SAND2018-7921C

Enabling uncertainty quantification and propagation in combustion models using statistical inference of unreported experimental data

Tiernan A. Casey, Habib N. Najm

Combustion Research Facility, Sandia National Laboratories, Livermore, CA, USA

TNF Workshop 2018

Email: tcasey@sandia.gov

Background

- Prescribing uncertainty measures to rate expressions is crucial for performing useful predictive combustion simulations
- Experimental measurement data are often unreported, information is reported in the form of model parameters fit to the data (e.g. rate constants, Arrhenius parameters) with error bars. Loss of correlation information between Arrhenius parameters
- Approach: use maximum entropy arguments to construct statistical inference on the unreported data itself using available information as constraints. Estimate parameters (with joint uncertainty) using this inferred data [1, 2]

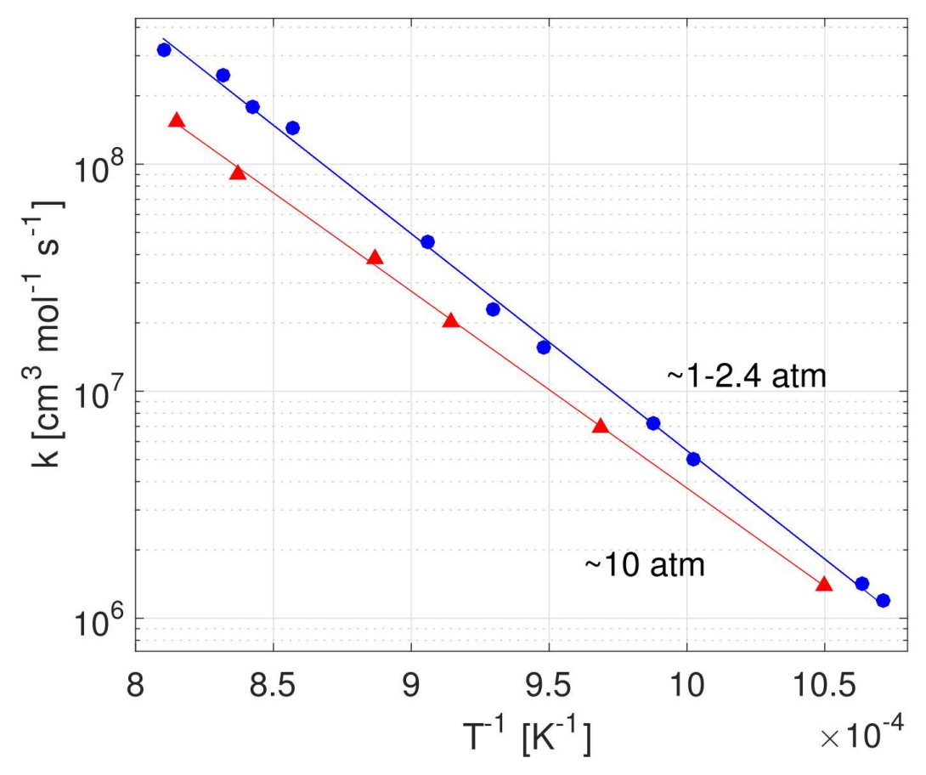
Experiments

Shock tube experiments of Sajid et al [3]:

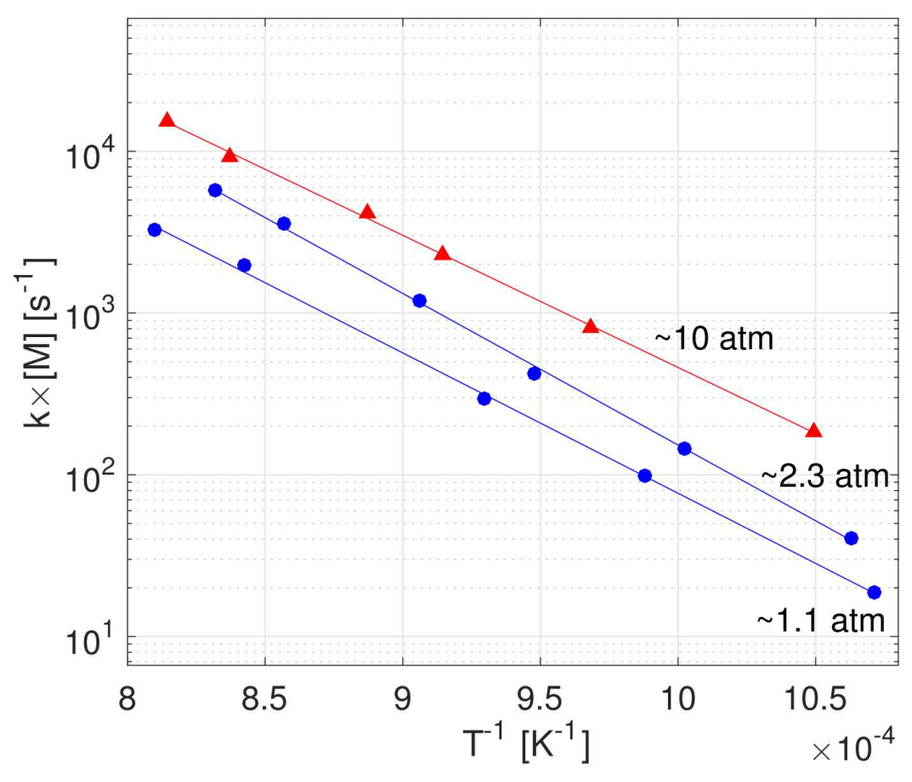
- Investigate thermal decomposition reaction:

$$H_2O_2 + M \longrightarrow 2\,OH + M \tag{1}$$

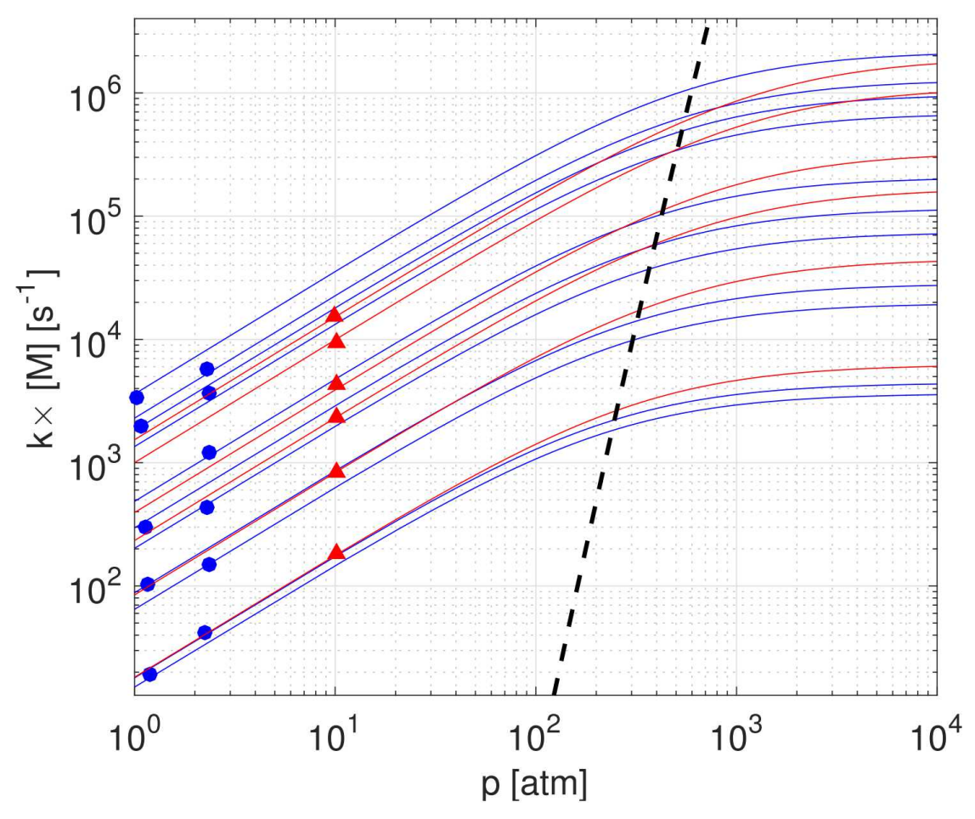
- Measured signals are noisy laser absorption time series, transformed to noisy H_2O_2 mole fraction vs. time profiles using the Beer-Lambert relation
- Shock tube conditions span temperatures of 930-1235 K, 1-2 atm
- Rate constant of target reaction (eq. 1) determined using homogeneous reactor simulations using a calibration chemical mechanism from Hong et al. [4]
- Raw data unreported, rate constants at design temperatures are, with attached error measures (symmetric error bars)



second order rate constants



first order rate constants



fall-off curves

Fall-off curves estimated using Lindemann theory with the high-pressure limit specified using the parameterization from the calibration mechanism used by the experimentalists to fit the unreported data

Bayesian inference

- Assuming an additive Gaussian noise model for the unreported noisy mole fraction vs. time data:

$$x\left(t_i, T_j\right) = f\left(t_i, T_j; \beta\right) + \sigma\varepsilon_{i,j} \tag{2}$$

where f is the output of the ODE system modeling the smooth H_2O_2 signal. The data points at all time (t) and temperature (T) locations define a data set $\mathbf{D}$. β are uncertain model parameters (rate constants) and σ is the standard deviation of an additive noise model. The posterior distribution of $\mathbf{D}$ given reported statistics S of the unreported data can be inverted:

Bayes' law:
$$p(\mathbf{D}|S) = \frac{p(S|\mathbf{D})\pi(\mathbf{D})}{p(S)} \tag{3}$$

- Consistency between proposed instances of $\mathbf{D}$ and S is enforced at the M temperatures using an approximate likelihood function that penalizes discrepancy using a Gaussian penalty kernel:

$$p(S|\mathbf{D}) \propto \prod_{j=1}^{M} \exp\left[-\delta_j \left(\frac{S_{j,\mathbf{D}} - S_j^*}{S_j^*}\right)^2\right] \tag{4}$$

- To facilitate evaluation of eq. 4 the statistics of proposed data $S_{\mathbf{D}}$ are evaluated from an associated parameter posterior PDF, which is itself evaluated using a Bayesian inference on parameters $\lambda = \{\beta, \sigma\}$:

$$p(\lambda|\mathbf{D}) = \frac{p(\mathbf{D}|\lambda)\pi(\lambda)}{p(\mathbf{D})} \tag{5}$$

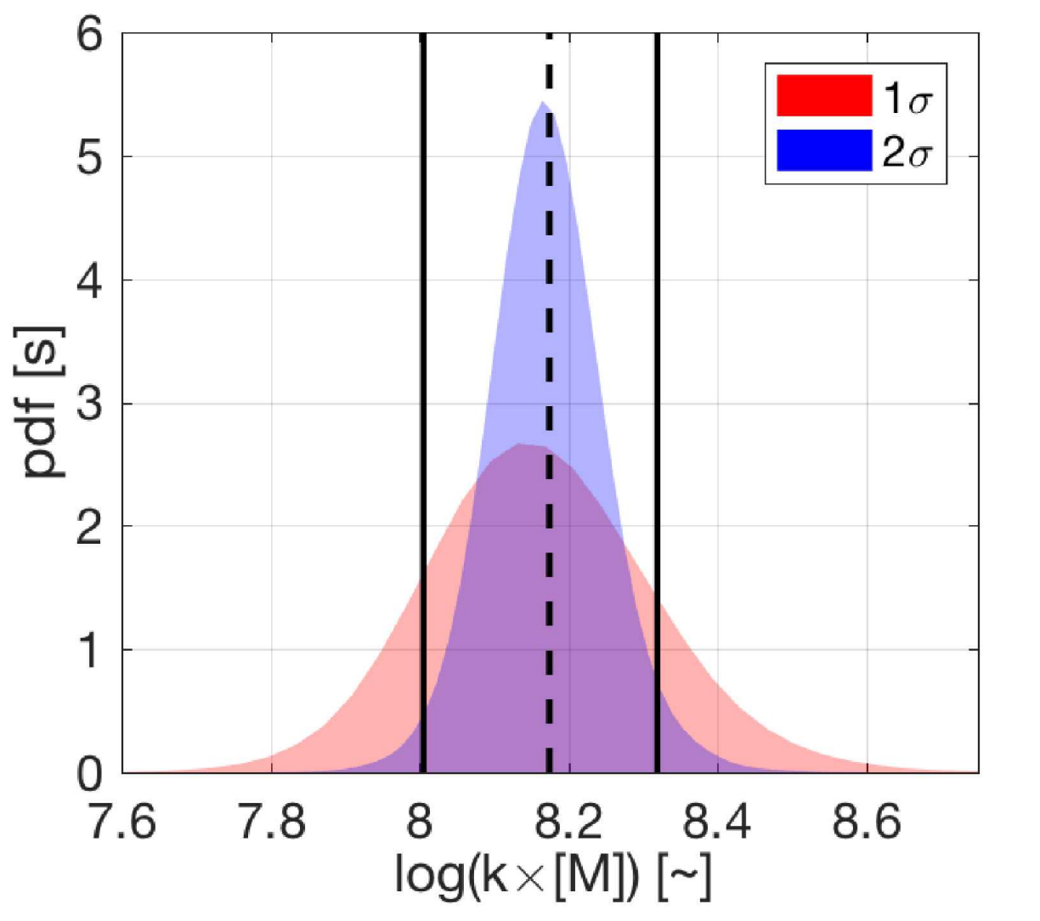
using a likelihood function implied by the form of eq. 2:

$$p(\mathbf{D}|\lambda) = \prod_{i=1}^{N}\prod_{j=1}^{M} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left[x\left(t_i, T_j\right) - f\left(t_i, T_j; \beta\right)\right]^2}{2\sigma^2}\right) \tag{6}$$

- As such the data inference algorithm explores a space of consistent data by approximately enforcing agreement between the statistics of hypothetical data with those of the the unreported true data using a nested Bayesian inference procedure employing an outer inference on data with associated inner inferences on parameters

Consistent data

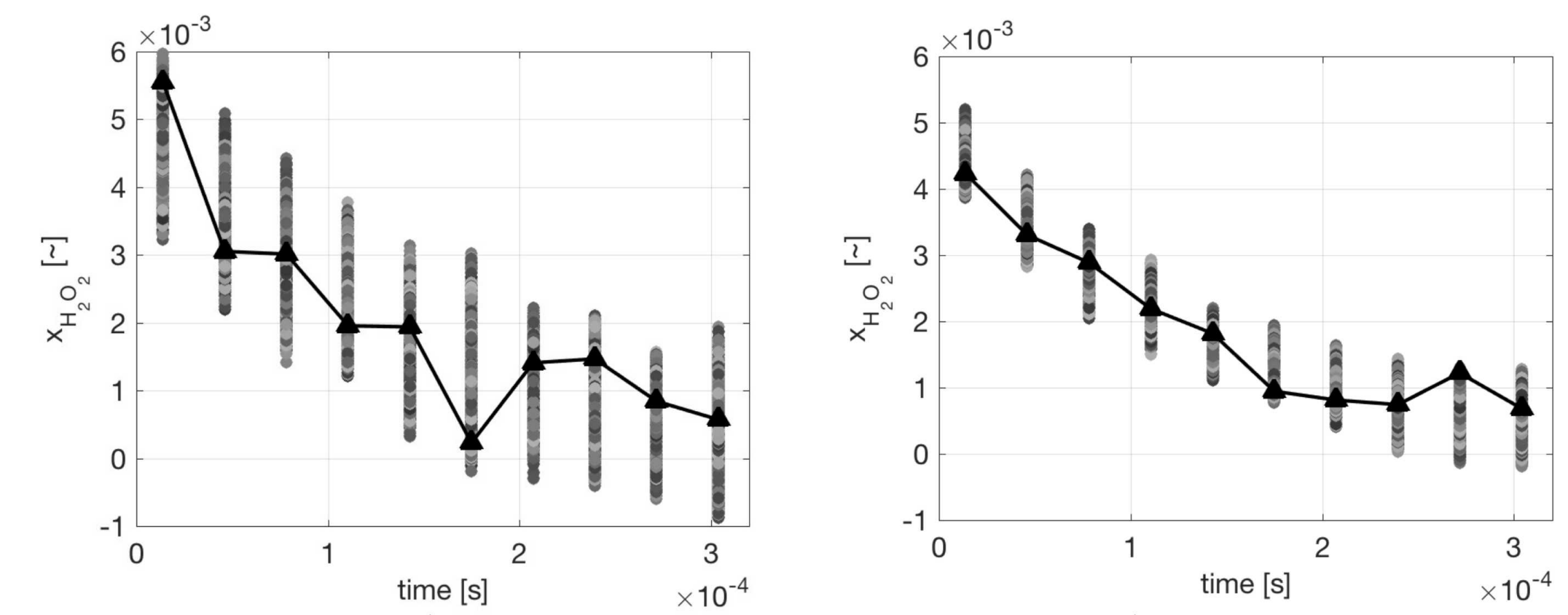
- The numerical solution of the parameter inference using Markov chain Monte Carlo (MCMC) methods delivers samples from the target parameter posterior PDF:



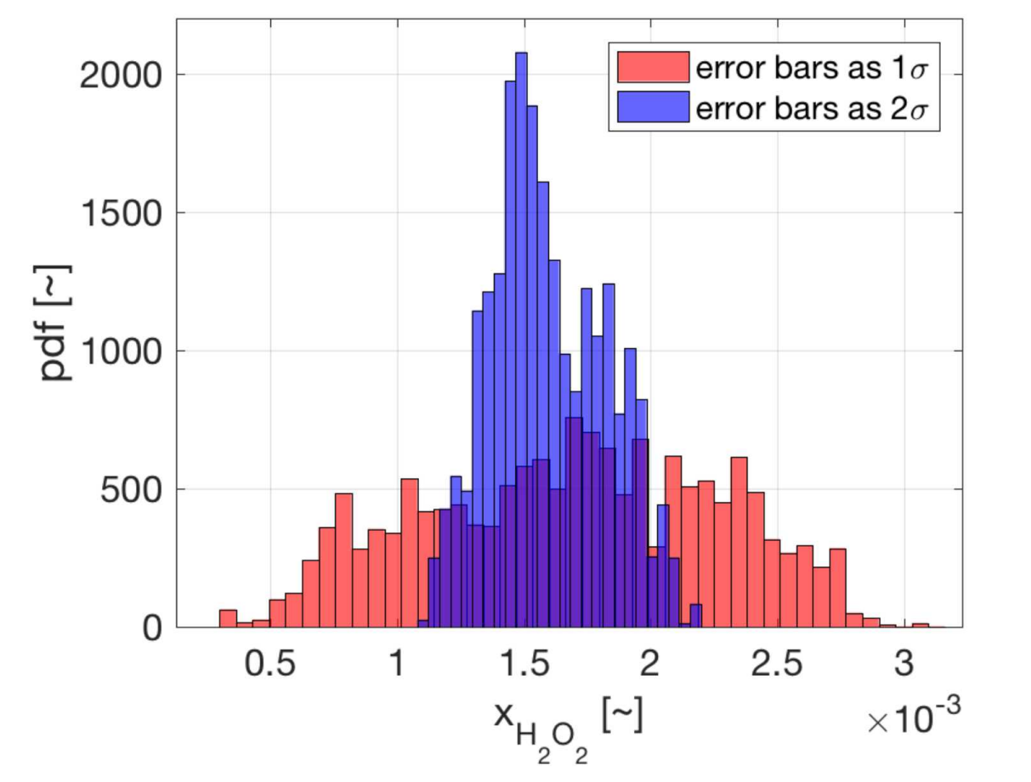
rate constant PDFs

  - comparison of inferred parameter distribution from a hypothetical data set and the reported error bars (constraints)
  - interpreting error bars as 1 or 2 std measures of uncertainty on rate constants implies higher or lower noise magnitude in the data respectively

- Similarly, inference on data delivers samples from the posterior PDF on data conditioned on the reported statistics (typically 1000s of data instances are sampled):



Data samples (i.e. hypothetical data sets) using a 1 std interpretation of rate constant error bars (left) and 2 std interpretation (right). Single examples of data sets are highlighted with solid lines

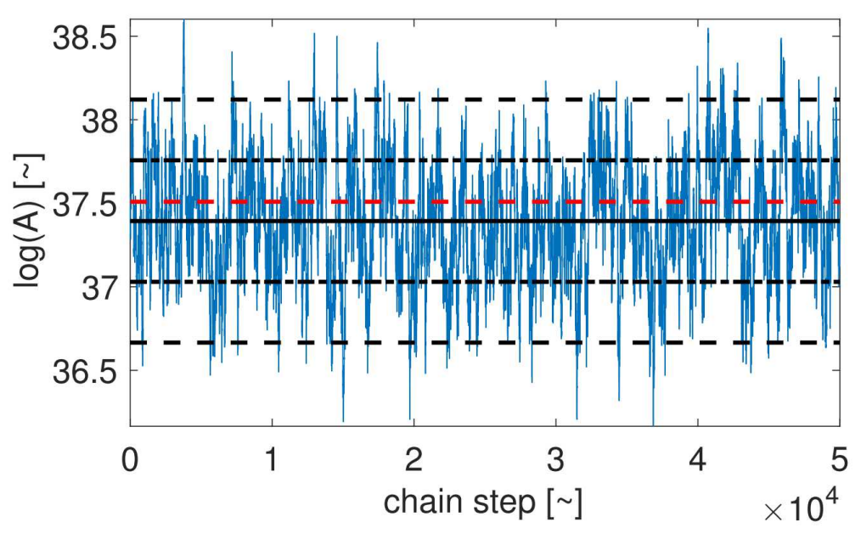


data PDF at a single time point (i.e. marginal distribution)

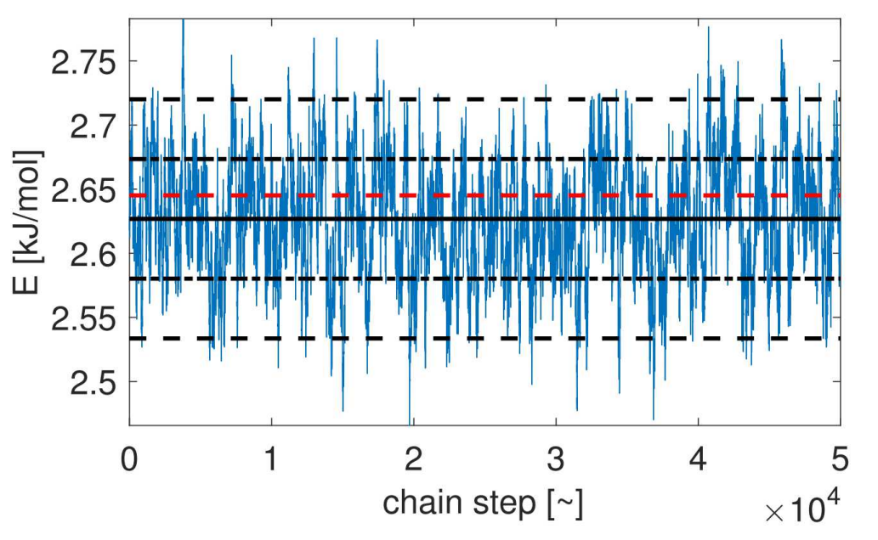
Distribution of noisy data at a single time point in the H_2O_2 decay, highlighting the increase in noise strength inferred when interpreting the error bars at the 1 std level

MaxEnt posterior PDFs

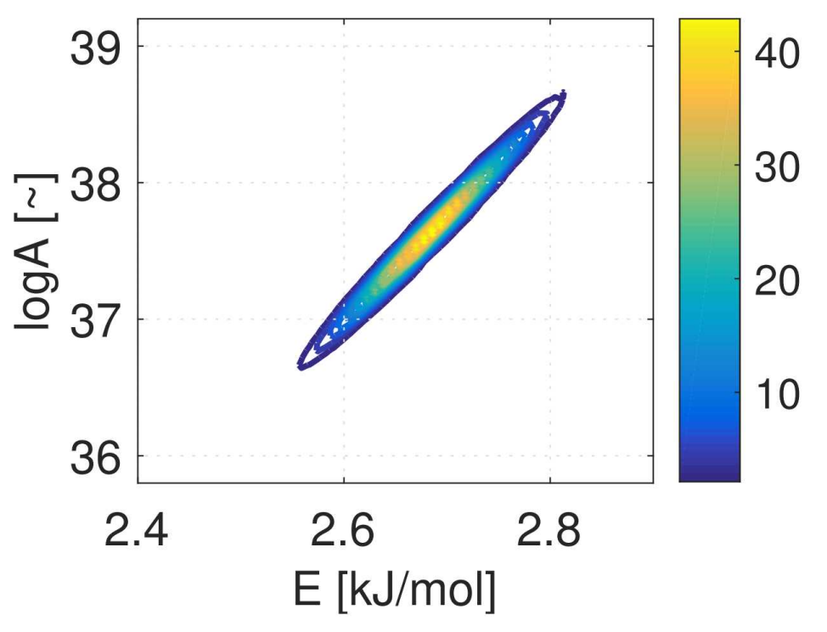
- Can learn joint PDFs of Arrhenius parameters from discovered hypothetical data employing Bayesian inference using data sets at all temperatures, constructing PDFs averaged over all data instances, complete with the previously missing correlation information



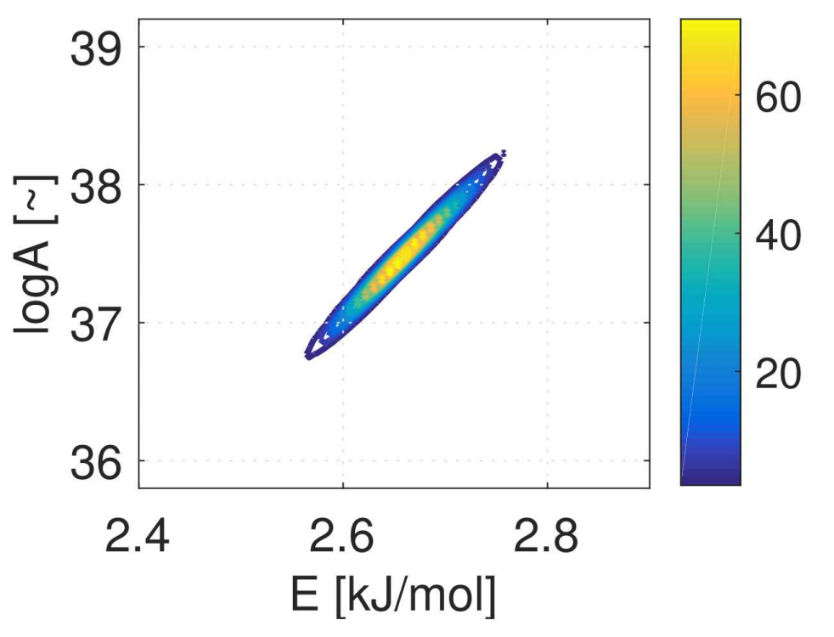
MCMC chain for pre-exponential factor (log(A))



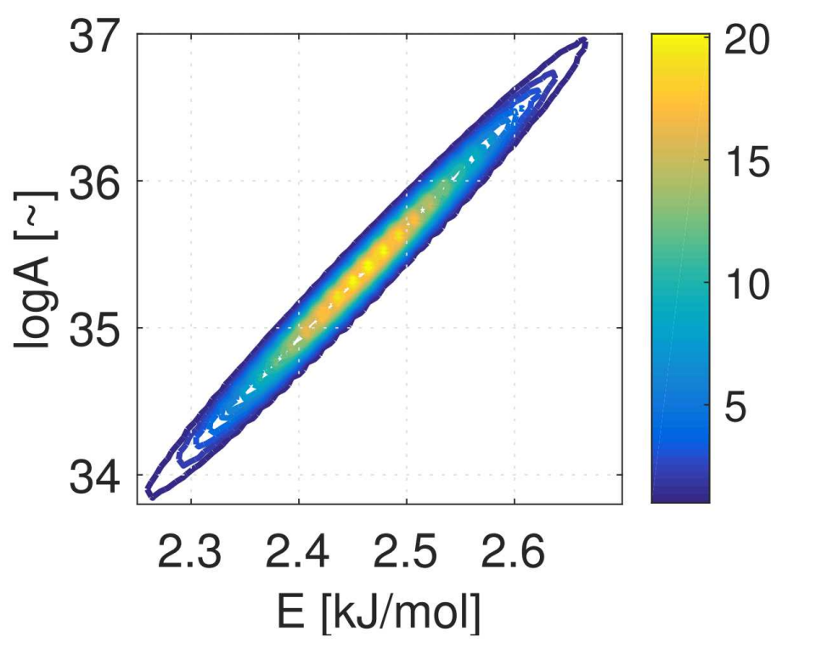
MCMC chain for activation energy (E)



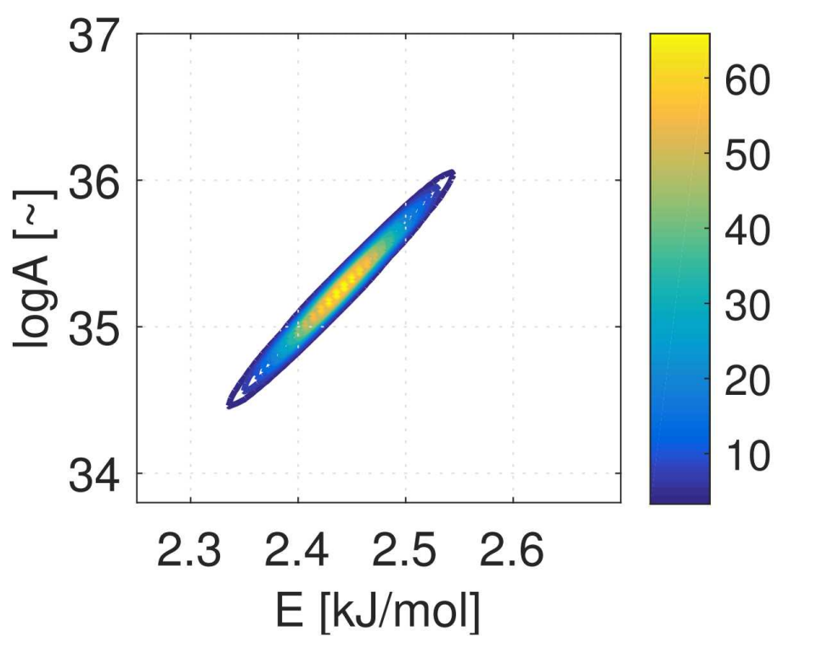
low pressure experiments (error bars at 1 std)



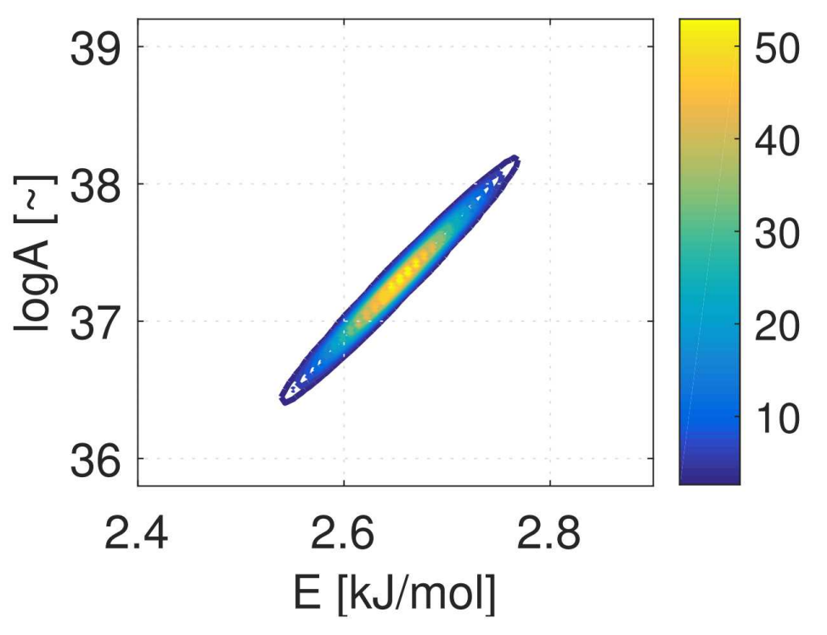
low pressure experiments (error bars at 2 std)



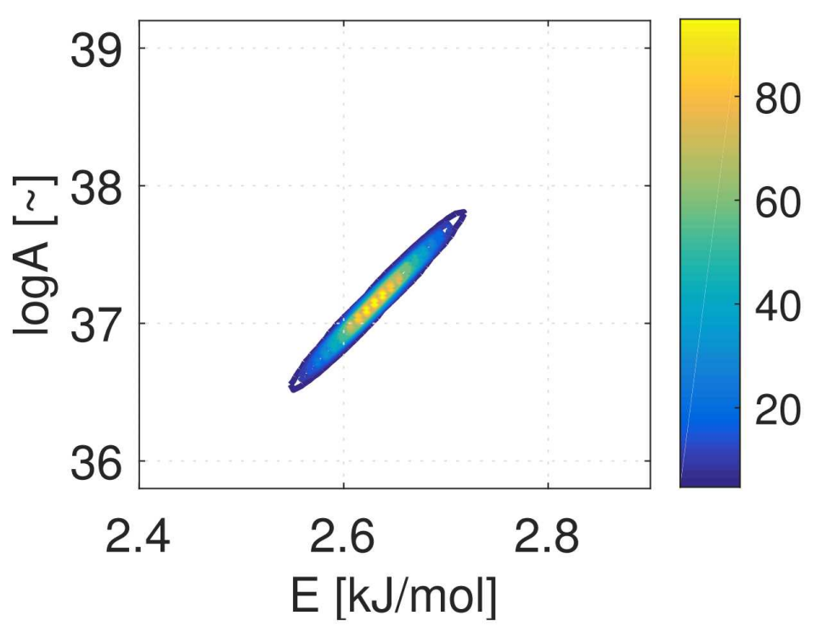
high pressure experiments (error bars at 1 std)



high pressure experiments (error bars at 2 std)



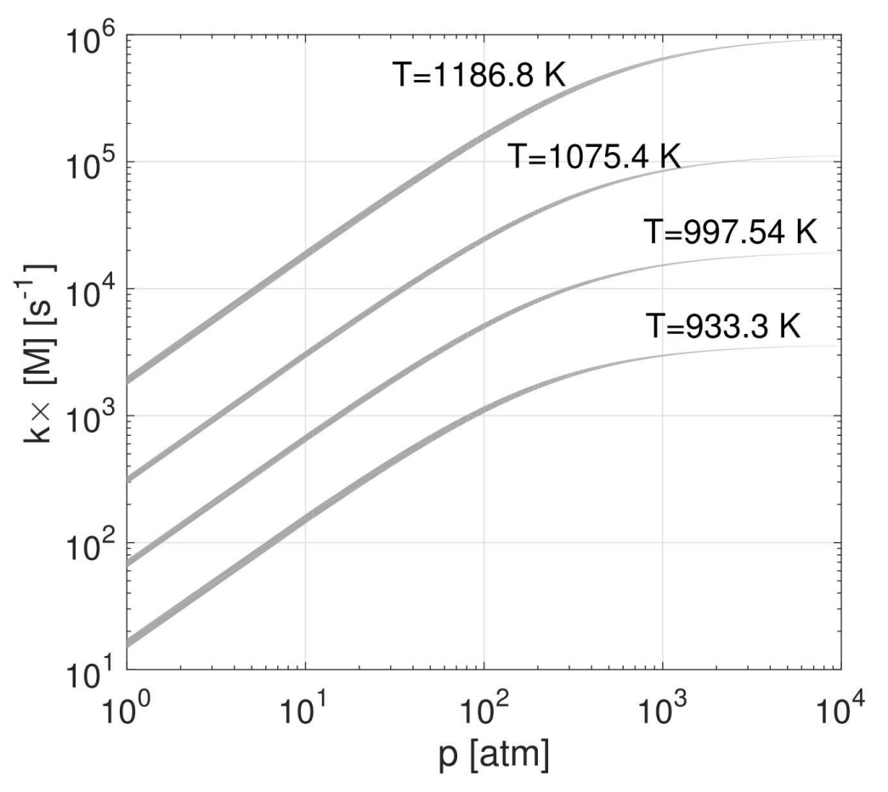
all experiments (error bars at 1 std)



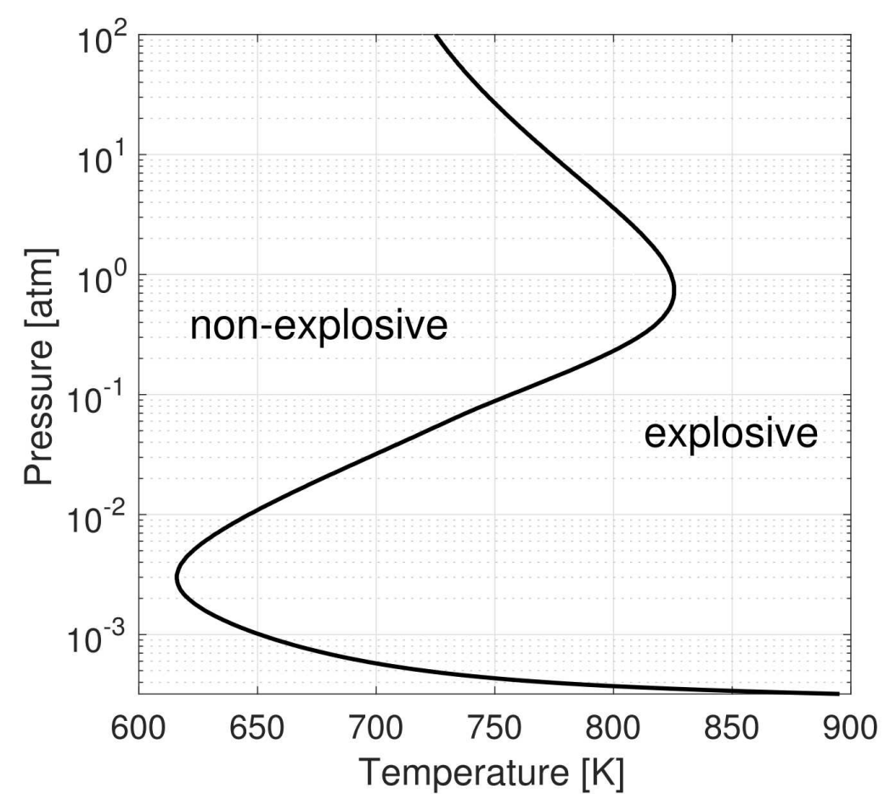
all experiments (error bars at 2 std)

Application - explosion limit curve

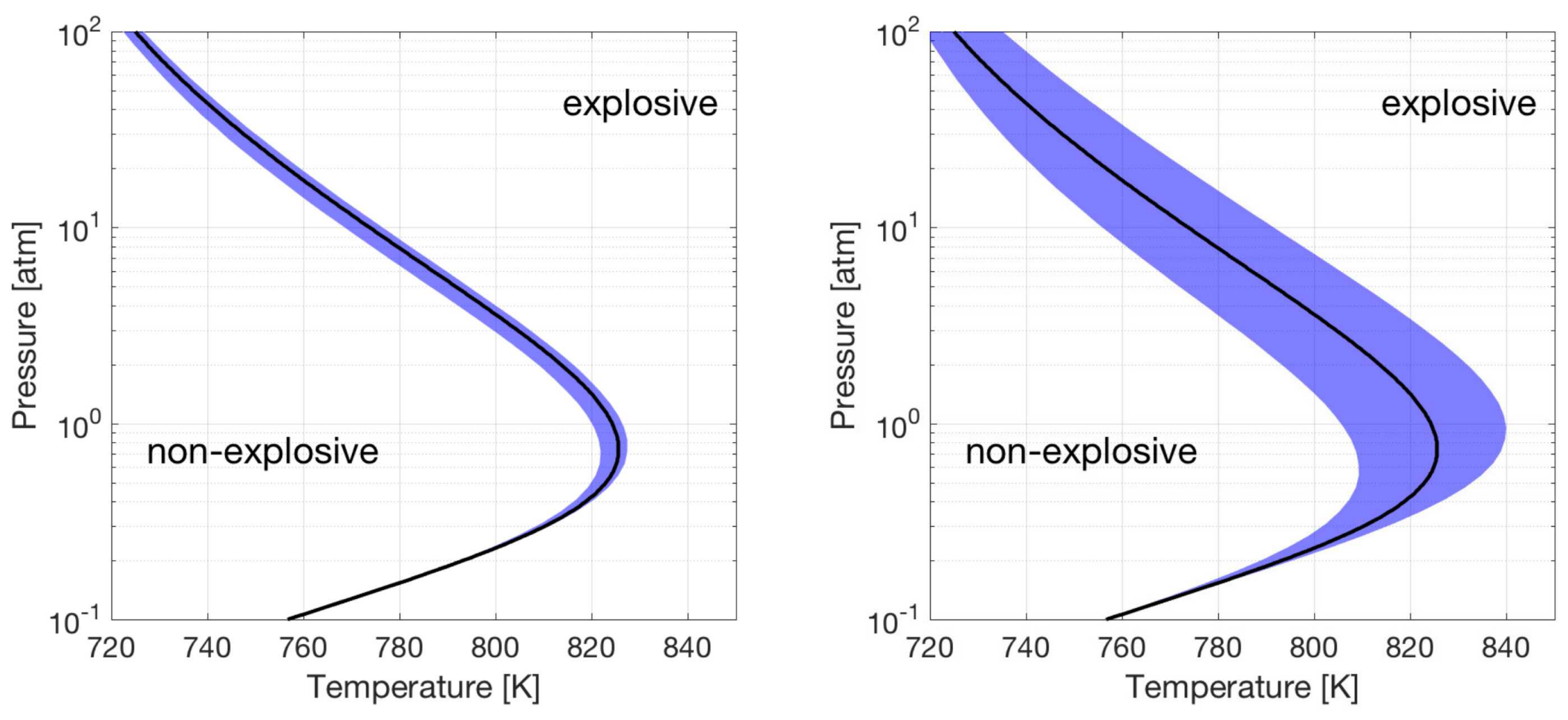
- Propagate uncertainty in a system of interest sensitive to H_2O_2 decomposition, e.g. the 3^{rd} explosion limit in H_2-O_2 mixtures, drawing samples from the joint Arrhenius parameter PDF



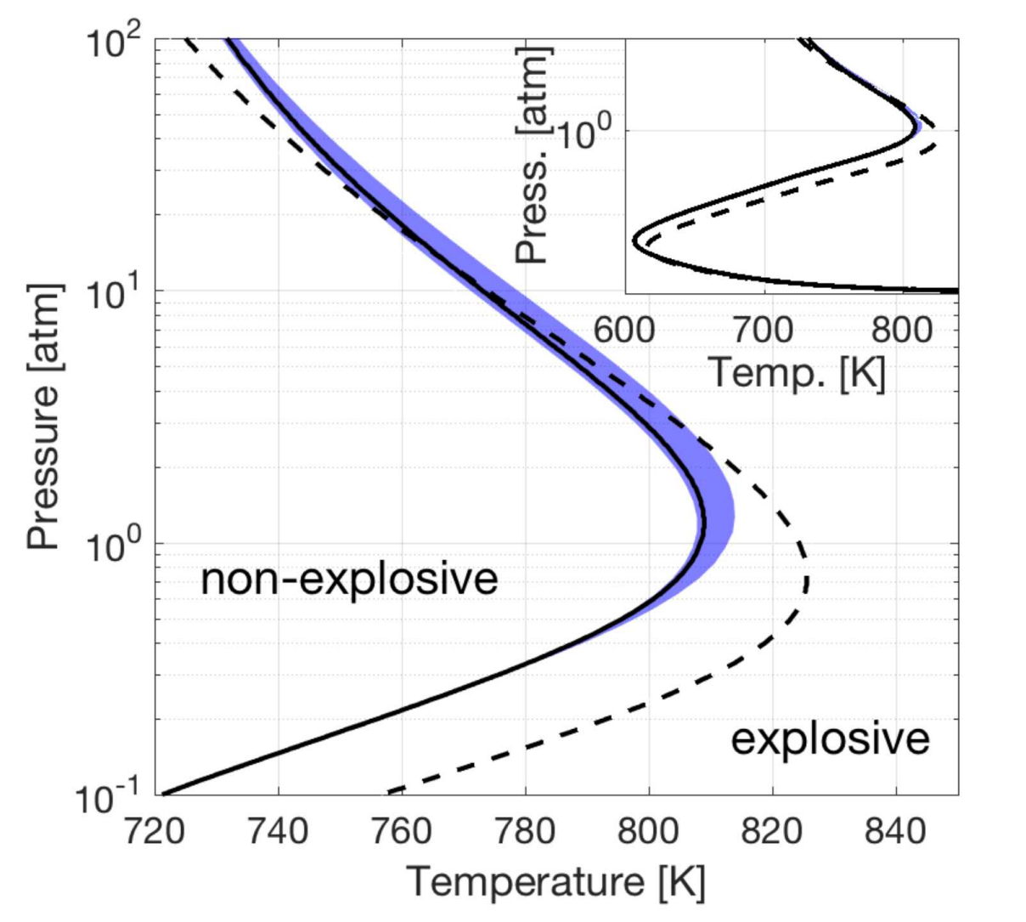
95% confidence interval of fall-off curves for selected temperatures



stoichiometric H_2-O_2 explosion limit curve using the nominal mechanism of Hong et al. [4]



Explosion limit Z-curve with 95% confidence interval near the third limit for stoichiometric H_2-O_2 mixtures sampling from the correlated joint PDF computed using data inference (left) compared to drawing samples from an equivalent uncorrelated PDF (right), highlighting the over-prediction of uncertainty when knowledge of the parameter correlation is lost. The solid black curves are the Z-curve computed using the nominal rate parameters determined by Sajid et al [3] using low pressure information only



- with access to a representation of the unreported data, can refit the data using any chemical mechanism of our choice
- e.g. refit using a consensus mechanism, designed for predictive combustion simulations (e.g. DNS, LES)

Comparison of predictions when fitting hypothetical data using a different chemical mechanism [5] and using said mechanism for predictive modeling. The dashed line is the nominal explosion Z-curve using the Hong et al. mechanism both to fit the Arrhenius parameters of reaction 1 to the data and to predict the Z-curve, indicating significant discrepancy between the two mechanisms.

Conclusions and future work

- Demonstrate the application of a data inference procedure to generate missing experimental data from combustion experiments for the purposes of refitting the data in a Bayesian framework to recover missing parameter correlation information
- Approach presents an objective framework for the assimilation of evidence from different sources to constrain the uncertain representation of a reaction of interest in the context of any choice of chemical fitting model
- Extend to all reactions for a given fuel to construct unbiased mechanisms incorporating both available and unreported data by combining information at the data level

References

[1] Estimating the joint distribution of rate parameters across multiple reactions in the absence of experimental data. *Proceedings of the Combustion Institute*, 37, 2018.

[2] 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*, 36(1):699 – 708, 2017.

[3] Muhammad Bilal Sajid, Et Es-sebbar, Tamour Javed, Christa Fittschen, and Aamir Farooq. Measurement of the rate of hydrogen peroxide thermal decomposition in a shock tube using quantum cascade laser absorption near 7.7 μm. *International Journal of Chemical Kinetics*, 46(5):275–284, 2014.

[4] Zekai Hong, David F Davidson, and Ronald K Hanson. An improved H_2/O_2 mechanism based on recent shock tube/laser absorption measurements. *Combustion and Flame*, 158(4):633–644, 2011.

[5] Juan Li, Zhenwei Zhao, Andrei Kazakov, and Frederick L Dryer. An updated comprehensive kinetic model of hydrogen combustion. *International journal of chemical kinetics*, 36(10):566–575, 2004.

This work was supported by the US Department of Energy (DOE), Office of Basic Energy Sciences (BES) Division of Chemical Sciences, Geosciences, and Biosciences. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.