Multi-phase chemistry surrogate modeling with a recurrent neural network

Atmospheric Chemical Mechanisms Conference 2022
Mechanism Development and Reduction

Xiaokai Yang (University of Illinois Urbana-Champaign)
Christopher Tessum, Nicole Riemer, Zhonghua Zheng, Guo Lin

Dec 2022
Atmospheric chemical mechanism

- **Gas-phase chemical mechanism**
  - Explicit mechanism
  - Lumped mechanism

  \[
  \text{ethylbenzene} = 1 \text{ TOL} + 1 \text{ PAR}
  \]

- **Multi-phase chemical mechanism**
  - Gas-particle partitioning

Box Model: PartMC-MOSAIC

- **PartMC**: Particle-resolved Monte Carlo aerosol model
- **MOSAIC**: Model for Simulating Aerosol Interactions and Chemistry
Update PartMC-MOSAIC Gas-phase Chemistry

- **MCM**: Master Chemical Mechanism
  - Near-explicit chemical mechanism
Update PartMC-MOSAIC SOA Chemistry

➢ **SORGAM**: Secondary Organic Aerosol Model
  - Organic gas/particle-partitioning thermodynamic equilibrium
  - Mapping:

<table>
<thead>
<tr>
<th>MCM Species</th>
<th>SORGAM Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>C198PAN</td>
<td>ALK1</td>
</tr>
<tr>
<td>NC102OOH</td>
<td>ARO1</td>
</tr>
<tr>
<td>C813OH</td>
<td>LIM1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ALK1</th>
<th>OLE1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARO1</td>
<td>ARO2</td>
</tr>
<tr>
<td>LIM1</td>
<td>LIM2</td>
</tr>
<tr>
<td>API1</td>
<td>API2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MCM Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>PXYLCO2H</td>
</tr>
<tr>
<td>MXYLCO2H</td>
</tr>
<tr>
<td>HOEMPHNO2</td>
</tr>
</tbody>
</table>

- Applied SIMPOL.1 to predict vapor pressure for species lacking in data
Coupled Box Model: MCM/PartMC-MOSAIC

Multi-phase Chemistry

- Inorganic Aerosol
  - MOSAIC
- Organic Aerosol
  - SORGAM

Gas-phase Chemistry

- MCM

Coupled Box Model

MCM/PartMC-MOSAIC
Machine-learned Surrogate Model

Reduce computational intensity by:

- **Unstiffening**: reduce stiffness of ODE system
- **Lumping**: reduce variables in the dynamic system
Training data: “teach” the model to make proper decisions

Machine learning algorithm: neural network, random forest etc.

Inductive biases: assumptions and constraints
Autoencoder-Decoder

➢ Compress chemical species (lumping)

Machine Learning Algorithm

**Neural ODE**: Neural Ordinary Differential Equation

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural network</td>
<td>$h_{t+1} = f(h_t, \theta_t)$</td>
</tr>
<tr>
<td>Recurrent network</td>
<td>$h_{t+1} = h_t + f(h_t, \theta_t)$</td>
</tr>
<tr>
<td><strong>Neural ODE</strong></td>
<td>$dh_t/dt = f(h_t, t, \theta_t)$</td>
</tr>
</tbody>
</table>

$h_t$ Hidden state at $t^{th}$ layer

$f$ Neural network

$\theta$ Parameter

Preliminary Training and Testing
Scaling up

<table>
<thead>
<tr>
<th>Reference</th>
<th>MCM/PartMC-MOSAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size/dimension</td>
<td>6,146 × 25</td>
</tr>
<tr>
<td>Scenario</td>
<td>3,000</td>
</tr>
</tbody>
</table>
Scaling up

<table>
<thead>
<tr>
<th>GPU 0</th>
<th>GPU 1</th>
<th>GPU 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Mini-batch</td>
<td>Gradient 0</td>
</tr>
<tr>
<td>Model</td>
<td>Mini-batch</td>
<td>Gradient 1</td>
</tr>
<tr>
<td>Model</td>
<td>Mini-batch</td>
<td>Gradient 2</td>
</tr>
</tbody>
</table>

**Dataset**
- Subset 0
- Subset 1
- Subset 2

**Training**
- Mean
- Pct90
- Pct95.0
- Pct99.0
- Pct100

**Testing**
- Mean
- Pct90
- Pct95.0
- Pct99.0
- Pct100

**Reference**
- MCM/PartMC-MOSAIC

**Size/dimension**
- 6,146 × 25

**Scenario**
- 3,000
Machine Learning Algorithm

➢ **ESN**: Echo state network

\[ x_{t+1} = (1 - \alpha)x_t + \alpha \cdot \tanh(W_x x_t + W_{in} u_t) \]

\[ u_{t+1} = W_{out} x_{t+1} \]

- \( u \): Input data
- \( x_t \): Hidden state
- \( \alpha \): Leaky coefficient
Preliminary Training and Testing

Latent species No.3

Time (hr)

0 50 100 150

Reference
Prediction

Size/dimension
6,146 × 10,080

Scenario
1
Machine Learning Algorithm

➢ **SINDy**: Sparse identification of nonlinear dynamics

Summary

- **Coupled box model: MCM/PartMC-MOSAIC**
  - A mass conserved multi-phase chemical mechanism
  - A detailed reference model for emulation

- **Machine learned surrogate model**
  - AutoEncoder: compress/lump the chemical system
  - Neural ODE: promising for small dataset but is computationally expensive
  - Echo state network: less computational expense but less generalization
  - SINDy: expected to be much faster and more generalizable
References


Thank you!