting points can well overcome the issue of nonconvexity in this problem.

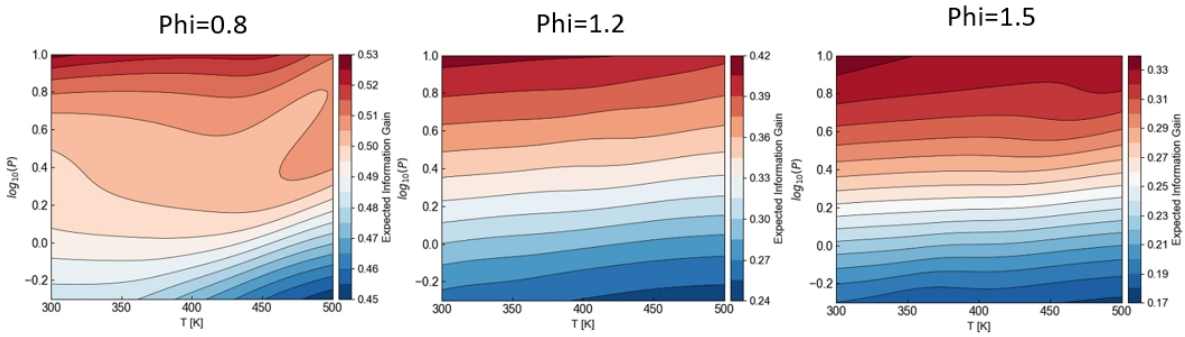


Figure 3.12: The iso-contours of EIG at ϕ=0.8, 1.2 and 1.5 in the first iteration.

Fig. 3.13 illustrates the prior and posterior uncertainty of target parameters after five designed experiments, as compared to those after five random experiments. It is clear that the posterior uncertainty after five designed experiments is smaller than that after five random experiments, indicating the current algorithm can pick the experimental conditions that are most informative to the specified parameters.

Ignition Delay Times

This part reports the same procedure implemented in the problem of IDT measurement for reducing the uncertainty of parameter 18, 71, and 175. Table 3.8 shows the five selected experimental conditions in the five design-inference iteration. Similarly, the EIG decreases as more experimental data is taken in. All trajectories have converged, as shown in Fig. 3.14, and none of them converges to the local maximum. In Fig. 3.15, there does exist a local maximum in the right upper corner of $\phi = 2$ sub-figure. However, the optimizer will converge to this local maximum only when the initialization is around this corner, i.e., T

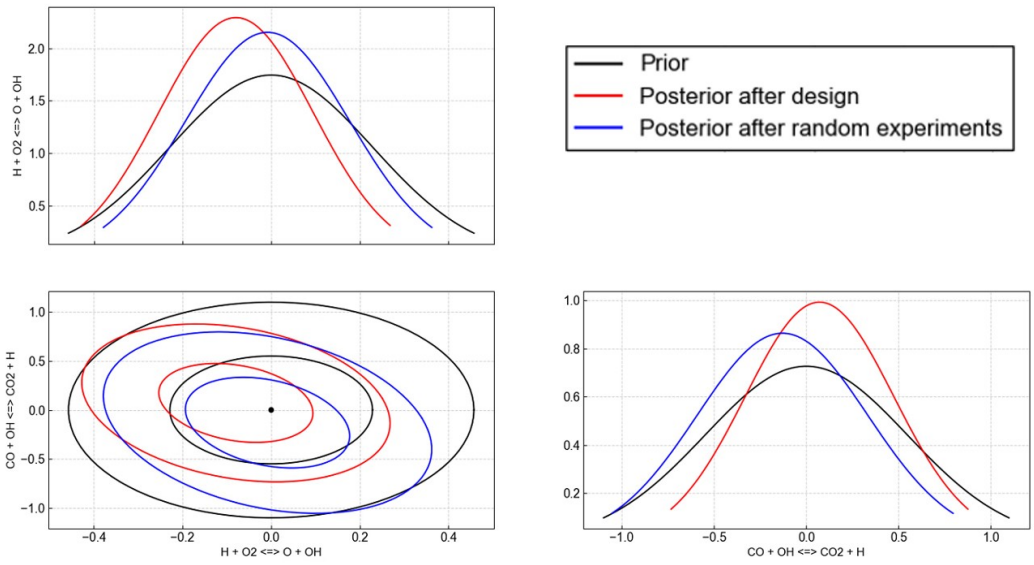


Figure 3.13: The prior and posterior uncertainty of parameters 153 and 188.

larger than 1550 K and ϕ around 2. It is possible that the number of initializations is too small to capture such a local maximum. In this case, we are lucky to capture the global maximum since the missed local maximum is not the global maximum. Thus, this method does not provide any guarantee of capturing global maximum, but increasing the number of initializations will increase the probability of arriving global maximum. If the optimization of different initialization is carried out in a serial computing manner, the computational time is proportional to number of initialization; if this is done in a parallel computing manner, however, the computational time will not show obvious increase as long as the number of initialization does not exceed the number of cores of Central Processing Units (CPU) or GPU. Currently, all the computation is done in a serial manner, and parallelization would be our next step. Thus, the number of initializations should be carefully determined in order to keep a balance between the probability of obtaining a global maximum and computational budget.

Table 3.8: The selected experimental condition and EIG for each iteration of the parameter-oriented design of IDT.

Iteration	T [K]	P [atm]	Equivalence Ratio	Expected Information Gain
1	800	30	5	1.37
2	1600	30	1.9	0.59
3	800	1	5	0.30
4	1327	30	2.95	0.17
5	1600	30	1.85	0.20

The results of posterior uncertainty is shown in Fig. 3.16. It is clear that the data from designed five experiments is more informative than that of five random experiments,

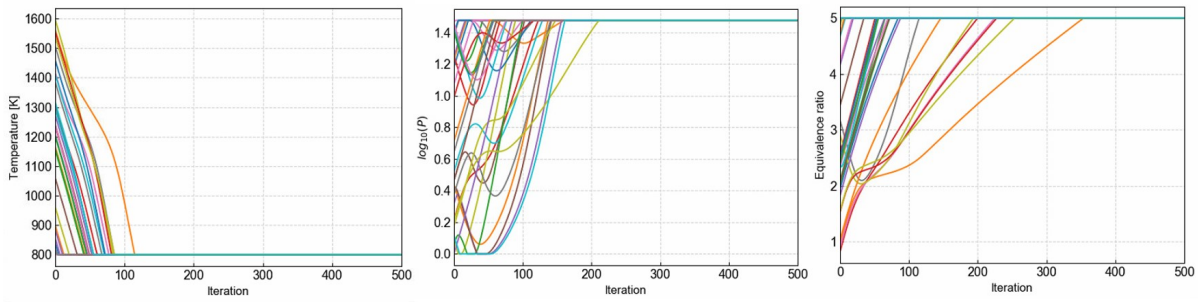


Figure 3.14: The 30 trajectories of temperature, $\log_{10}$ of pressure, and equivalence ratios during gradient ascent iterations.

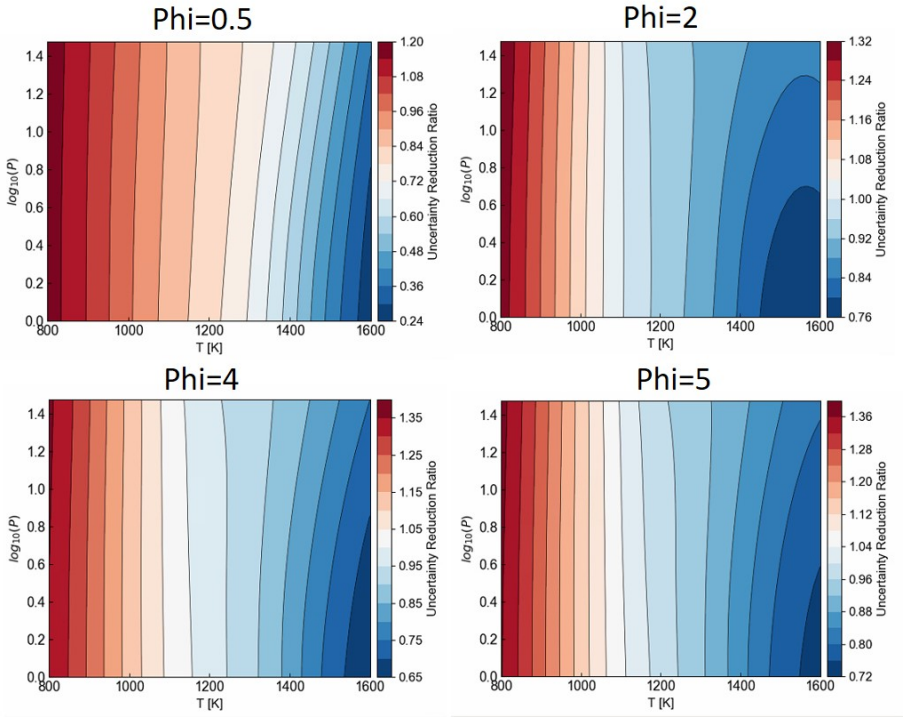


Figure 3.15: The iso-contours of EIG in the first iteration at ϕ=0.5, 2, 4, and 5.

since the model updated by five designed experiments has smaller posterior uncertainty on specified parameters.

3.3.3 Target-Oriented Experimental Design

The eventual goal of uncertainty reduction is to reduce the prediction uncertainty of the model. If we focus on the prediction uncertainty of a certain (or a certain group of) target(s) and do not care about the uncertainty of certain parameters, we can design experiments with reduced prediction uncertainty of these targets as objective. When there is only one target and we can conduct experiments at this condition, the strategy would be pretty straightforward: repeating experiments at this target. However, if experiments are hard

to conduct under the target conditions due to the constraints of instruments, we need to leverage algorithms to design target-oriented experiments.

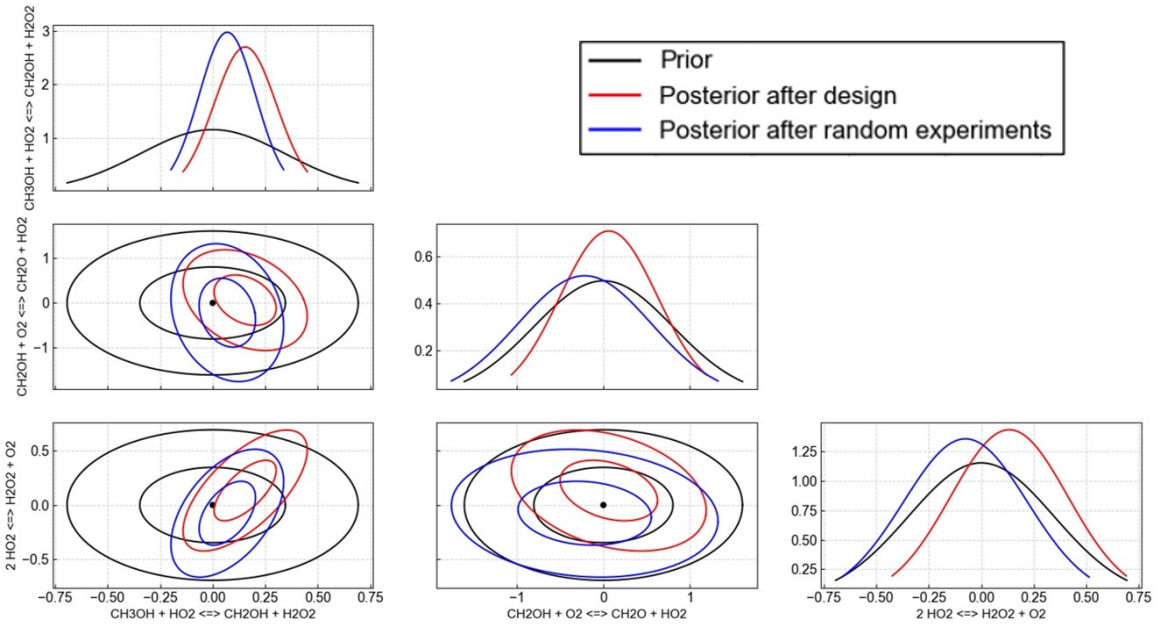


Figure 3.16: The prior and posterior uncertainty of parameters 18, 71, and 175.

Ignition Delay Times

In this part, we will demonstrate designing five experiments within the range of P=1-15 atm, T=800-1200 K, ϕ=0.5-2, while our goal is to minimize the prediction uncertainty of IDT at P=30 atm, T=1200-1600K, ϕ=5. The Jacobian matrix J_t in Eq. 3.6 is chosen as the derivative of IDT with respect to kinetic parameters at 81 temperature points (with 5K increment). The sum of prediction variances (SPV) of all 81 points is minimized. Note the here the variance is for $\log_{10}$ of IDT.

Table 3.9: The selected experimental condition and uncertainty reduction factor for each iteration of the target-oriented design of IDT.

Iteration	T [K]	P [atm]	Equivalence Ratio	Uncertainty Reduction Factor
1	1200	1.0	1.98	3.78
2	1200	1.0	2	1.51
3	1200	1.0	1.97	1.25
4	1200	1.0	1.98	1.15
5	1200	1.0	1.98	1.10

The conditions chosen by the algorithm and the corresponding uncertainty reduction factors are shown in Table 3.9. Here the uncertainty reduction factor (URF) is defined as

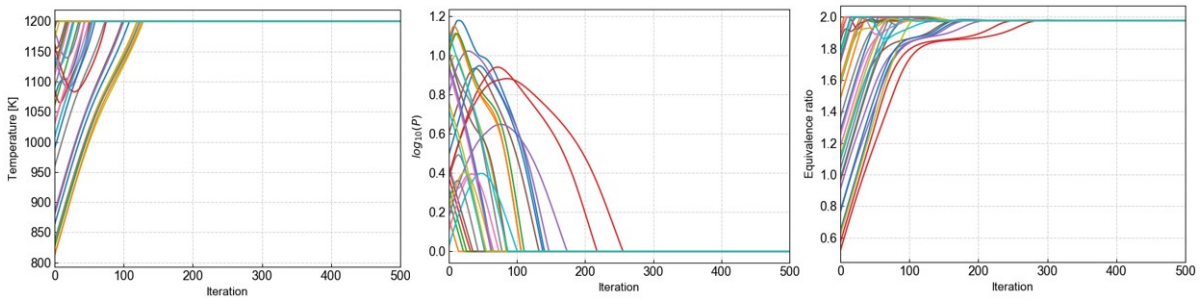


Figure 3.17: The 30 trajectories of temperature, $\log_{10}$ of pressure, and equivalence ratios during gradient ascent iterations.

the ratio of SPV of the prior model over SPV of the posterior model. From the table, we can observe that all iterations picked almost the same experimental conditions. After the first iteration, the uncertainty can be reduced by a factor of nearly four, but as we obtain more experimental data, the URF decreases for each iteration, similar to the change of EIG in previous cases. The trajectories of design during gradient ascent is shown in Fig. 3.17. It is clear that every case of random initialization converges to the same stationary point. From Fig. 3.18, we can clearly see that as ϕ increases, the maximum URF increases (note that the range of color bars is different for different sub-figures). The maximum of URF does lie in the converged conditions. From the 2-D contours shown in Fig. 3.18, we can conjecture that the contour of URF in 3-D is roughly a series of oblique ellipses, and thus near convex. This could be a potential reason why all initializations converge to the same maximum. The comparison of prior and posterior prediction uncertainty is shown in Fig. 3.19. It can be clearly seen that the posterior uncertainty of both posterior mechanisms is smaller than that of the prior model. The model after designed experiments shows smaller prediction uncertainty, especially at the low-temperature side.

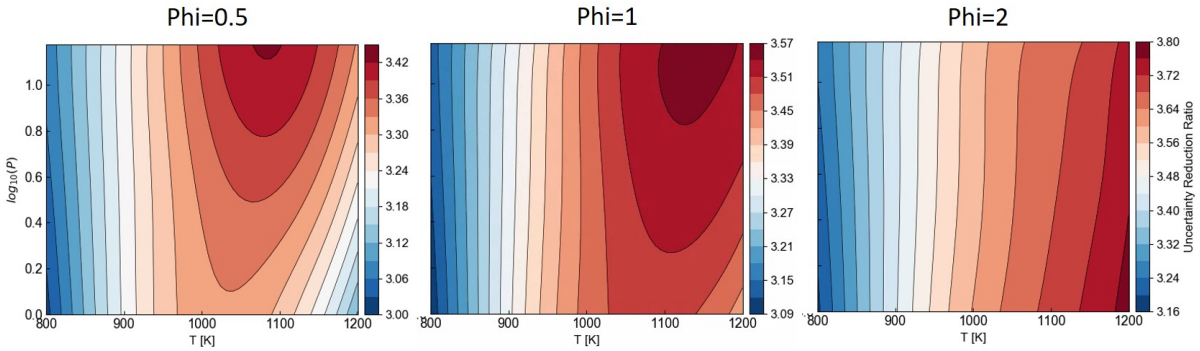


Figure 3.18: The iso-contours of URF at ϕ=0.5, 1 and 2 in the first iteration.

Laminar Flame Speeds

In this part, we will illustrate the design of five LFS experiments within the ranges P=0.5-5 atm, T=300-400 K, ϕ=0.8-1.2. Our objective is to minimize the prediction uncertainty of

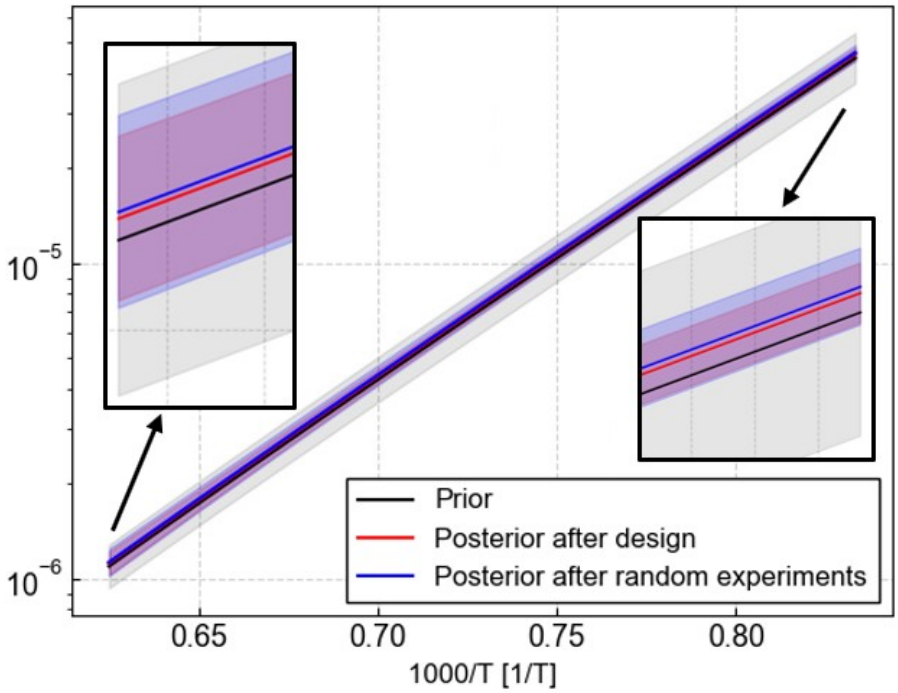


Figure 3.19: The prior and posterior 1-σ prediction uncertainty and mean values of IDT predictions at T=1200-1600K, P=30 atm, ϕ=5 after 5 random experiments and designed experiments, respectively. The black line and (gray) shadow are for the prior model, the red line and shadow for the posterior model after obtaining data from 5 designed experiments, and the blue line and shadow for the posterior model after obtaining data from 5 uniformly distributed random experiments.

LFS at P=10 atm, T=500K, ϕ=0.8-1.5. The Jacobian matrix J_t in Eq. 3.6 is calculated at 31 equivalence ratio points (equally spaced). The SPV of all 31 points is minimized.

Table 3.10: The selected experimental condition and URF for each iteration of the target-oriented design of LFS.

Iteration	T [K]	P [atm]	Equivalence Ratio	Uncertainty Reduction Factor
1	300	5	1.13	3.22
2	300	5	0.9	1.48
3	300	5	1.2	1.32
4	300	5	1.2	1.19
5	300	5	0.8	1.17

The algorithmically chosen conditions and their corresponding URF are presented in Table 3.10. The table reveals that all iterations select the same temperature and pressure, while the equivalence ratios are distributed across the entire design space. Similar to the IDT scenario, as additional experimental data is acquired, the URF decreases with each

iteration. The trajectories of the design during gradient ascent is depicted in Fig. 3.20. It is evident that every case of random initialization converges to the same stationary point. Fig. 3.21 illustrates that the landscape is nearly linear. The contrast between the prior and posterior prediction uncertainty is depicted in Fig. 3.22. It is evident that the posterior uncertainty for both models is reduced compared to that of the prior model. The posterior region after designed experiments is smaller than that after random experiments, especially at the fuel-rich side, where the experimental condition is not accessible.

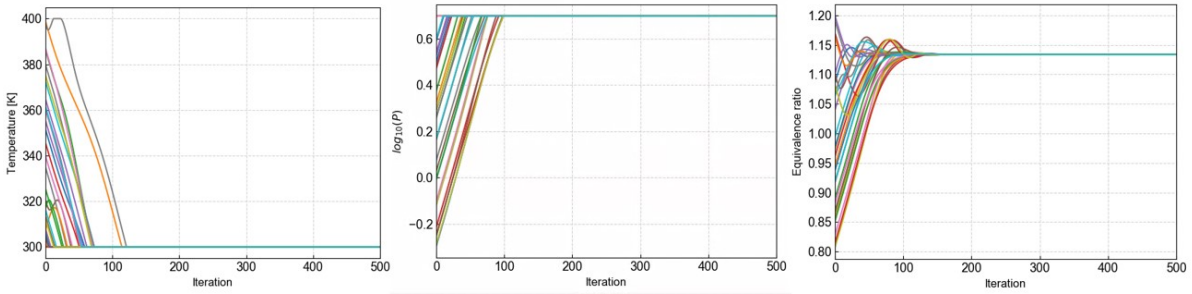


Figure 3.20: The 30 trajectories of temperature, $\log_{10}$ of pressure, and equivalence ratios during gradient ascent iterations.

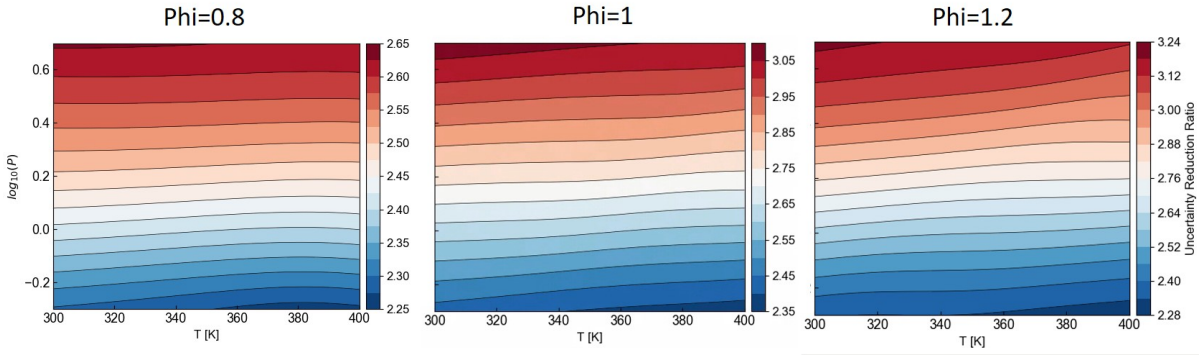


Figure 3.21: The iso-contours of URF at ϕ=0.8, 1 and 1.2 in the gradient ascent iterations.

3.3.4 Parallel Experimental Design

In real practice, it is possible that we cannot conduct one experiment and then do inference and design the next experiment. The first typical scenario is that the experiment is parallel in nature, such as the sensor placement, where multiple sensors can be placed, or multiple facilities that can be operated simultaneously. The second typical scenario is that the experimental data is acquired in high frequency so that the duration of experiments is far shorter than the time of one design-inference iteration. In this case, we can design multiple experiments at once. Of course, this approach will increase the dimension of the design space, where more initialization is required to converge to the global minimum in a more complex geometry. Thus, it assumes that we can parallelize different initializations so that the increase in initialization number will not cause an increase in running time. In this thesis, we

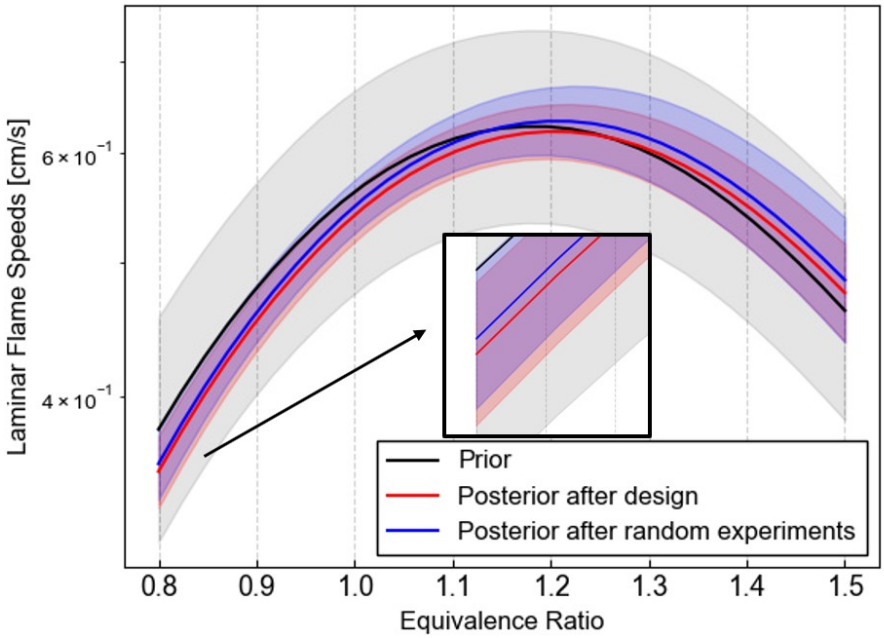


Figure 3.22: The prior and posterior prediction uncertainty ranges and mean values of LFS predictions at T=500K, P=10 atm, ϕ=0.8-1.5 after 5 random experiments and designed experiments, respectively.

still run parallel experimental design in a sequential manner, while the parallelization is left as the task of future works. The number of random initialization is set as 60. It is worth noting that the parallel design does not refer to a new design target, so it can be combined with model-oriented, target-oriented, and parameter-oriented design. In this subsection, we will use IDT to demonstrate the parameter-oriented parallel design and use LFS to demonstrate the combination of target-oriented design and parallel design. Due to the high-dimension nature of parallel design, we will not show the convergence trajectories of each dimension. We already verified that all the cases have converged to stationary points. Also, it is hard to visualize the landscape of EIG or URF. We only present the selected experimental conditions and the comparison of random experiments and designed experiments.

Ignition Delay Times

In this part, we try to minimize the volume of the hyper-ellipse of parameters 18, 71, and 175 by conducting five IDT measurements, and we design the five experiments before any data is acquired. The selected experiments and the final EIG are listed in Table 3.11. As a comparison, the EIG of five random experiments is 2.07. This means that the volume of hyper-ellipse is smaller by a factor of 1.5 compared with the case with random experiments. The prior and posterior uncertainty contours are shown in Fig. 3.23. It is clearly shown that although the posterior uncertainty of parameter 18 is smaller for random experiments, the volume of all three parameters for designed experiments is smaller than for random experiments.

Table 3.11: The selected experimental conditions and EIG for each iteration of the parallel parameter-oriented design of IDT.

No.	T [K]	P [atm]	Equivalence Ratio	Expected Information Gain
1	800	1	5	-
2	1374	1	5	-
3	800	30	5	-
4	1600	30	1.93	-
5	1600	30	1.93	2.6

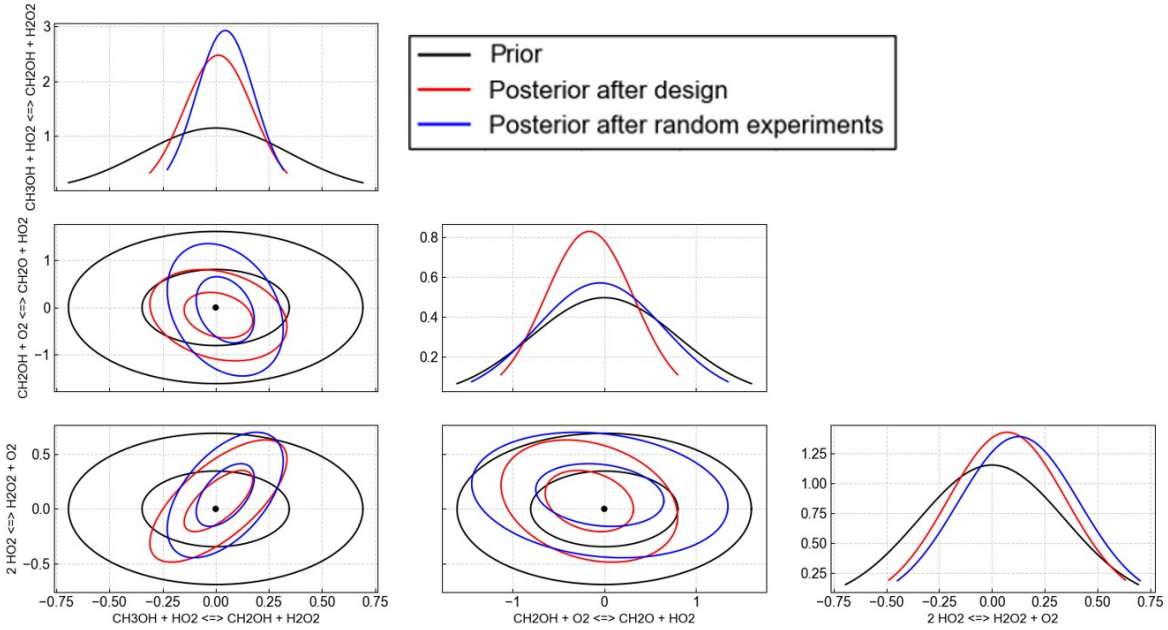


Figure 3.23: The prior and posterior uncertainty of parameters 18, 71, and 175.

Laminar Flame Speeds

In this part, similar to part 3.3.3, our goal is to minimize the prediction uncertainty of LFS at P=10 atm, T=500K, ϕ=0.8-1.5 by conducting pre-designed five LFS experiments within the ranges P=0.5-5 atm, T=300-400 K, ϕ=0.8-1.2. The design is carried out before data is collected. The experiments chosen and the URF after five experiments are detailed in Table 3.12. In contrast, the URF for five randomly selected experiments is 6.4. Figures 3.24 illustrate the prior and posterior prediction uncertainty. It is evident that the designed experiments are more effective in terms of reducing the specified prediction uncertainty. Similar to part 3.3.3, such an out-performance is more obvious on the rich side.

Table 3.12: The selected experimental conditions and URF for each iteration of the parallel target-oriented design of LFS.

No.	T [K]	P [atm]	Equivalence Ratio	Uncertainty Reduction Factor
1	300	5	1.2	-
2	341	5	0.8	-
3	341	5	0.8	-
4	300	5	1.2	-
5	300	5	1.2	9.9

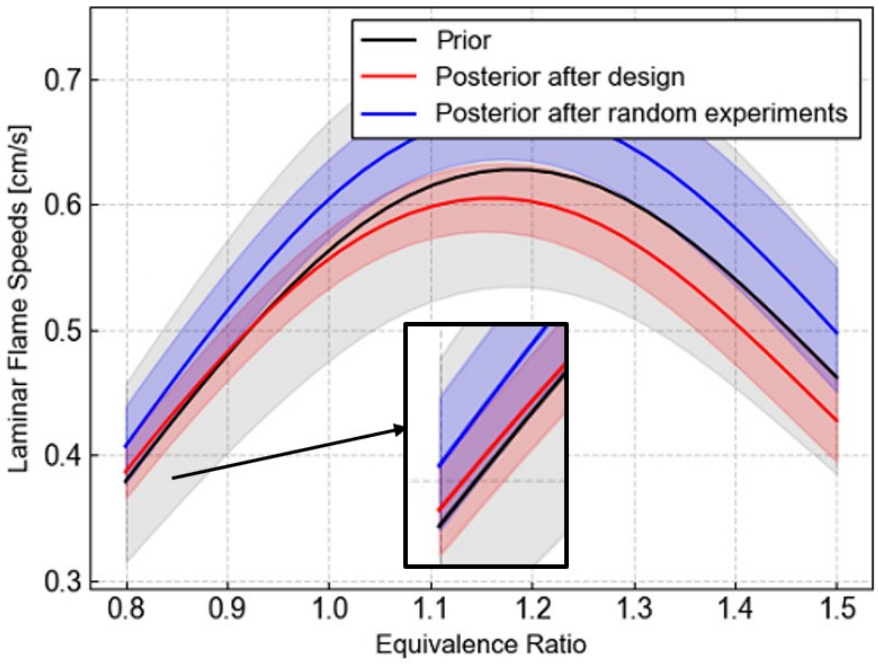


Figure 3.24: The prior and posterior prediction uncertainty ranges and mean values of LFS predictions at T=500K, P=10 atm, ϕ=0.8-1.5 after 5 random experiments and designed parallel experiments, respectively.

3.4 Conclusion

In this chapter, we propose an efficient Bayesian experimental design algorithm combining Laplacian approximation-based experimental design and gradient-based design optimization with neural network response surfaces. Instead of obtaining information gain by Laplacian approximation and integrating it over y, we take the simplified version same as [89]. We optimize the design space by projected gradient ascent, and point-by-point enumeration is avoided. In order to avoid trapping in local maxima of information gain, random initialization of multiple optimization cases is employed. The resultant algorithm shows great efficiency and robustness against local maxima. A novel neural network response surface is designed, where the inputs include both kinetic parameters and experimental conditions. This neural

network can provide accurate prediction across the entire wide design space, so that multiple response surfaces are not needed at different experimental conditions.

In addition, we further develop different experimental design targets based on the posterior covariance matrix computed by Laplacian approximation, including model-oriented, parameter-oriented, target-oriented, and parallel experimental design. In previous works, they either just presented model-oriented experimental design [85], [89], or addressed them on a more computationally expensive framework [87], [117], or naively fixed the disregarded parameters [85]. In this work, we demonstrate these design targets in a more efficient framework based on a full posterior covariance matrix. The numerical experiments show that the proposed method can work very well in terms of reducing the uncertainty of the specified parameters, reducing the uncertainty of the specified targets, or designing multiple experiments at once. The proposed method achieves more than ten times acceleration compared with the same information gain evaluation method but with a point-wise enumeration of 50 grids in each dimension in a 3-D design space.

Chapter 4

Future Work

In Chapter 2, we compared the impact of effective independent-data number and target selection on the results of parameter identification and posterior model prediction. In the effective independent-data number part, the role of ill-posedness is identified. Some other reasons, such as model error and inaccurate likelihood, are discussed. In the target selection part, the potential reason for the similarity of C-1 and LastP, in terms of point estimation, is discussed. However, the answers to several questions are still unclear. First, the role of model error and inaccurate likelihood functions in the Bayesian inference of time-series data is still unclear. Given that the model data is well-posed, whether effective independent-data number depends on the imperfectness of models and likelihoods still needs further studies. Second, whether several points can represent the information of the entire curve, in terms of point estimation, still needs further verification. In our case, the curves of pyrolysis usually have simple and monotonic shapes, and it is not clear whether the observations in this case are still valid in more complex curves, such as species profiles of autoignition. Third, the reason why the prediction uncertainty of LastP is larger than that of C-1, even in the last point, is undisclosed. Fourth, in C-1, the mean values of posterior parameter uncertainty are different from the MAP values for many parameters, while in LastP, the mean values are almost the same as MAP values except for one parameter. The posterior uncertainty contours of C-1 are also different from the prior ones, while the posterior contours of LastP are very similar to its priors. Such a difference is still unexplained.

In Chapter 3, we develop a gradient-based efficient algorithm for Bayesian experimental design for chemical kinetics. The estimation of information gain is done by a simplified version of Laplacian approximation without taking expectations with respect to possible experimental data. The model-oriented, target-oriented, parameter-oriented, and parallel experimental designs are demonstrated facilitated by novel neural network response surfaces, showing the efficiency and multi-function nature of the proposed methods. Several future directions are worth exploring. First, the error of such a simplification should be related to the Hessian matrix of the response surface, since if the second-order gradient is zero, the approximation is exact. Thus, it should be straightforward to derive the order of error in terms of the second-order derivatives based on the theoretical results of [87], where the integral of information gain is computed and the error analysis is carried out based on the assumption that a large amount of data can be acquired in each time of experiments. Second, the parallel computing of different initialization in design optimization is worth trying, since

they are suitable for parallel computing by nature and the strong parallel ability of GPU cannot be fully exploited. Third, several advanced Bayesian experimental design algorithms without linearizing the response surfaces [113] can be implemented in combustion kinetics, in case the model is highly nonlinear. Fourth, a reduced-order modeling technique can be used to decrease the dimension of model inputs, hence reducing the size of surrogate neural networks. This will accelerate the design process since a smaller network size has a shorter time of inference and gradient calculation.

Appendix A

Additional Tables

Table A.1: The cosine similarity of gradient computed by networks and Cantera for eight species and four time points under five temperatures at nominal values of kinetic parameters. The smallest cosine value for each temperature is marked in red.

		H_2	C_2H_2	CH_4	C_2H_4	C_2H_6	pC_3H_4	C_3H_6	C_3H_8
1250K	0.4ms	0.999524	0.971768	0.998505	0.998949	0.996809	**0.963821**	0.999412	0.999723
	0.8ms	0.999652	0.995641	0.999361	0.999198	0.997512	0.997518	0.999597	0.999787
	1.2ms	0.999587	0.995605	0.999381	0.99943	0.998651	0.997665	0.999513	0.999762
	1.6ms	0.99948	0.993731	0.999116	0.999508	0.999248	0.996961	0.999352	0.999744
1290K	0.4ms	0.999401	**0.981546**	0.998813	0.998375	0.997261	0.981585	0.999082	0.999528
	0.8ms	0.999384	0.994307	0.999512	0.998937	0.998544	0.99629	0.999085	0.999387
	1.2ms	0.999457	0.99574	0.999632	0.999509	0.999547	0.996022	0.999338	0.999545
	1.6ms	0.99947	0.995385	0.999457	0.999669	0.999519	0.995288	0.999351	0.999604
1330K	0.4ms	0.999384	**0.989643**	0.99877	0.998042	0.995958	0.99329	0.998741	0.999206
	0.8ms	0.999385	0.996273	0.999386	0.999021	0.998319	0.995837	0.999155	0.999371
	1.2ms	0.999427	0.996173	0.999091	0.99932	0.998687	0.994377	0.999492	0.999624
	1.6ms	0.999259	0.995123	0.998756	0.999177	0.998091	0.992253	0.999376	0.999564
1370K	0.4ms	0.999422	**0.990121**	0.998782	0.997812	0.9981	0.996025	0.997829	0.999309
	0.8ms	0.99954	0.994189	0.999108	0.999194	0.999155	0.994004	0.999376	0.999673
	1.2ms	0.999098	0.99511	0.998863	0.999262	0.998415	0.99303	0.999709	0.999554
	1.6ms	0.998011	0.995358	0.998648	0.998924	0.997076	0.992195	0.999242	0.999344
1410K	0.4ms	0.999352	0.995818	0.998977	0.999005	0.998795	0.995501	0.998773	0.999686
	0.8ms	0.9987	0.997211	0.999068	0.999219	0.998591	0.993401	0.999269	0.999447
	1.2ms	0.99733	0.996992	0.99857	0.999128	0.998097	**0.993081**	0.99944	0.999227
	1.6ms	0.995921	0.996608	0.997407	0.998485	0.997422	0.993881	0.999234	0.998464

Table A.2: The relative error of gradient vector length ($|\text{Grad}|_{NN}$ -$|\text{Grad}|_{Cantera}$)/ $|\text{Grad}|_{NN}$ computed by networks and Cantera for eight species and four time points under five temperatures. The smallest cosine value for each temperature is marked in red.It is observed that for most point the error is smaller than 10%. There is only one point with large error, 0.68. A potential reason is that the gradient itself is pretty small, so even small error of computation can cause large relative error. In fact, as long as the gradient directions agree well, the length does not matter, since we can tune the step size to control descent speed of the optimizer.

		H_2	C_2H_2	CH_4	C_2H_4	C_2H_6	pC_3H_4	C_3H_6	C_3H_8
1250K	0.4ms	-0.04261	0.07623	-0.03718	-0.02552	-0.0637	0.011884	-0.07523	-0.06384
	0.8ms	0.018307	0.002879	0.005637	0.012365	-0.00657	-0.00257	-0.002	-0.02493
	1.2ms	0.032117	0.051381	0.031779	0.027898	0.017274	0.113878	0.014417	-0.00525
	1.6ms	0.028151	0.067531	0.042481	0.032958	0.033475	0.135964	0.012662	-0.00185
1290K	0.4ms	0.026954	-0.31131	-0.01317	-0.00417	-0.02936	-0.68009	0.041615	-0.00983
	0.8ms	0.05877	-0.11058	0.002488	0.01611	-0.00334	-0.01458	0.074294	0.013045
	1.2ms	0.047494	-0.0274	0.006942	0.018126	0.006635	0.092667	0.060501	0.011771
	1.6ms	0.035608	0.002123	0.007322	0.022987	0.022953	0.101313	0.04648	0.002728
1330K	0.4ms	0.057903	-0.03479	-0.0064	-0.00365	-0.02275	-0.03699	0.044506	0.033288
	0.8ms	0.064359	0.018983	-0.0002	0.012363	-0.00913	0.103288	0.053338	0.028339
	1.2ms	0.045899	0.051497	0.008234	0.021176	0.004642	0.102816	0.035329	0.019915
	1.6ms	0.031287	0.059802	0.012152	0.027387	0.011556	0.077832	0.017074	0.010687
1370K	0.4ms	0.078727	-0.09163	0.006312	-0.0001	-0.00972	0.046629	0.076466	0.058476
	0.8ms	0.044779	0.004639	0.016437	0.005577	0.000669	0.092411	0.049062	0.03088
	1.2ms	0.024171	0.047901	0.035761	0.023727	0.014776	0.092921	0.033645	0.035759
	1.6ms	-0.01913	0.042325	0.024985	0.011226	-0.00576	0.076541	-0.00935	0.0349
1410K	0.4ms	0.056951	-0.03034	0.01868	0.016169	0.014421	0.023269	0.05117	0.04857
	0.8ms	0.027899	0.055021	0.047256	0.014062	0.032621	0.034637	0.022931	0.037312
	1.2ms	0.013136	0.086568	0.065721	0.012974	0.042735	0.048791	0.01351	0.048304
	1.6ms	0.001265	0.071913	0.049872	-0.01384	0.016931	0.045427	-0.01602	0.073247

Table A.3: The differences between NEs of original mechanism and NEs of LastP. Negative differences are in bold.

	H_2	C_2H_2	CH_4	C_2H_4	C_2H_6	pC_3H_4	C_3H_6	C_3H_8
1250 K	0.6040	**-0.0114**	0.2538	0.7644	0.2596	0.0019	0.7490	1.5679
1290 K	0.7939	0.0225	0.3609	1.0303	0.3308	**-0.0021**	0.8560	2.1016
1330 K	1.0258	**-0.0224**	0.4093	0.9531	0.3222	**-0.0345**	0.7115	2.1127
1370 K	0.9864	**-0.0890**	0.3763	0.2028	0.2575	**-0.0442**	0.6529	1.8252
1410 K	0.8291	**-0.0395**	0.2649	**-0.3527**	0.1780	**-0.0394**	0.7368	1.4990

Table A.4: The differences between NEs of original mechanism and NEs of C-1. Negative differences are in bold.

	H_2	C_2H_2	CH_4	C_2H_4	C_2H_6	pC_3H_4	C_3H_6	C_3H_8
1250 k	0.5624	**-0.0101**	0.2264	0.6479	0.2165	0.0019	0.6910	1.3547
1290 k	0.7982	0.0218	0.3249	0.8747	0.2688	**-0.0015**	0.9200	1.8717
1330 k	0.9417	**-0.0203**	0.3754	0.9287	0.2515	**-0.0328**	0.8575	2.1013
1370 k	0.9184	**-0.0889**	0.3577	0.3451	0.1849	**-0.0420**	0.7978	1.9363
1410 k	0.7760	**-0.0449**	0.2757	**-0.2495**	0.1081	**-0.0367**	0.7968	1.5470

Table A.5: The differences between NEs of original mechanism and NEs of C-160. Negative differences are in bold.

	H_2	C_2H_2	CH_4	C_2H_4	C_2H_6	pC_3H_4	C_3H_6	C_3H_8
1250 K	0.5442	**-0.0091**	0.3904	0.7247	0.1960	0.0016	0.6992	1.4355
1290 K	0.7949	0.0217	0.5457	0.9639	0.2754	**-0.0090**	0.9291	2.0145
1330 K	0.9675	**-0.0224**	0.5455	0.9541	0.3237	**-0.0642**	0.8005	2.1588
1370 K	0.9902	**-0.1133**	0.5135	0.3572	0.3568	**-0.1115**	0.7415	1.8649
1410 K	0.8619	**-0.1114**	0.3552	**-0.0078**	0.4042	**-0.1552**	0.7761	1.5341

Table A.6: Inner product of the normalized sensitivity vectors of pC_3H_4 for 1370K at 1.6 ms and at 0.4 ms/0.8 ms/1.2 ms. Values below 0.9 are in bold.

Species/Time	0.4 ms	0.8 ms	1.2 ms
H_2	0.965193	0.986311	0.999998
C_2H_2	0.953348	0.987103	0.999999
CH_4	0.931971	0.983121	0.999999
C_2H_4	0.911628	0.978563	0.999999
C_2H_6	**0.871753**	0.971632	0.999998
pC_3H_4	**0.880802**	0.961229	0.999997
C_3H_6	0.940318	0.982524	0.999999
C_3H_8	0.947249	0.982917	0.999998

Table A.7: Inner product of the normalized sensitivity vectors of pC_3H_4 for 1330K at 1.6 ms and at 0.4 ms/0.8 ms/1.2 ms. Values below 0.9 are in bold.

Species/Time	0.4 ms	0.8 ms	1.2 ms
H_2	0.984654	0.99402	0.999999
C_2H_2	0.962212	0.988744	0.999999
CH_4	0.943885	0.983331	0.999999
C_2H_4	0.920733	0.976031	0.999998
C_2H_6	**0.86429**	0.962936	0.999997
pC_3H_4	0.927775	0.975203	0.999997
C_3H_6	0.961514	0.987019	0.999999
C_3H_8	0.971453	0.989322	0.999999

Table A.8: Inner product of the normalized sensitivity vectors of pC_3H_4 for 1290K at 1.6 ms and at 0.4 ms/0.8 ms/1.2 ms. No Value is below 0.9.

Species/Time	0.4 ms	0.8 ms	1.2 ms
H_2	0.993482	0.997403	1
C_2H_2	0.975403	0.9916	0.999999
CH_4	0.968949	0.989268	0.999999
C_2H_4	0.954687	0.983247	0.999998
C_2H_6	0.908057	0.966932	0.999997
pC_3H_4	0.957058	0.986341	0.999999
C_3H_6	0.982199	0.993385	0.999999
C_3H_8	0.98882	0.995429	0.999999

Table A.9: Inner product of the normalized sensitivity vectors of pC_3H_4 for 1250K at 1.6 ms and at 0.4 ms/0.8 ms/1.2 ms. No Value is below 0.9.

Species/Time	0.4 ms	0.8 ms	1.2 ms
H_2	0.997563	0.999012	1
C_2H_2	0.988674	0.995734	0.999999
CH_4	0.98806	0.995659	0.999999
C_2H_4	0.984118	0.993341	0.999999
C_2H_6	0.964769	0.984702	0.999998
pC_3H_4	0.971277	0.990949	0.999999
C_3H_6	0.99377	0.997611	1
C_3H_8	0.996474	0.998535	1

Appendix B

Additional Figures

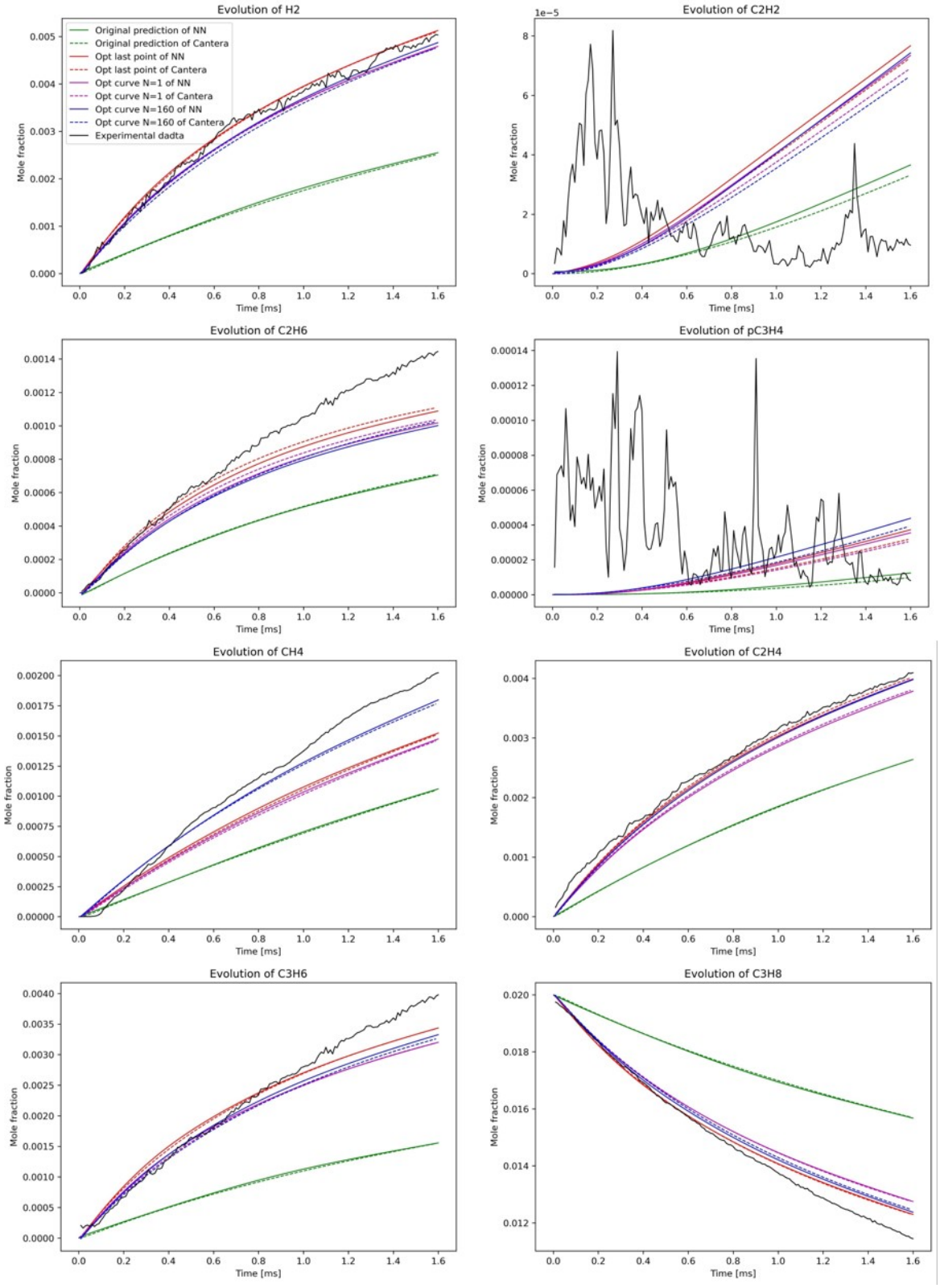


Figure B.1: The comparison between the prediction of neural networks and Cantera for eight species at 1250K.

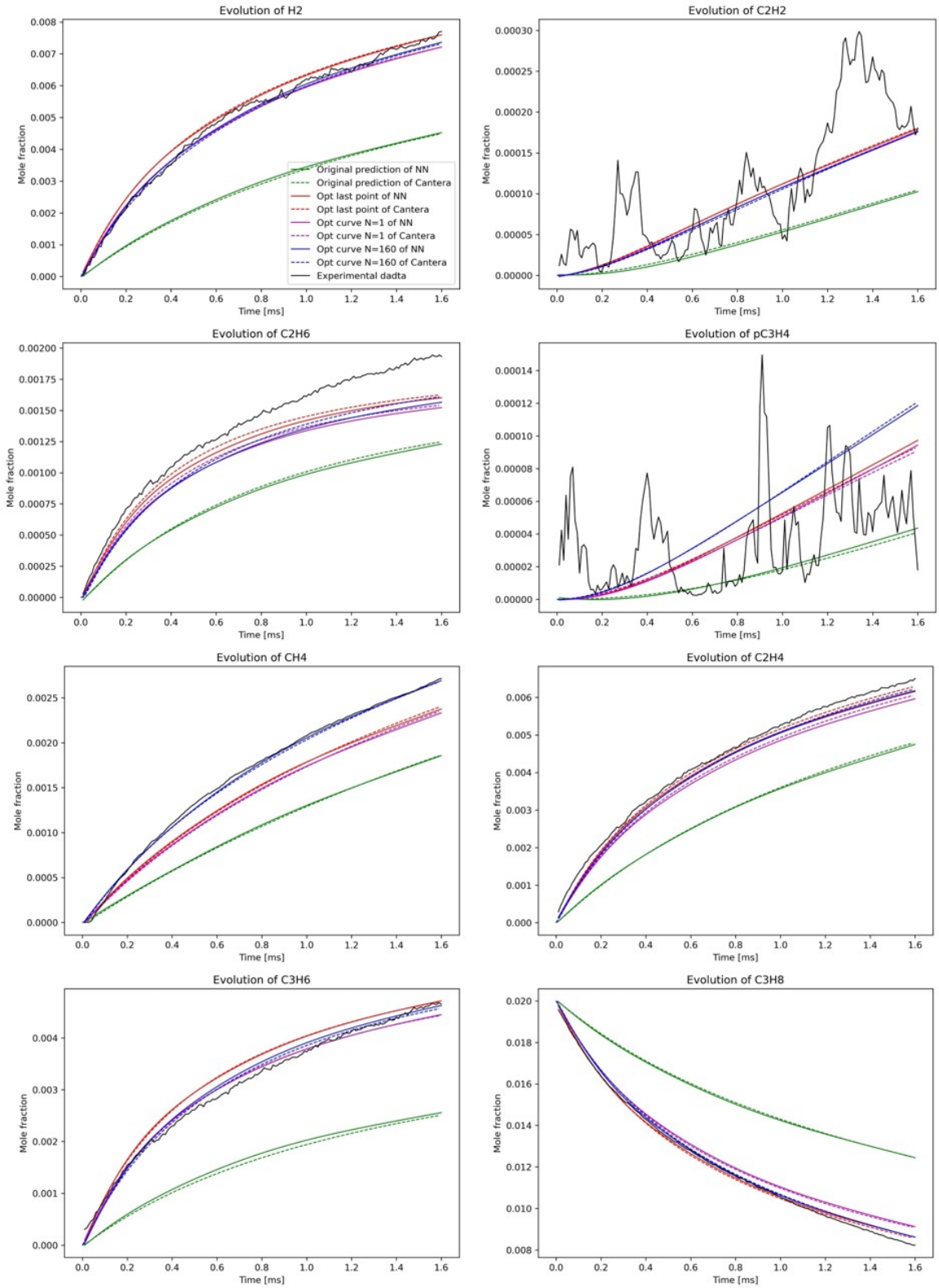


Figure B.2: The comparison between the prediction of neural networks and Cantera for eight species at 1290K.

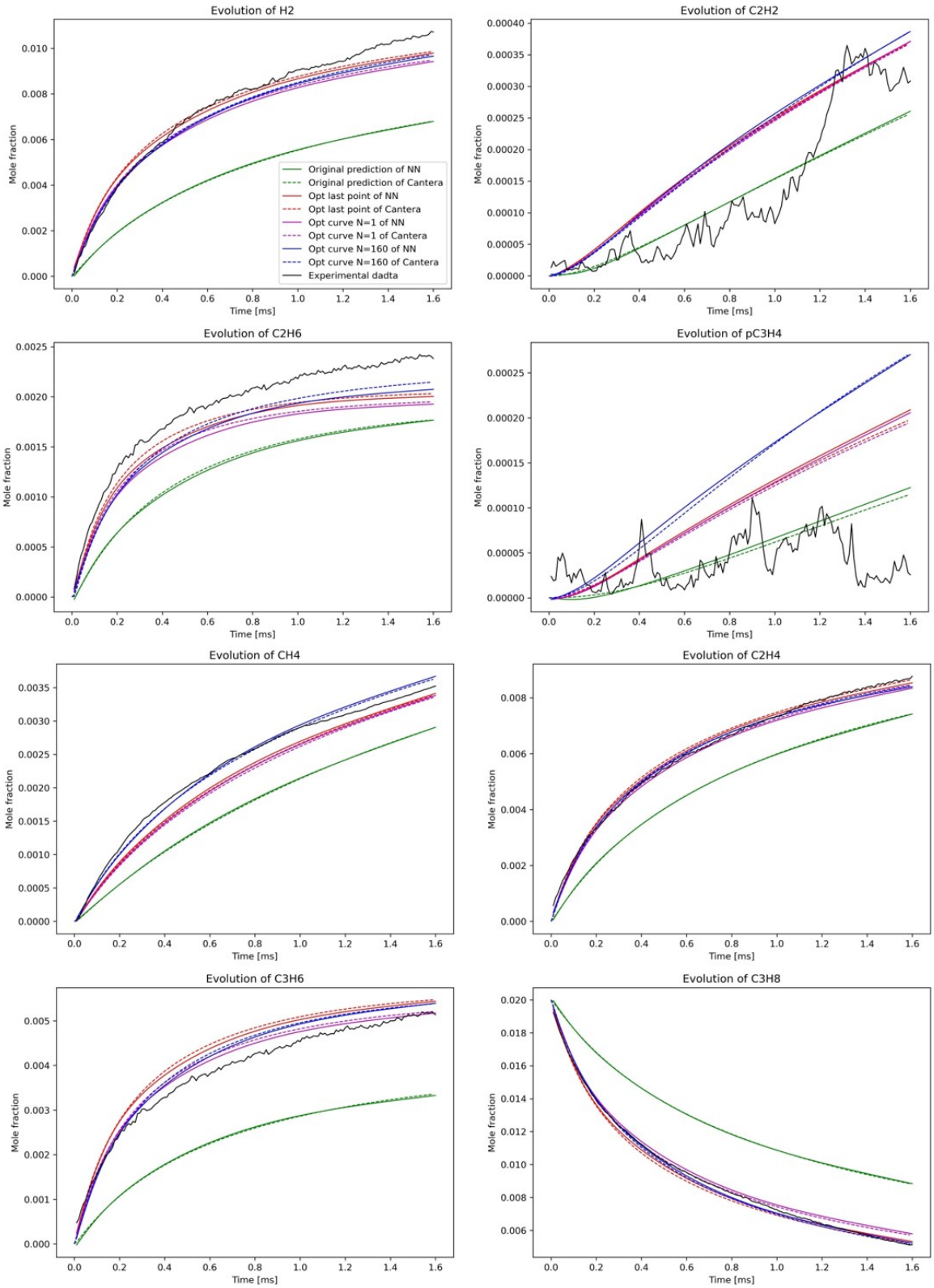


Figure B.3: The comparison between the prediction of neural networks and Cantera for eight species at 1330K.

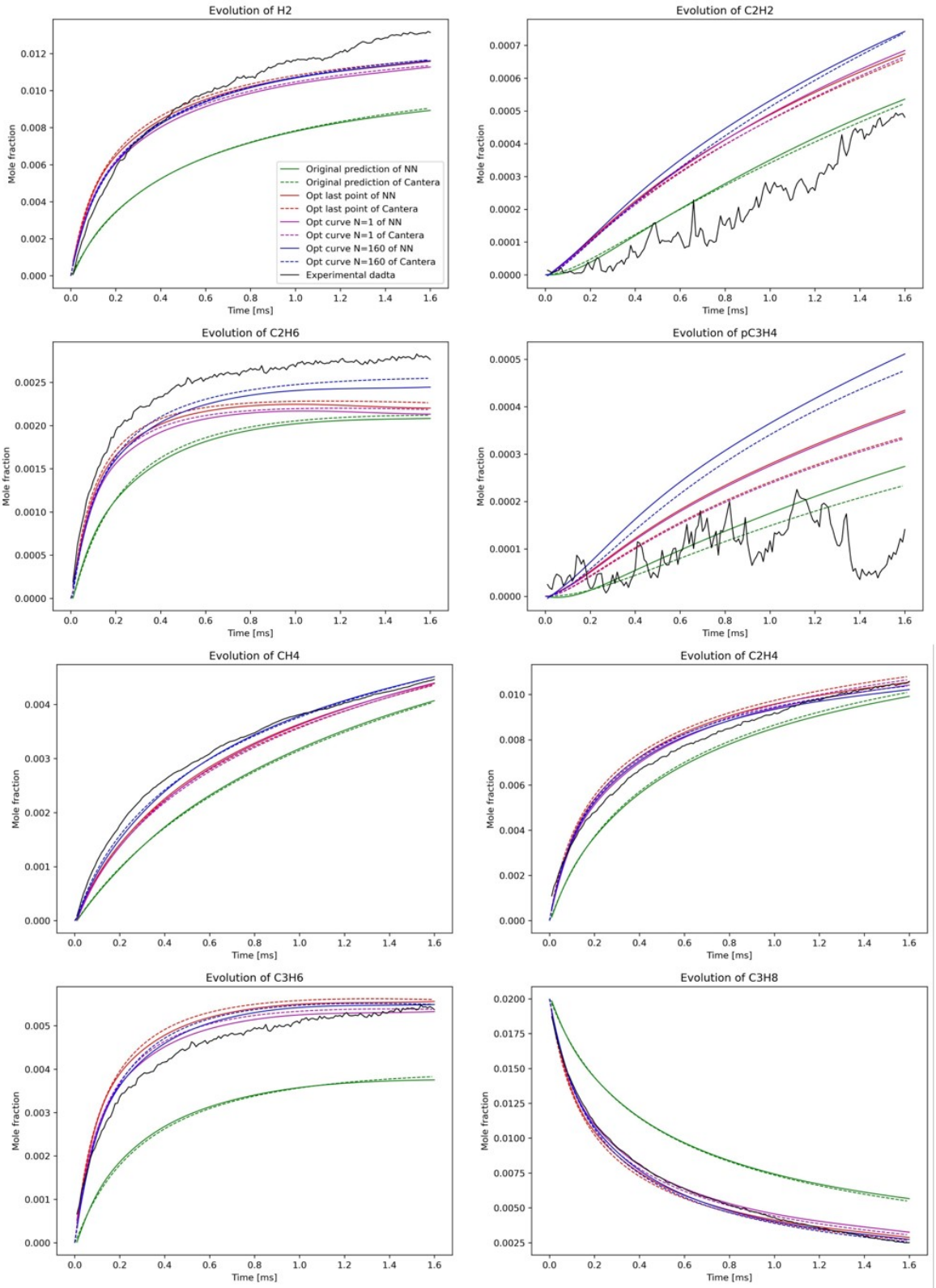


Figure B.4: The comparison between the prediction of neural networks and Cantera for eight species at 1370K.

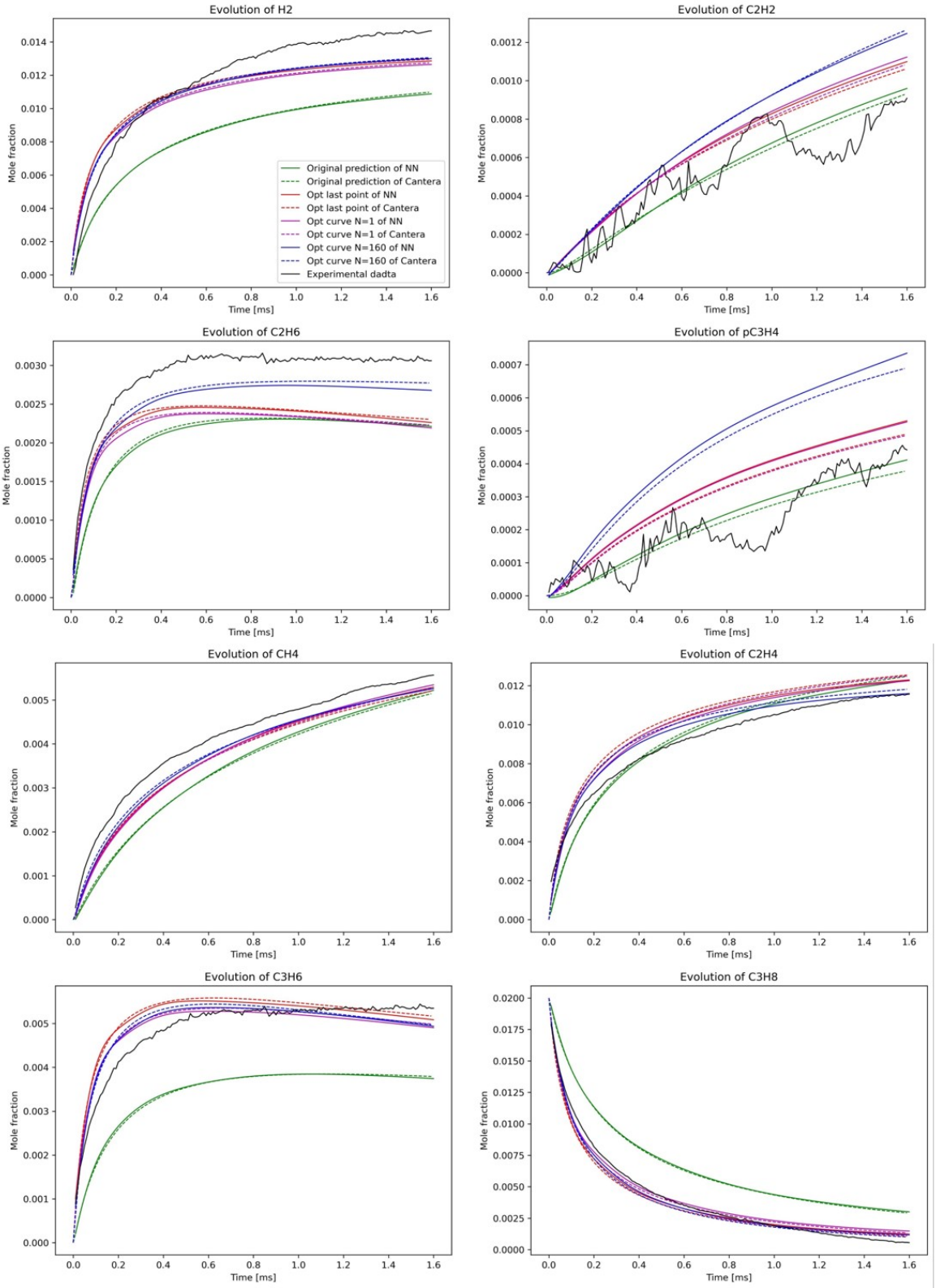


Figure B.5: The comparison between the prediction of neural networks and Cantera for eight species at 1410K.

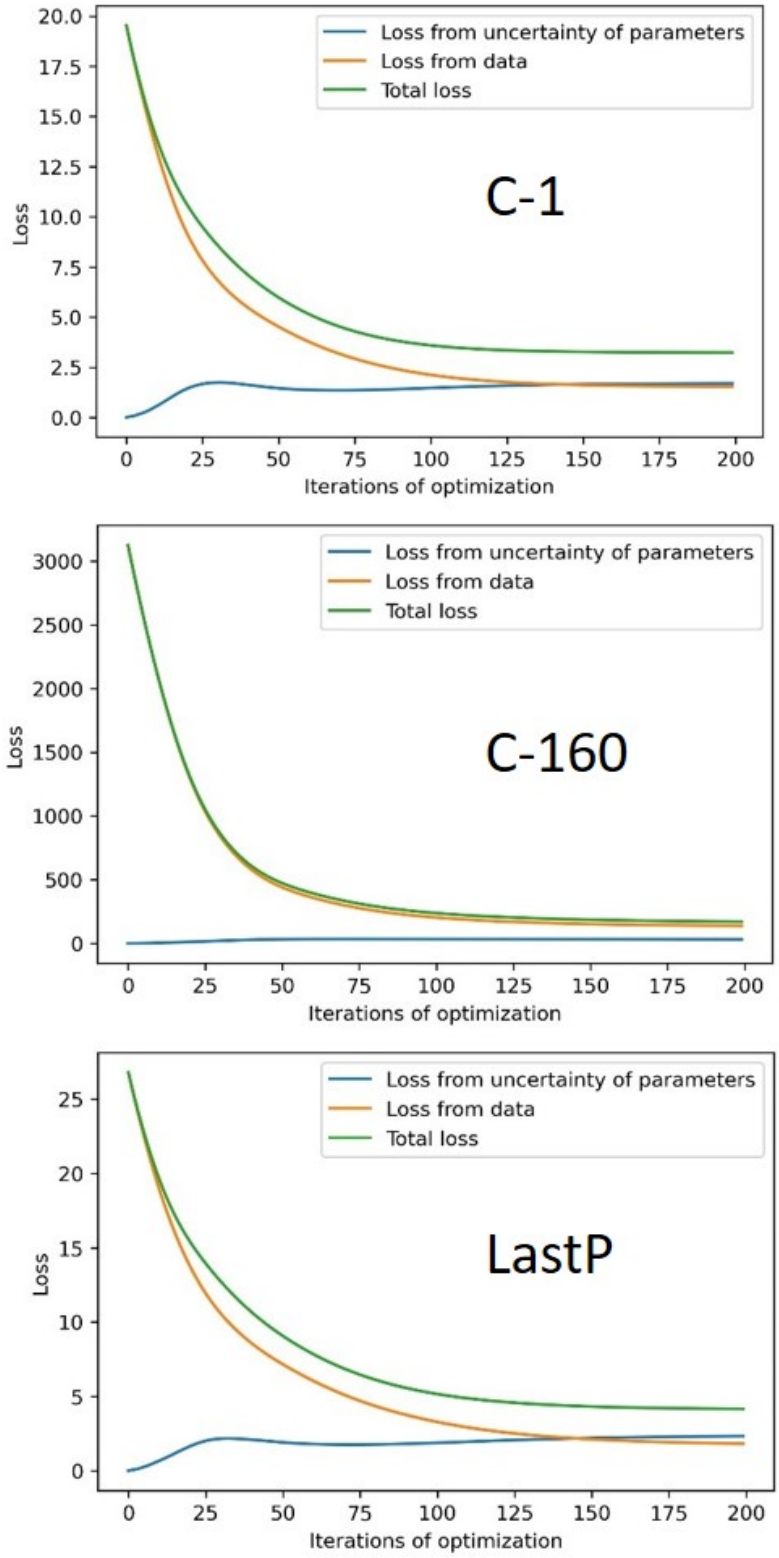


Figure B.6: The evolution loss function versus iteration for C-1, C-160 and LastP.

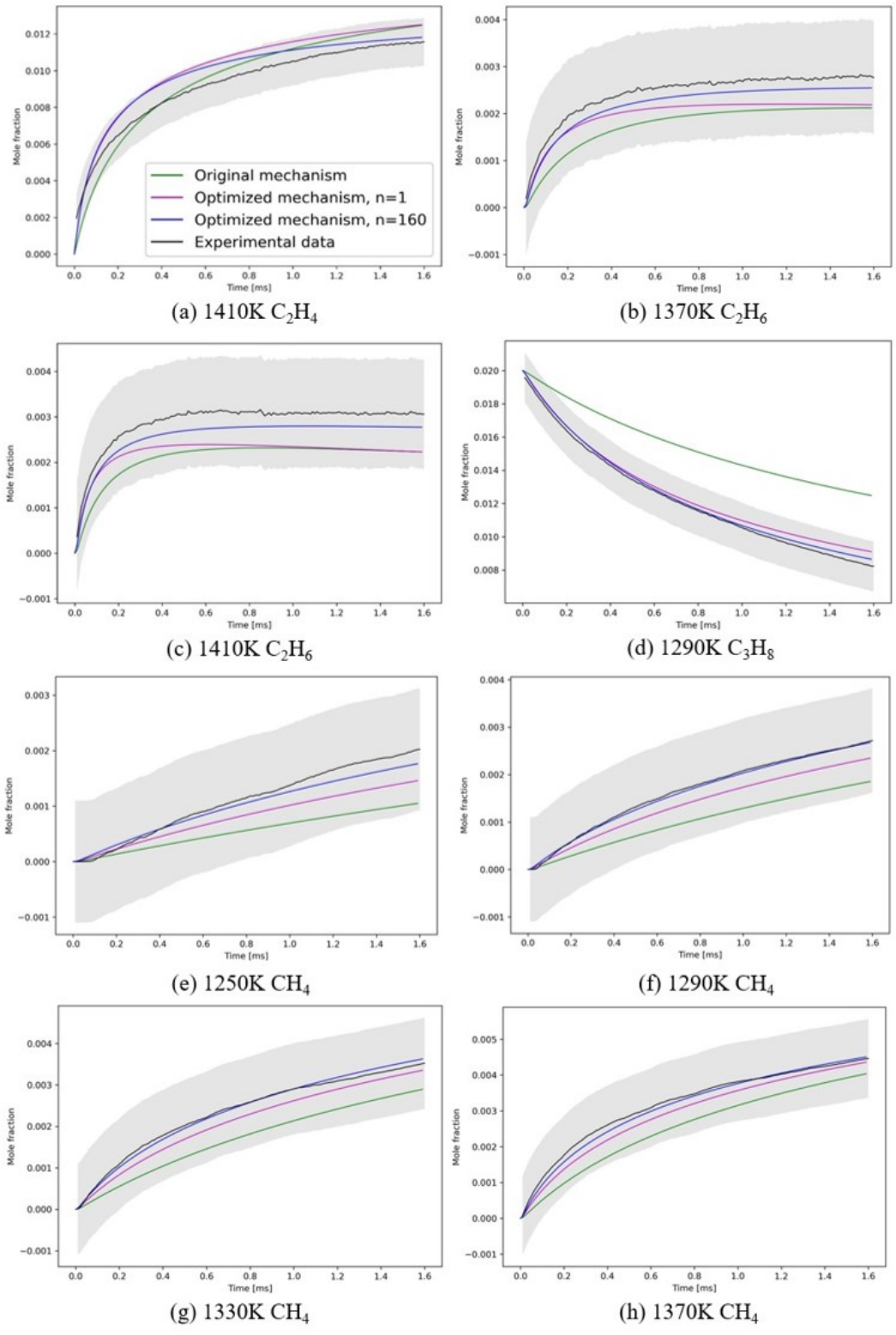
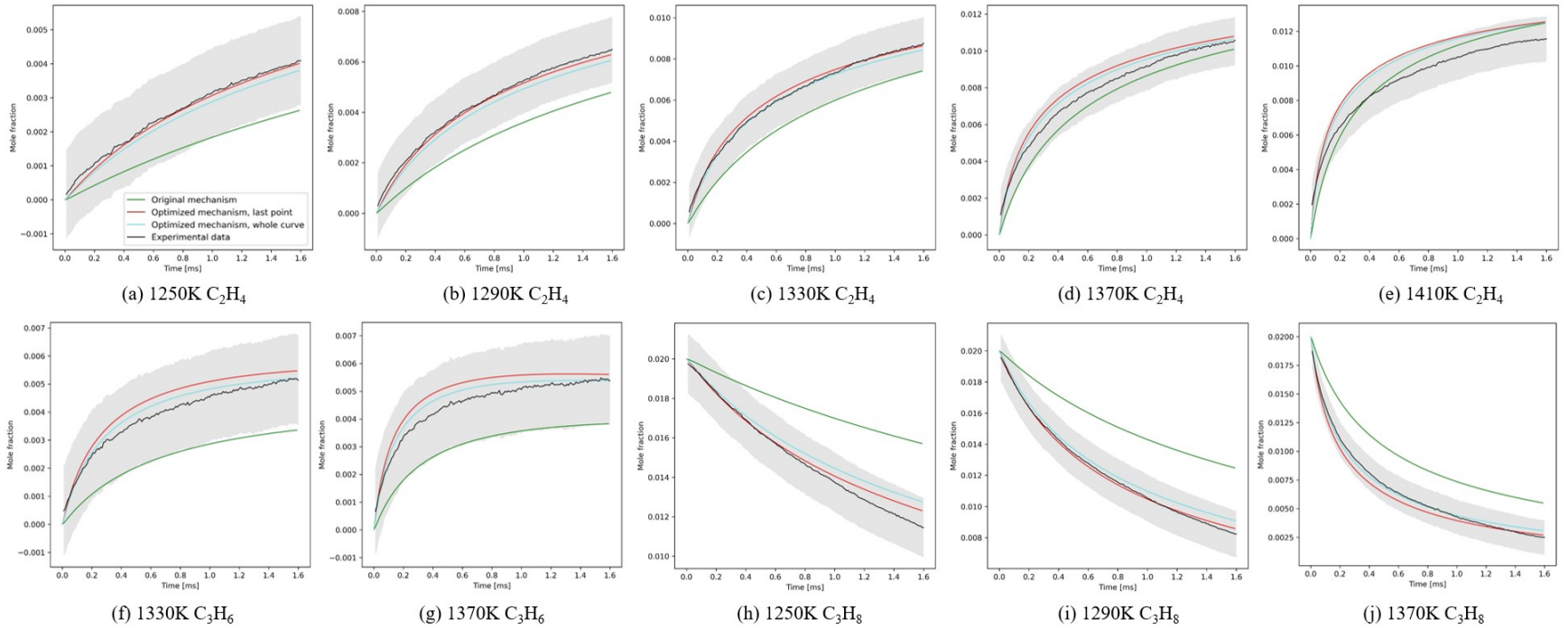


Figure B.7: The comparison of the species evolution of C-160 and C-1 under the scenarios where the difference between NEs of C-160 and C-1 is larger than 10% of corresponding experimental uncertainties (in bold in Table 2.2)



(a) 1250K C_2H_4 (b) 1290K C_2H_4 (c) 1330K C_2H_4 (d) 1370K C_2H_4 (e) 1410K C_2H_4

(f) 1330K C_3H_6 (g) 1370K C_3H_6 (h) 1250K C_3H_8 (i) 1290K C_3H_8 (j) 1370K C_3H_8

Figure B.8: The comparison of the species evolution of LastP and C-1 under the scenarios where the difference between NEs of C-160 and C-1 is larger than 10% of corresponding experimental uncertainties (in bold in Table 2.4), plus C_2H_4 at 1330K

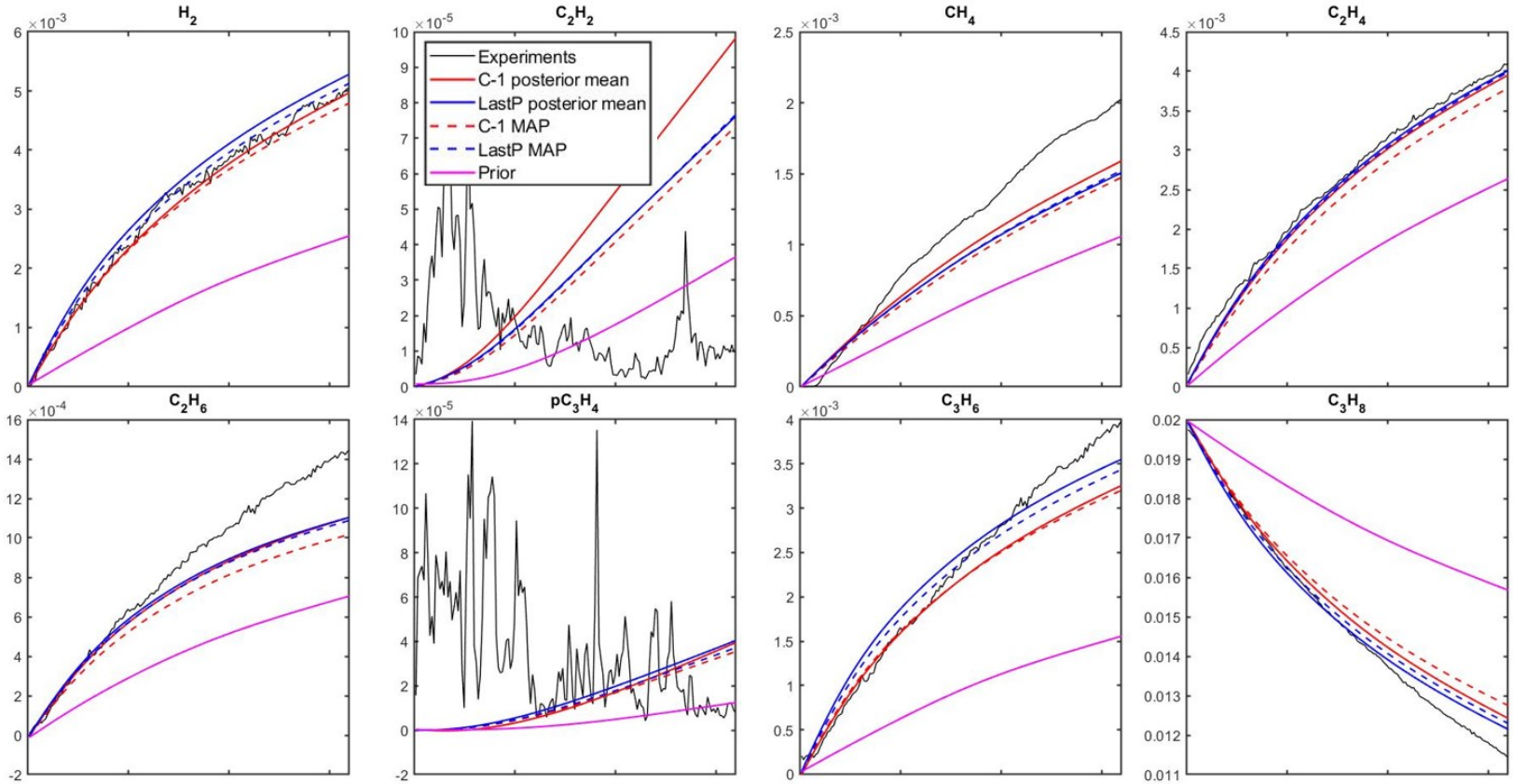


Figure B.9: Comparison of eight species profiles predicted by posterior mean and MAP parameters for LastP, C-1 and prior models at 1250K.

References

[1] S. B. Pope, "Small scales, many species and the manifold challenges of turbulent combustion," *Proceedings of the Combustion Institute*, vol. 34, no. 1, pp. 1–31, 2013.

[2] H. Chen, M. Tao, Q. Yang, H. Ge, and P. Zhao, "Two-stage autoignition and combustion mode evolution in boundary layer flows above a cold flat plate," *Proceedings of the Combustion Institute*, vol. 38, no. 1, pp. 767–776, 2021.

[3] Y. Shi, H.-W. Ge, and R. D. Reitz, *Computational optimization of internal combustion engines*. Springer Science & Business Media, 2011.

[4] H. Xiao and P. Cinnella, "Quantification of model uncertainty in rans simulations: A review," *Progress in Aerospace Sciences*, vol. 108, pp. 1–31, 2019.

[5] H. Wang and D. A. Sheen, "Combustion kinetic model uncertainty quantification, propagation and minimization," *Progress in Energy and Combustion Science*, vol. 47, pp. 1–31, 2015.

[6] C. K. Law, *Combustion Physics*. Cambridge university press, 2010.

[7] B. Yang, "Towards predictive combustion kinetic models: Progress in model analysis and informative experiments," *Proceedings of the Combustion Institute*, vol. 38, no. 1, pp. 199–222, 2021.

[8] S. Wang, Y. Ding, J. Miao, and R. K. Hanson, "Shock tube and multi-species laser absorption measurements of rate constants for methanol pyrolysis," *Proceedings of the Combustion Institute*, vol. 39, no. 1, pp. 755–763, 2023.

[9] J. A. Miller, R. Sivaramakrishnan, Y. Tao, C. F. Goldsmith, M. P. Burke, A. W. Jasper, N. Hansen, N. J. Labbe, P. Glarborg, and J. Zádor, "Combustion chemistry in the twenty-first century: Developing theory-informed chemical kinetics models," *Progress in Energy and Combustion Science*, vol. 83, p. 100 886, 2021.

[10] H. Wang, X. You, A. V. Joshi, S. G. David, A. Laskin, F. Egolfopoulos, and C. K. Law, *USC Mech version II: High-temperature combustion reaction model of H_2*, http://ignis.usc.edu/USC_Mech_II.htm, May, 2007.

[11] S. G. Davis, A. V. Joshi, H. Wang, and F. Egolfopoulos, "An optimized kinetic model of h2/co combustion," *Proceedings of the combustion Institute*, vol. 30, no. 1, pp. 1283–1292, 2005.

[12] M. O'Conaire, H. J. Curran, J. M. Simmie, W. J. Pitz, and C. K. Westbrook, "A comprehensive modeling study of hydrogen oxidation," *International journal of chemical kinetics*, vol. 36, no. 11, pp. 603–622, 2004.

[13] P. Saxena and F. A. Williams, "Testing a small detailed chemical-kinetic mechanism for the combustion of hydrogen and carbon monoxide," *Combustion and Flame*, vol. 145, no. 1-2, pp. 316–323, 2006.

[14] H. Sun, S. Yang, G. Jomaas, and C. K. Law, "High-pressure laminar flame speeds and kinetic modeling of carbon monoxide/hydrogen combustion," *Proceedings of the Combustion Institute*, vol. 31, no. 1, pp. 439–446, 2007.

[15] J. Li, Z. Zhao, A. Kazakov, M. Chaos, F. L. Dryer, and J. J. Scire Jr, "A comprehensive kinetic mechanism for co, ch2o, and ch3oh combustion," *International Journal of Chemical Kinetics*, vol. 39, no. 3, pp. 109–136, 2007.

[16] A. A. Konnov, "Remaining uncertainties in the kinetic mechanism of hydrogen combustion," *Combustion and flame*, vol. 152, no. 4, pp. 507–528, 2008.

[17] M. P. Burke, M. Chaos, F. L. Dryer, and Y. Ju, "Negative pressure dependence of mass burning rates of h2/co/o2/diluent flames at low flame temperatures," *Combustion and Flame*, vol. 157, no. 4, pp. 618–631, 2010.

[18] T. Lu and C. K. Law, "Toward accommodating realistic fuel chemistry in large-scale computations," *Progress in Energy and Combustion Science*, vol. 35, no. 2, pp. 192–215, 2009, ISSN: 0360-1285.

[19] K. Braman, T. A. Oliver, and V. Raman, "Bayesian analysis of syngas chemistry models," *Combustion Theory and Modelling*, vol. 17, no. 5, pp. 858–887, 2013.

[20] J. Wang, Z. Zhou, K. Lin, C. K. Law, and B. Yang, "Facilitating bayesian analysis of combustion kinetic models with artificial neural network," *Combustion and Flame*, vol. 213, pp. 87–97, 2020.

[21] D. Miller and M. Frenklach, "Sensitivity analysis and parameter estimation in dynamic modeling of chemical kinetics," *International Journal of Chemical Kinetics*, vol. 15, no. 7, pp. 677–696, 1983.

[22] M. Frenklach, H. Wang, and M. J. Rabinowitz, "Optimization and analysis of large chemical kinetic mechanisms using the solution mapping method—combustion of methane," *Progress in Energy and Combustion Science*, vol. 18, no. 1, pp. 47–73, 1992, ISSN: 0360-1285.

[23] G. E. Box, W. H. Hunter, and S. Hunter, *Statistics for Experimenters: Design, Innovation, and Discovery*. John Wiley and Sons New York, 2005, vol. 664.

[24] G. P. Smith, D. M. Golden, M. Frenklach, *et al.*, *GRI-Mech 3.0*, http://www.me.berkley.edu/gri-mech/, 1999.

[25] M. Frenklach, A. Packard, P. Seiler, and R. Feeley, "Collaborative data processing in developing predictive models of complex reaction systems," *International journal of chemical kinetics*, vol. 36, no. 1, pp. 57–66, 2004.

[26] R. Feeley, P. Seiler, A. Packard, and M. Frenklach, "Consistency of a reaction dataset," *The Journal of Physical Chemistry A*, vol. 108, no. 44, pp. 9573–9583, 2004.

[27] R. Feeley, M. Frenklach, M. Onsum, T. Russi, A. Arkin, and A. Packard, "Model discrimination using data collaboration," *The Journal of Physical Chemistry A*, vol. 110, no. 21, pp. 6803–6813, 2006.

[28] W. Li, A. Hegde, J. Oreluk, A. Packard, and M. Frenklach, "Representing model discrepancy in bound-to-bound data collaboration," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 9, no. 1, pp. 231–259, 2021.

[29] D. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*. Athena Scientific, 2008.

[30] D. A. Sheen, X. You, H. Wang, and T. Løvås, "Spectral uncertainty quantification, propagation and optimization of a detailed kinetic model for ethylene combustion," *Proceedings of the Combustion Institute*, vol. 32, no. 1, pp. 535–542, 2009.

[31] V. Peterka, "Bayesian approach to system identification," in *Trends and Progress in System identification*, Elsevier, 1981, pp. 239–304.

[32] J. L. Beck, "Statistical system identification of structures," in *Proceedings of international conference on structural safety and reliability*, ASCE, 1989, pp. 1395–1402.

[33] J. L. Beck and L. S. Katafygiotis, "Updating models and their uncertainties. i: Bayesian statistical framework," *Journal of Engineering Mechanics*, vol. 124, no. 4, pp. 455–461, 1998.

[34] A. Vikhansky, M. Kraft, M. Simon, S. Schmidt, and H.-J. Bart, "Droplets population balance in a rotating disc contactor: An inverse problem approach," *AIChE Journal*, vol. 52, no. 4, pp. 1441–1450, 2006.

[35] H. N. Najm, B. J. Debusschere, Y. M. Marzouk, S. Widmer, and O. Le Maıtre, "Uncertainty quantification in chemical systems," *International journal for numerical methods in engineering*, vol. 80, no. 6-7, pp. 789–814, 2009.

[36] K. Miki, S. H. Cheung, E. E. Prudencio, and P. L. Varghese, "Bayesian uncertainty quantification of recent shock tube determinations of the rate coefficient of reaction h+ o2→ oh+ o," *International Journal of Chemical Kinetics*, vol. 44, no. 9, pp. 586–597, 2012.

[37] K. Miki, E. E. Prudencio, S. H. Cheung, and G. Terejanu, "Using bayesian analysis to quantify uncertainties in the h+ o2→ oh+ o reaction," *Combustion and flame*, vol. 160, no. 5, pp. 861–869, 2013.

[38] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006.

[39] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[40] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[41] L. Cai, H. Pitsch, S. Y. Mohamed, V. Raman, J. Bugler, H. Curran, and S. M. Sarathy, "Optimized reaction mechanism rate rules for ignition of normal alkanes," *Combustion and Flame*, vol. 173, pp. 468–482, 2016.

[42] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid monte carlo," *Physics Letters B*, vol. 195, no. 2, pp. 216–222, 1987.

[43] Q. Li, H. Chen, B. C. Koenig, and S. Deng, "Bayesian chemical reaction neural network for autonomous kinetic uncertainty quantification," *Physical Chemistry Chemical Physics*, vol. 25, no. 5, pp. 3707–3717, 2023.

[44] W. Ji and S. Deng, "Autonomous discovery of unknown reaction pathways from data by chemical reaction neural network," *The Journal of Physical Chemistry A*, vol. 125, no. 4, pp. 1082–1092, 2021.

[45] D. A. Sheen and H. Wang, "The method of uncertainty quantification and minimization using polynomial chaos expansions," *Combustion and Flame*, vol. 158, no. 12, pp. 2358–2374, 2011.

[46] Q. Huang, C. Pu, K. Khosoussi, D. M. Rosen, D. Fourie, J. P. How, and J. J. Leonard, "Incremental non-gaussian inference for slam using normalizing flows," *IEEE Transactions on Robotics*, vol. 39, no. 2, pp. 1458–1475, 2022.

[47] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American Statistical Association*, vol. 112, pp. 859–877, 2017.

[48] J. Na, J. H. Bak, and N. V. Sahinidis, "Efficient bayesian inference using adversarial machine learning and low-complexity surrogate models," *Computers & Chemical Engineering*, vol. 151, p. 107 322, 2021.

[49] M. Ohen and D. G. Vlachos, "Chemical kinetics bayesian inference toolbox (ckbit)," *Computer Physics Communications*, vol. 265, p. 107 989, 2021.

[50] S. G. Davis, A. B. Mhadeshwar, D. G. Vlachos, and H. Wang, "A new approach to response surface development for detailed gas-phase and surface reaction kinetic model optimization," *International Journal of Chemical Kinetics*, vol. 36, no. 2, pp. 94–106, 2004.

[51] R. Ghanem, "Stochastic finite element expansion for random media," *Journal of Engineering Mechanics*, vol. 115, no. 5, pp. 1035–1053, 1989.

[52] M. T. Reagana, H. N. Najm, R. G. Ghanem, and O. M. Knio, "Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection," *Combustion and flame*, vol. 132, no. 3, pp. 545–555, 2003.

[53] M. Reagan, H. Najm, B. Debusschere, O. Le Maıtre, O. Knio, and R. Ghanem, "Spectral stochastic uncertainty quantification in chemical systems," *Combustion Theory and Modelling*, vol. 8, no. 3, p. 607, 2004.

[54] B. D. Phenix, J. L. Dinaro, M. A. Tatang, J. W. Tester, J. B. Howard, and G. J. McRae, "Incorporation of parametric uncertainty into complex kinetic mechanisms: Application to hydrogen oxidation in supercritical water," *Combustion and Flame*, vol. 112, no. 1-2, pp. 132–146, 1998.

[55] D. Xiu and G. E. Karniadakis, "The wiener–askey polynomial chaos for stochastic differential equations," *SIAM journal on scientific computing*, vol. 24, no. 2, pp. 619–644, 2002.

[56] B. Debusschere, "Intrusive polynomial chaos methods for forward uncertainty propagation," in *Handbook of Uncertainty Quantification*, R. Ghanem, D. Higdon, and H. Owhadi, Eds., Springer International Publishing, 2017, pp. 617–636.

[57] H. Rabitz and Ö. F. Aliş, "General foundations of high-dimensional model representations," *Journal of Mathematical Chemistry*, vol. 25, no. 2-3, pp. 197–233, 1999.

[58] G. Li, C. Rosenthal, and H. Rabitz, "High dimensional model representations," *The Journal of Physical Chemistry A*, vol. 105, no. 33, pp. 7765–7777, 2001.

[59] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[60] W. Jiao, W. Wang, J.-t. Huang, X. Wang, and Z. Tu, "Is chatgpt a good translator? a preliminary study," *arXiv preprint arXiv:2301.08745*, 2023.

[61] F. Agostinelli, S. McAleer, A. Shmakov, and P. Baldi, "Solving the rubik's cube with deep reinforcement learning and search," *Nature Machine Intelligence*, vol. 1, no. 8, pp. 356–363, 2019.

[62] A. Krizhevsky, G. Hinton, *et al.*, "Learning multiple layers of features from tiny images," 2009.

[63] R. K. Tripathy and I. Bilionis, "Deep uq: Learning deep neural network surrogate models for high dimensional uncertainty quantification," *Journal of Computational Physics*, vol. 375, pp. 565–588, 2018.

[64] Y. Zhu and N. Zabaras, "Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification," *Journal of Computational Physics*, vol. 366, pp. 415–447, 2018.

[65] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators," *Nature Machine Intelligence*, vol. 3, no. 3, pp. 218–229, 2021, ISSN: 2522-5839.

[66] Z. Li, N. B. Kovachki, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, A. Anandkumar, *et al.*, "Fourier neural operator for parametric partial differential equations," in *International Conference on Learning Representations*, 2020.

[67] Y. Zhang, W. Dong, L. A. Vandewalle, R. Xu, G. P. Smith, and H. Wang, "Neural network approach to response surface development for reaction model optimization and uncertainty minimization," *Combustion and Flame*, vol. 251, p. 112 679, 2023.

[68] B. C. Koenig, W. Ji, and S. Deng, "Kinetic subspace investigation using neural network for uncertainty quantification in nonpremixed flamelets," *Proceedings of the Combustion Institute*, vol. 39, no. 4, pp. 5229–5238, 2023.

[69] H. Chen, W. Ji, S. J. Cassady, A. M. Ferris, R. K. Hanson, and S. Deng, "Using shock tube species time-histories in bayesian parameter estimation: Effective independent-data number and target selection," *Proceedings of the Combustion Institute*, vol. 39, no. 4, pp. 5299–5308, 2023.

[70] J. Duvall, K. Duraisamy, and S. Pan, "Non-linear Independent Dual System (NIDS) for Discretization-independent Surrogate Modeling over Complex Geometries," *arXiv:2109.07018 [physics]*, 2021.

[71] Z. Hong, K.-Y. Lam, R. Sur, S. Wang, D. F. Davidson, and R. K. Hanson, "On the rate constants of oh+ ho2 and ho2+ ho2: A comprehensive study of h2o2 thermal decomposition using multi-species laser absorption," *Proceedings of the Combustion Institute*, vol. 34, no. 1, pp. 565–571, 2013.

[72] R. Hanson and D. Davidson, "Recent advances in laser absorption and shock tube methods for studies of combustion chemistry," *Progress in Energy and Combustion Science*, vol. 44, pp. 103–114, 2014, ISSN: 0360-1285.

[73] S. S. Goldsborough, S. Hochgreb, G. Vanhove, M. S. Wooldridge, H. J. Curran, and C.-J. Sung, "Advances in rapid compression machine studies of low-and intermediate-temperature autoignition phenomena," *Progress in Energy and Combustion Science*, vol. 63, pp. 1–78, 2017.

[74] T. Rainforth, A. Foster, D. R. Ivanova, and F. B. Smith, "Modern bayesian experimental design," *arXiv preprint arXiv:2302.14545*, 2023.

[75] E. G. Ryan, C. C. Drovandi, J. M. McGree, and A. N. Pettitt, "A review of modern computational algorithms for bayesian optimal design," *International Statistical Review*, vol. 84, no. 1, pp. 128–154, 2016.

[76] A. S. Tomlin, "The role of sensitivity and uncertainty analysis in combustion modelling," *Proceedings of the Combustion Institute*, vol. 34, no. 1, pp. 159–176, 2013.

[77] T. Turányi, "Sensitivity analysis of complex kinetic systems. tools and applications," *Journal of Mathematical Chemistry*, vol. 5, no. 3, pp. 203–248, 1990.

[78] T. Russi, A. Packard, R. Feeley, and M. Frenklach, "Sensitivity analysis of uncertainty in model prediction," *The Journal of Physical Chemistry A*, vol. 112, no. 12, pp. 2579–2588, 2008.

[79] D. A. Sheen and J. A. Manion, "Kinetics of the reactions of h and ch3 radicals with n-butane: An experimental design study using reaction network analysis," *The Journal of Physical Chemistry A*, vol. 118, no. 27, pp. 4929–4941, 2014.

[80] É. Valkó, M. Papp, M. Kovács, T. Varga, I. G. Zsély, T. Nagy, and T. Turányi, "Design of combustion experiments using differential entropy," *Combustion Theory and Modelling*, vol. 26, no. 1, pp. 67–90, 2022.

[81] Z. Zhou, K. Lin, Y. Wang, J. Wang, C. K. Law, and B. Yang, "Optex: An integrated framework for experimental design and combustion kinetic model optimization," *Combustion and Flame*, vol. 245, p. 112 298, 2022.

[82] S. Li, T. Tao, J. Wang, B. Yang, C. K. Law, and F. Qi, "Using sensitivity entropy in experimental design for uncertainty minimization of combustion kinetic models," *Proceedings of the Combustion Institute*, vol. 36, no. 1, pp. 709–716, 2017.

[83] J. Wang, S. Li, and B. Yang, "Combustion kinetic model development using surrogate model similarity method," *Combustion Theory and Modelling*, vol. 22, no. 4, pp. 777–794, 2018.

[84] K. Lin, Z. Zhou, Y. Wang, C. K. Law, and B. Yang, "Using active subspace-based similarity analysis for design of combustion experiments," *Proceedings of the Combustion Institute*, vol. 39, no. 4, pp. 5177–5186, 2023.

[85] X. Huan and Y. M. Marzouk, "Simulation-based optimal bayesian experimental design for nonlinear systems," *Journal of Computational Physics*, vol. 232, no. 1, pp. 288–317, 2013.

[86] T. Rainforth, R. Cornish, H. Yang, A. Warrington, and F. Wood, "On nesting monte carlo estimators," in *International Conference on Machine Learning*, PMLR, 2018, pp. 4267–4276.

[87] Q. Long, M. Scavino, R. Tempone, and S. Wang, "Fast estimation of expected information gains for bayesian experimental designs based on laplace approximations," *Computer Methods in Applied Mechanics and Engineering*, vol. 259, pp. 24–39, 2013.

[88] S. Mosbach, A. Braumann, P. L. Man, C. A. Kastner, G. P. Brownbridge, and M. Kraft, "Iterative improvement of bayesian parameter estimates for an engine model by means of experimental design," *Combustion and Flame*, vol. 159, no. 3, pp. 1303–1313, 2012.

[89] F. vom Lehn, L. Cai, and H. Pitsch, "Iterative model-based experimental design for efficient uncertainty minimization of chemical mechanisms," *Proceedings of the Combustion Institute*, vol. 38, no. 1, pp. 1033–1042, 2021.

[90] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Statistical Science*, pp. 273–304, 1995.

[91] D. F. Davidson and R. Hanson, "Recent advances in shock tube/laser diagnostic methods for improved chemical kinetics measurements," *Shock Waves*, vol. 19, pp. 271–283, 2009.

[92] D. A. Sheen and H. Wang, "Combustion kinetic modeling using multispecies time histories in shock-tube oxidation of heptane," *Combustion and Flame*, vol. 158, no. 4, pp. 645–656, 2011.

[93] T. Atherley, S. De Persis, N. Chaumeix, Y. Fernandes, A. Bry, A. Comandini, O. Mathieu, S. Alturaifi, C. Mulvihill, and E. Petersen, "Laminar flame speed and shock-tube multi-species laser absorption measurements of dimethyl carbonate oxidation and pyrolysis near 1 atm," *Proceedings of the Combustion Institute*, vol. 38, no. 1, pp. 977–985, 2021.

[94] S. J. Cassady, R. Choudhary, V. Boddapati, N. H. Pinkowski, D. F. Davidson, and R. K. Hanson, "The pyrolysis of propane," *International Journal of Chemical Kinetics*, vol. 52, no. 11, pp. 725–738, 2020.

[95] S. Banerjee, R. Tangko, D. A. Sheen, H. Wang, and C. T. Bowman, "An experimental and kinetic modeling study of n-dodecane pyrolysis and oxidation," *Combustion and Flame*, vol. 163, pp. 12–30, 2016.

[96] C. Olm, T. Varga, É. Valkó, H. J. Curran, and T. Turányi, "Uncertainty quantification of a newly optimized methanol and formaldehyde combustion mechanism," *Combustion and Flame*, vol. 186, pp. 45–64, 2017.

[97] T. Varga, I. G. Zsély, T. Turányi, T. Bentz, and M. Olzmann, "Kinetic analysis of ethyl iodide pyrolysis based on shock tube measurements," *International Journal of Chemical Kinetics*, vol. 46, no. 6, pp. 295–304, 2014.

[98] M. Khalil, K. Chowdhary, C. Safta, K. Sargsyan, and H. N. Najm, "Inference of reaction rate parameters based on summary statistics from experiments," *Proceedings of the Combustion Institute*, vol. 36, no. 1, pp. 699–708, 2017.

[99] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, and N. D. Goodman, "Pyro: Deep universal probabilistic programming," *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 973–978, 2019.

[100] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[101] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[102] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, no. 4, pp. 457–472, 1992.

[103] J. Bell, M. Day, J. Goodman, R. Grout, and M. Morzfeld, "A bayesian approach to calibrating hydrogen flame kinetics using many experiments and parameters," *Combustion and Flame*, vol. 205, pp. 305–315, 2019.

[104] D. G. Goodwin, R. L. Speth, H. K. Moffat, and 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.

[105] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 42, no. 1, pp. 55–61, 2000.

[106] K. Sargsyan, H. Najm, and R. Ghanem, "On the statistical calibration of physical models," *International Journal of Chemical Kinetics*, vol. 47, no. 4, pp. 246–276, 2015.

[107] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.

[108] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Problems*, vol. 33, no. 12, p. 124 007, 2017.

[109] S. Mosbach, J. H. Hong, G. P. Brownbridge, M. Kraft, S. Gudiyella, and K. Brezinsky, "Bayesian error propagation for a kinetic model of n-propylbenzene oxidation in a shock tube," *International Journal of Chemical Kinetics*, vol. 46, no. 7, pp. 389–404, 2014.

[110] P. C. Hansen and D. P. O'Leary, "The use of the l-curve in the regularization of discrete ill-posed problems," *SIAM journal on scientific computing*, vol. 14, no. 6, pp. 1487–1503, 1993.

[111] W. Ji, T. Yang, Z. Ren, and S. Deng, "Dependence of kinetic sensitivity direction in premixed flames," *Combustion and Flame*, vol. 220, pp. 16–22, 2020.

[112] X. Huan and Y. Marzouk, "Gradient-based stochastic optimization methods in bayesian experimental design," *International Journal for Uncertainty Quantification*, vol. 4, no. 6, 2014.

[113] A. Foster, M. Jankowiak, M. O'Meara, Y. W. Teh, and T. Rainforth, "A unified stochastic gradient approach to designing bayesian-optimal experiments," in *International Conference on Artificial Intelligence and Statistics*, PMLR, 2020, pp. 2959–2969.

[114] T. Goda, T. Hironaka, W. Kitade, and A. Foster, "Unbiased mlmc stochastic gradient-based optimization of bayesian experimental designs," *SIAM Journal on Scientific Computing*, vol. 44, no. 1, A286–A311, 2022.

[115] A. G. Carlon, B. M. Dia, L. Espath, R. H. Lopez, and R. Tempone, "Nesterov-aided stochastic gradient methods using laplace approximation for bayesian design optimization," *Computer Methods in Applied Mechanics and Engineering*, vol. 363, p. 112 909, 2020.

[116] J.-H. Oh, P. Wiersema, K. Kim, E. Mayhew, J. Temme, C.-B. Kweon, and T. Lee, "Fast uncertainty reduction of chemical kinetic models with complex spaces using hybrid response-surface networks," *Combustion and Flame*, vol. 253, p. 112 772, 2023.

[117] F. B. Smith, A. Kirsch, S. Farquhar, Y. Gal, A. Foster, and T. Rainforth, "Prediction-oriented bayesian active learning," in *International Conference on Artificial Intelligence and Statistics*, PMLR, 2023, pp. 7331–7348.

[118] G. P. Smith, Y. Tao, and H. Wang, *Foundational fuel chemistry model version 1.0 (FFCM-1)*, https://web.stanford.edu/group/haiwanglab/FFCM1/pages/FFCM1.html, 2016.

[119] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[120] X. Zhang, G. Wang, J. Zou, Y. Li, W. Li, T. Li, H. Jin, Z. Zhou, and Y.-Y. Lee, "Investigation on the oxidation chemistry of methanol in laminar premixed flames," *Combustion and Flame*, vol. 180, pp. 20–31, 2017.

[121] M. Belkin, D. Hsu, S. Ma, and S. Mandal, "Reconciling modern machine-learning practice and the classical bias–variance trade-off," *Proceedings of the National Academy of Sciences*, vol. 116, no. 32, pp. 15 849–15 854, 2019.

[122] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, JMLR Workshop and Conference Proceedings, 2011, pp. 315–323.

[123] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.