Learning-based Reconstruction of Chemical Reaction Networks

Juechun Tang[†] Hongtao Zhong[‡]
[†]Dept of Chemical and Biological Engineering [‡]Dept of Mechanical and Aerospace Engineering

Introduction

Dynamics in chemical reaction networks (CRNs) are of central importance for understanding various phenomena including explosion, catalysis, and biological systems.

CRN Formulation

$$\dot{\mathbf{Y}}_\mathbf{i}(t) = \frac{d\mathbf{Y}_\mathbf{i}}{dt} = \sum_j \mathbf{w}_{\mathbf{i},\mathbf{j}} = \mathbf{f}(\mathbf{Y}_\mathbf{i}, \mathbf{k}_\mathbf{j}) \qquad i \in \mathbf{S}, \quad j \in \mathbf{R}$$

$\mathbf{Y}$: species concentrations $\mathbf{S}$: a set of chemical species
k_j: reaction rate constant $\mathbf{R}$: a set of reactions

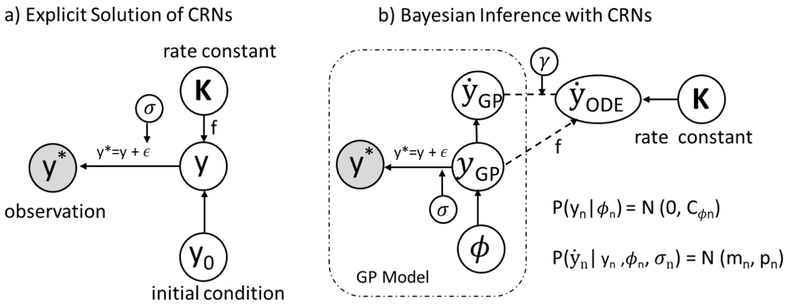
CRN Reconstruction: $\mathbf{k} = \mathscr{F}(\mathbf{S}, \mathbf{R}, \mathbf{Y}^*_\mathbf{i}(\mathbf{t}^*))$
Given experimental species temporal profiles as $\mathbf{Y}^*_\mathbf{i}(\mathbf{t}^*)$, can we determine the rate constants $\mathbf{k}$?

Limitations of Current Methods: manual fitting; complex numerical integration; much prior knowledge.
Why Machine Learning?: More efficient than numerical integration; May provide a distribution of inferred rate constants.

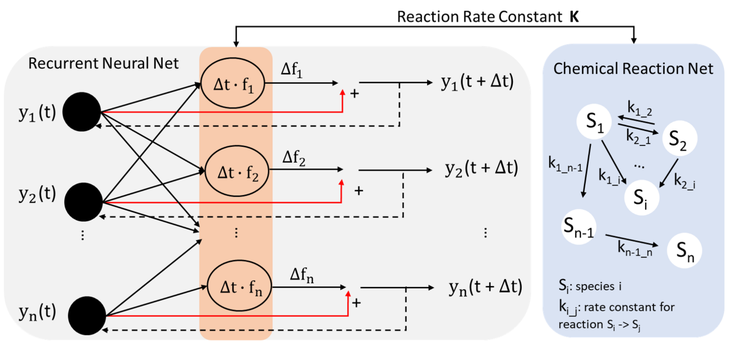
Our Goal: Apply machine learning techniques to reconstruct chemical reaction networks from discrete experimental data, facilitating research in complex chemical kinetics.

Methods and Approach

Adaptive Gradient Matching with Gaussian Process [2]



Recurrent Neural Net [3]

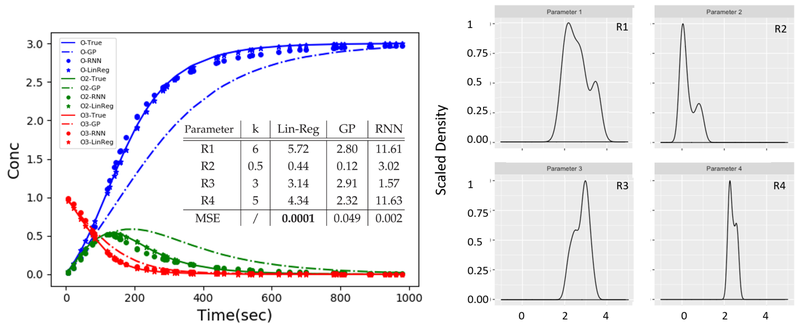


Case 1: Atmospheric O_2-O_3-O Kinetics

O_2 - O_3 - O kinetics is an indispensable component of atmospheric chemistry [1]. The reduced mechanism is written as

$O_2 \xrightarrow{h\nu} O\cdot + O\cdot$ (R1) $O\cdot + O_2 \longrightarrow O_3$ (R2)
$O_3 \xrightarrow{h\nu} O_2 + O\cdot$ (R3) $O_3 + O\cdot \longrightarrow O_2 + O_2$ (R4)

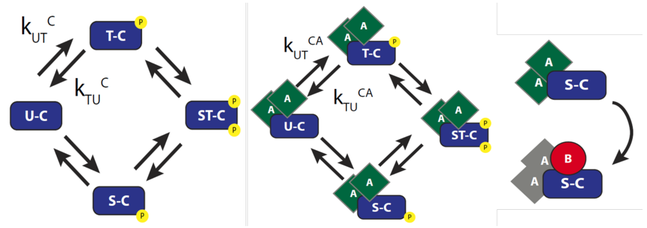
Model Performance and Rate Constant Distribution:



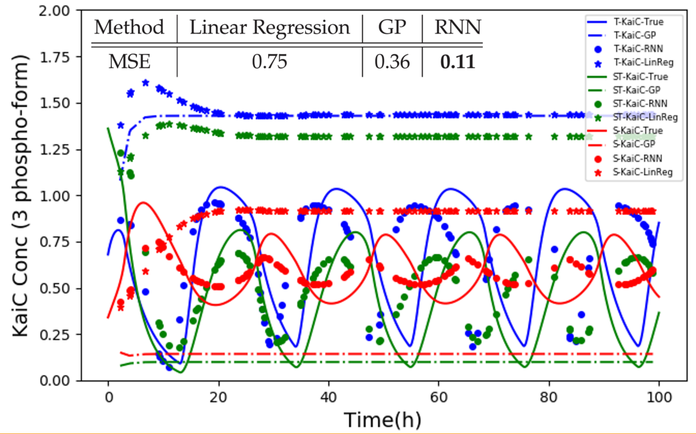
Parameter	k	Lin-Reg	GP	RNN
R1	6	5.72	2.80	11.61
R2	0.5	0.44	0.12	3.02
R3	3	3.14	2.91	1.57
R4	5	4.34	2.32	11.63
MSE	/	**0.0001**	0.049	0.002

Case 2: Biochemical Kinetic Oscillation

We demonstrate the parameter learning process with a real 22h-period circadian oscillator found in cyanobacteria [4].



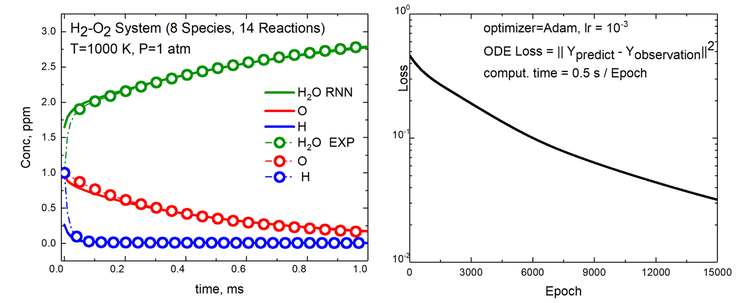
Model Performance with Inferred Rate Constants :



Method	Linear Regression	GP	RNN
MSE	0.75	0.36	**0.11**

Case 3: Towards Large-Scale Kinetics

Chemical reaction networks are intrinsically large-scale. Here we present RNN reconstruction of a H_2−O_2 combustion mechanism with 14 reactions.



Model Performance with Inferred Rate Constants:

	Reactions	k	Initial ϵ	RNN ϵ
1	$H_2 + OH \longrightarrow H_2O + H$	1.3×10^{-12}	0.39	0.02
2	$O + H_2O \longrightarrow OH + OH$	1.0×10^{-13}	0.30	0.36
3	$HO_2 + OH \longrightarrow H_2O + O_2$	4.0×10^{-11}	1.90	1.91
...	...	...	..	...
14	$H_2O_2 + H \longrightarrow HO_2 + H_2$	1.0×10^{-12}	0.32	0.02

k (cm^3/molecule s) retrived from NIST chemical kinetics database. ϵ is the deviation from the true rate constants defined as $\|(k - k_{true})/k_{true}\|$

Summary and Conclusion

Conclusion

- RNN best reserves the system dynamics from small to large-scale systems
- Lin-Reg (with poly-terms) performs best in simple kinetics but fails to capture nonlinear dynamics in real systems

Discussion and Extensions

- This study is limited to a fully observed system. CRN reconstruction from a partial observed system would be more applicable and interesting
- Rate constants are assumed to be constant in this work. In reality, there is a strong nonlinear dependence between rate constants and temperature, pressure, etc.
- Large stiffness is a distinctive feature of CRNs. Rate constants may vary up to several orders of magnitudes. Model reduction is necessary before any learning-based reconstruction.

References

[1] L. Barrie, J. Bottenheim, R. Schnell, P. Crutzen, and R. Rasmussen. Ozone destruction and photochemical reactions at polar sunrise in the lower arctic atmosphere.Nature, 334(6178):138, 1988. [2] B. Calderhead, M. Girolami, and N. D. Lawrence. Accelerating bayesian inference over nonlinear differential equations with gaussian processes. InAdvances in neural information processing systems, pages 217–224, 2009. [3] H. Ling, S. Samarasinghe, and D. Kulasiri. Novel recurrent neural network for modelling biological networks: oscillatory p53 interaction dynamics.Biosystems, 114(3):191–205, 2013. [4] M. Rust, J. Markson, W. Lane, D. Fisher, and E. O'shea. Ordered phosphorylation governs oscillation of a three-protein circadian clock.Science, 318(5851):809–812, 2007.