Advancing Atmospheric Chemistry Workflows in Python

https://linktr.ee/musica_documentation

Angela Pak, Aidan Winney
Mentors: Jian Sun, Kyle Shores, Matthew Dawson
Summer Internships in Parallel Computational Science (SIParCS)

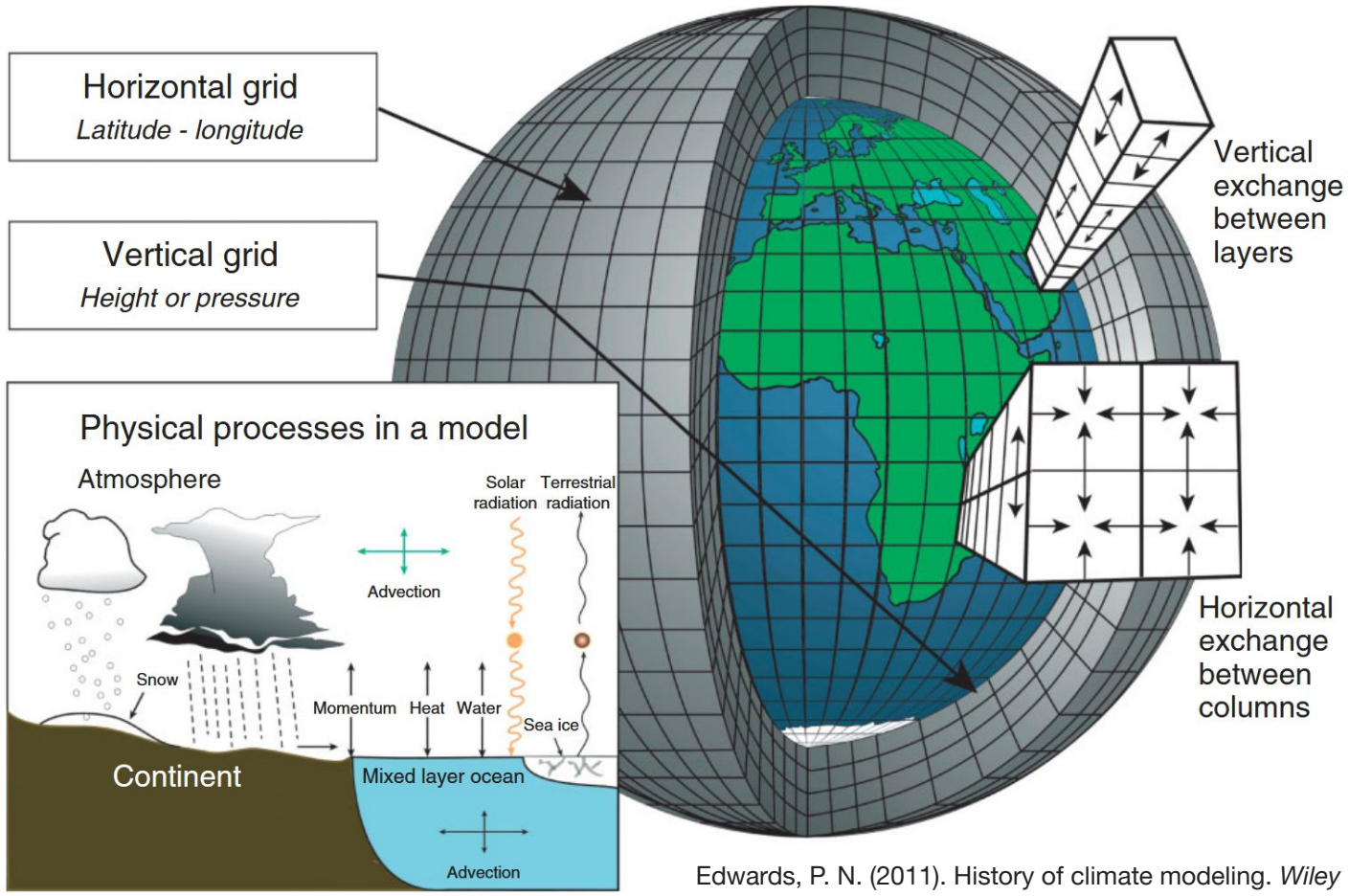
July 29, 2025

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

MUSICA Simulation Overview

MUSICA allows users to simulate chemical reactions in the atmosphere

- These simulations assume that the **air mass is isolated** and **well-mixed**.
- There are two types of grid cells in atmospheric models:
  - Horizontal grid cells
  - Vertical grid cells

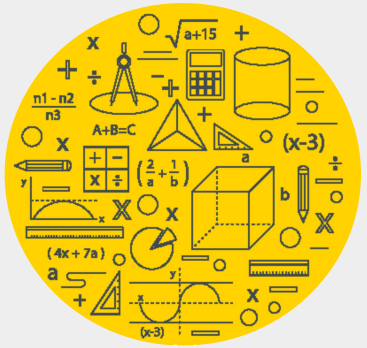


Edwards, P. N. (2011). History of climate modeling. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(1), 128-139.

Intro to MUSICA System

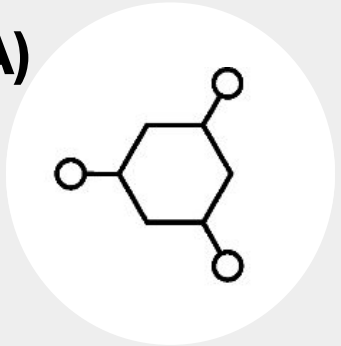
Model-Independent Chemistry Module (MICM)

- Chemistry solver
- Ordinary Differential Equation (ODE) integration



Multi-scale Infrastructure for Chemistry and Aerosols (MUSICA)

- Chemistry set up
- Mechanism configuration

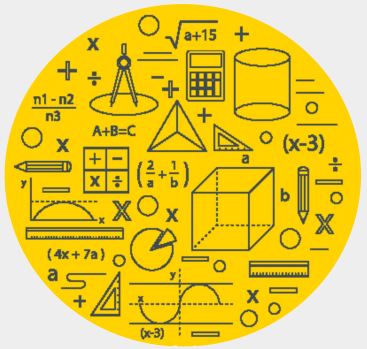


MusicBox

- Box model
- User friendly entry point

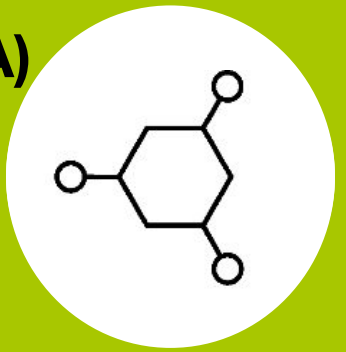
Intro to MUSICA System

Model-Independent Chemistry Module (MICM)



- Chemistry solver
- Ordinary Differential Equation (ODE) integration

Multi-scale Infrastructure for Chemistry and Aerosols (MUSICA)



- Chemistry set up
- Mechanism configuration

MusicBox

- Box model
- User friendly entry point

MUSICA/MusicBox needs improved user guidance

Motivations:

- Incomplete documentation → limited user base ❌
- Gaps in support resources → complex use cases ❌
- Poor user knowledge outside the development team ❌

Objectives:

1. Clear and accessible documentation
2. Interactive tutorials with added performance features
3. A method for user engagement

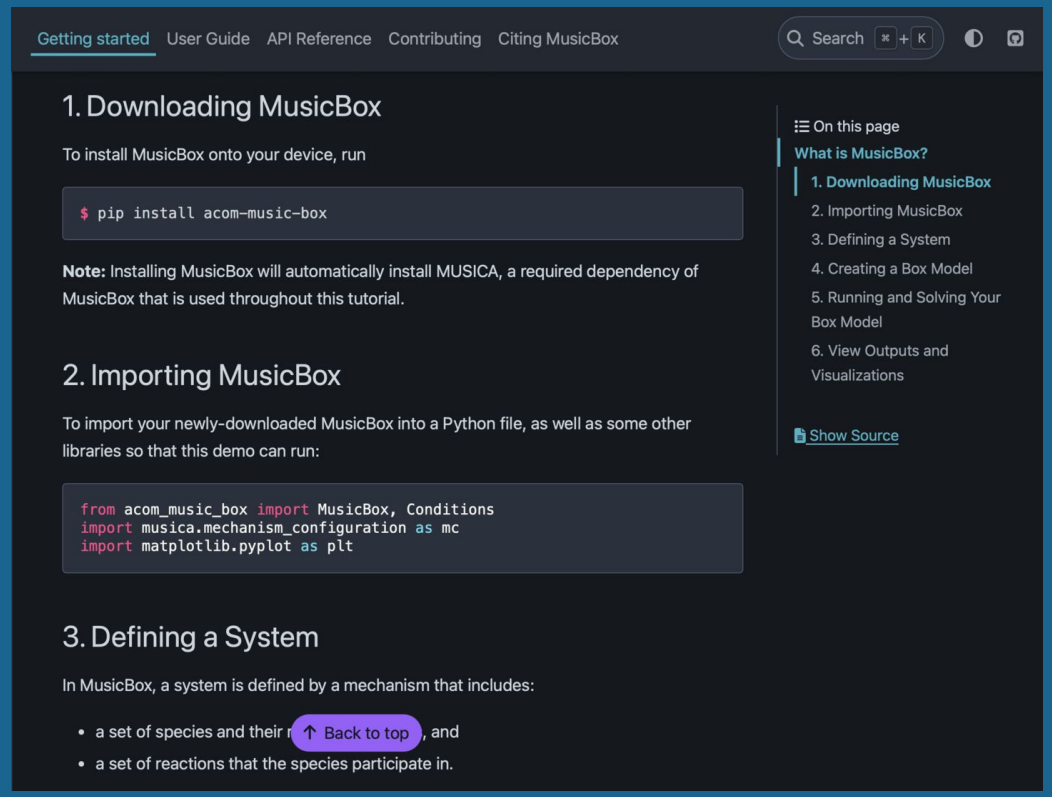
Enhance Documentation

- “Essentials” getting started guide
- Detailed user guide
- Python API reference
- Contributing guide

Build Tutorials

- MusicBox essentials
- Multiple grid cells
- Latin hypercube sampling
- Machine learning with Scikit-learn

MusicBox Essentials

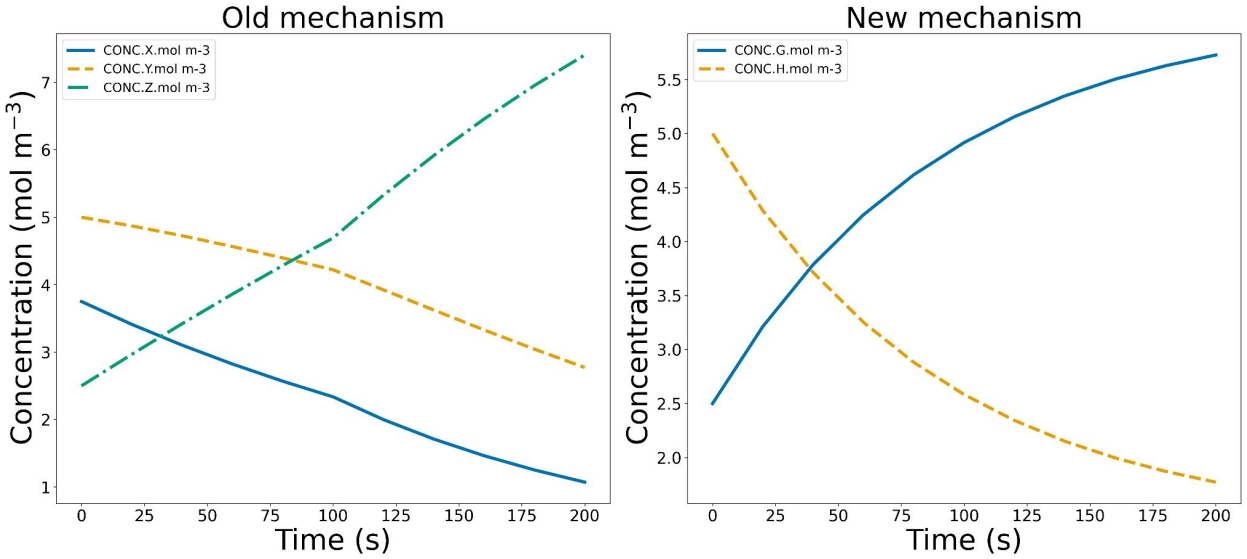


- Introducing the user to MusicBox
- Defining the species (chemicals), their phases (gas), reactions, and the mechanism
- Creating and solving the box model

view in browser **http://bit.ly/4nOMb6T**

interactive notebook **https://bit.ly/4IZSFOZ**

MusicBox Details: Overriding Mechanisms and Loading Box Models

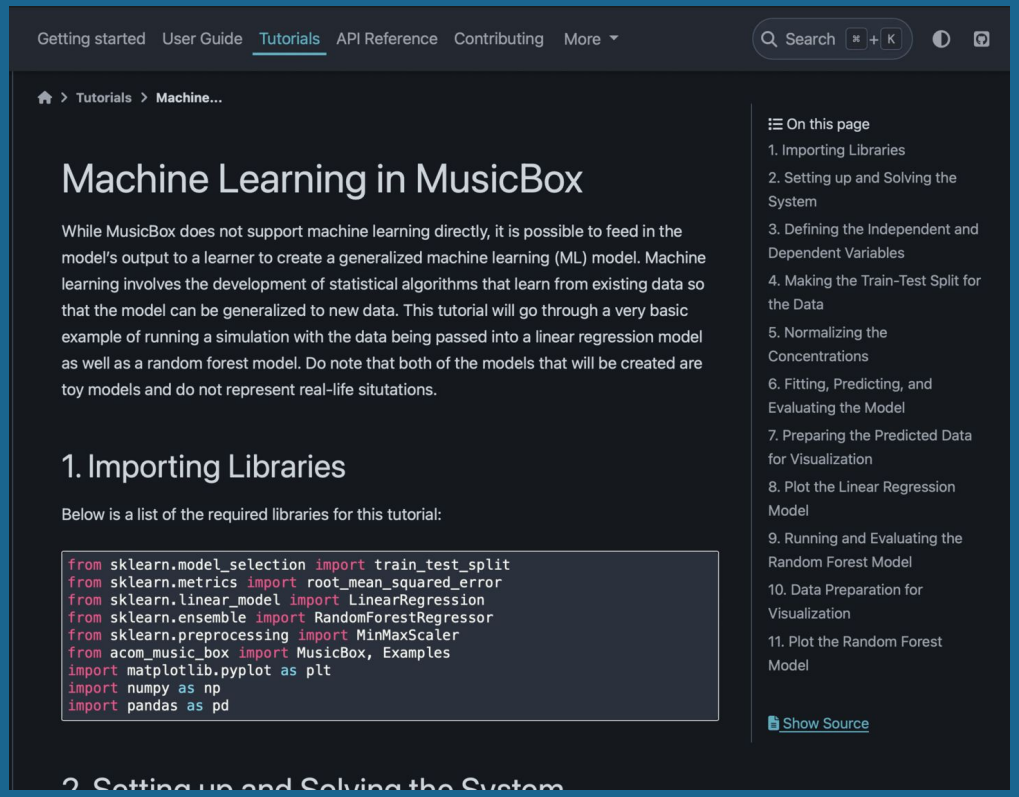


- Overwriting a mechanism replaces the entire chemical system:
  - species, conditions, simulation parameters

- Overwriting an old mechanism with a new one
- Loading a pre-made box model example
- Importing custom box model configurations

interactive notebook **https://bit.ly/3IDBIek**

MusicBox Details: Machine Learning with Scikit-learn

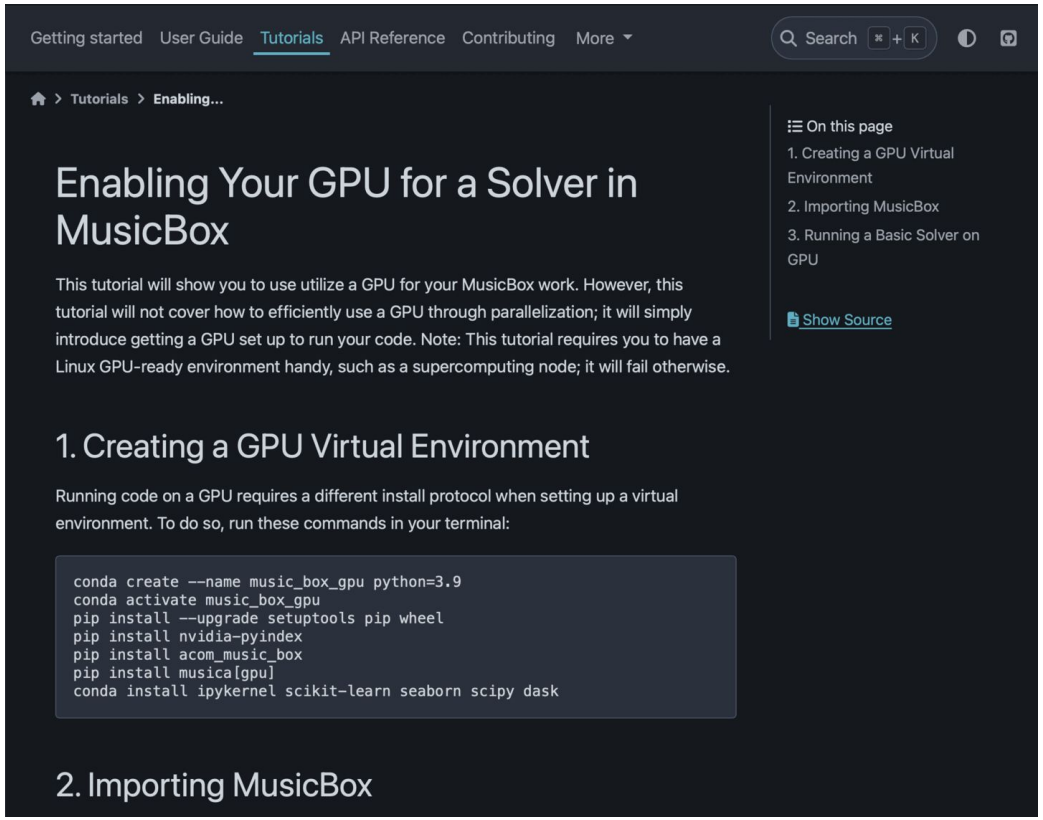


view in browser **http://bit.ly/3GQ5tYQ**

- Splitting the data into training and testing data
- Normalizing the concentrations
- Fitting, predicting, and evaluating with a linear regression and a random forest model
- Visualizing the predicted and actual concentration curves

interactive notebook **https://bit.ly/4m5Lx2U**

MusicBox & MUSICA Details: GPU Solver

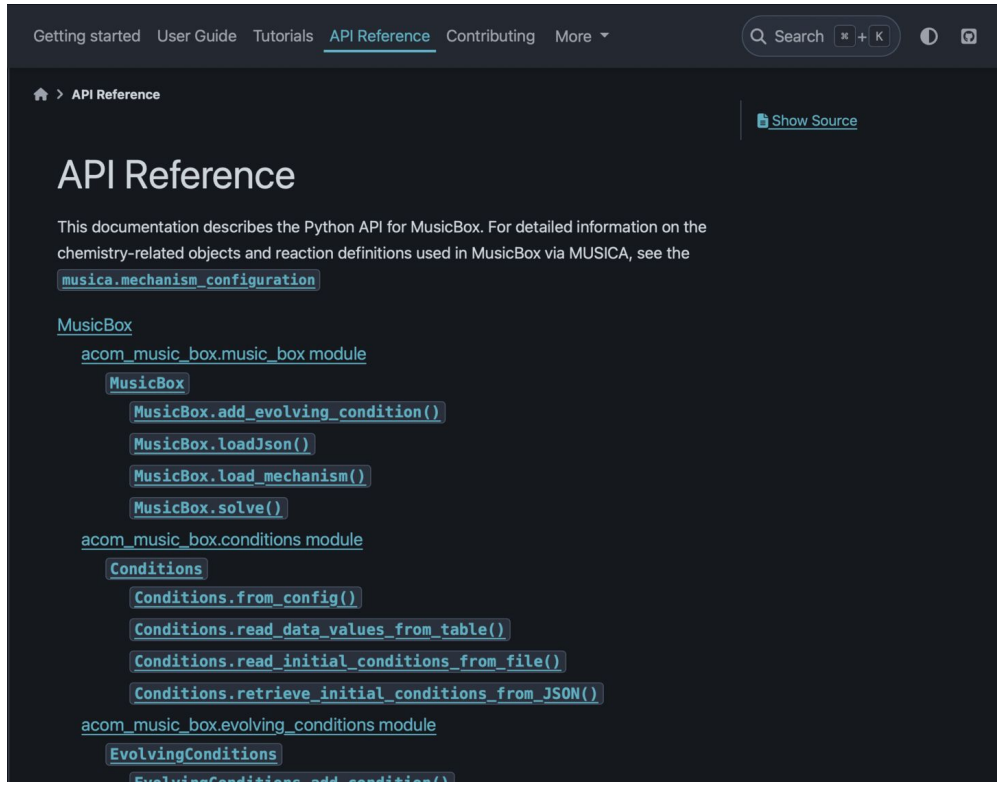


view in browser **http://bit.ly/3GDZVR8**

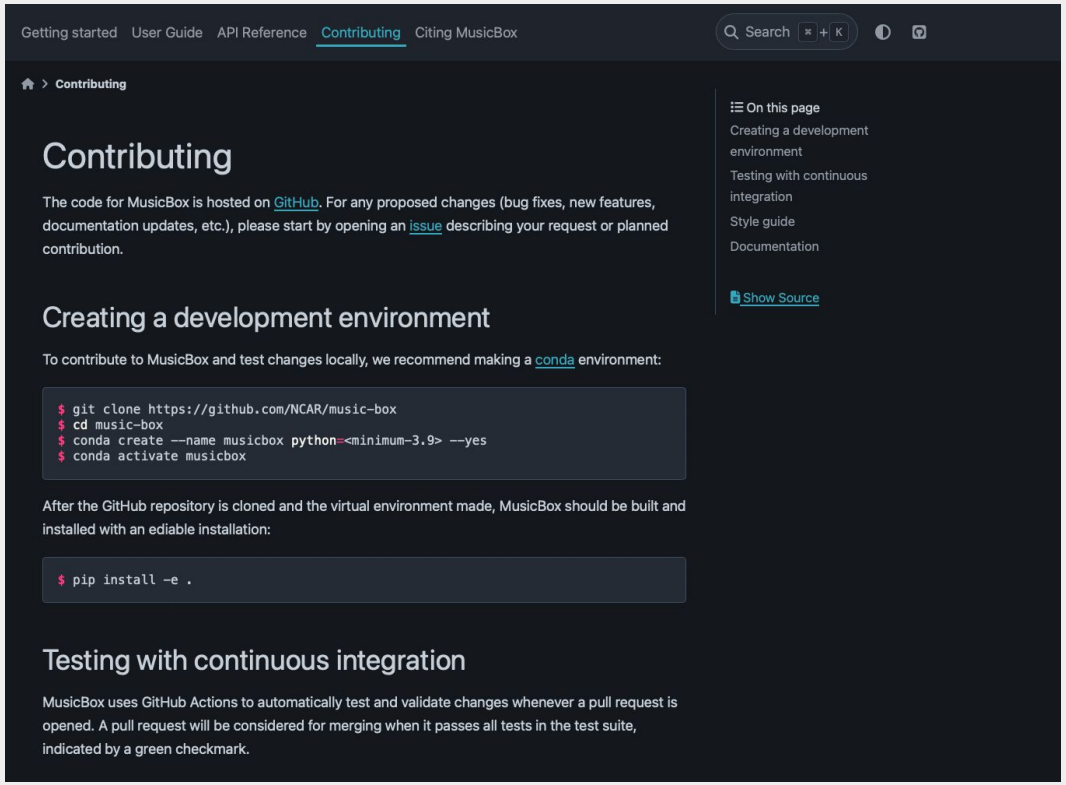
- Creating a new Conda GPU environment
- Verifying that a GPU is available to run on
- Running and visualizing a basic simulation

interactive notebook **https://bit.ly/4lPhpt0**

Documentation: API Reference and Contributing Guide

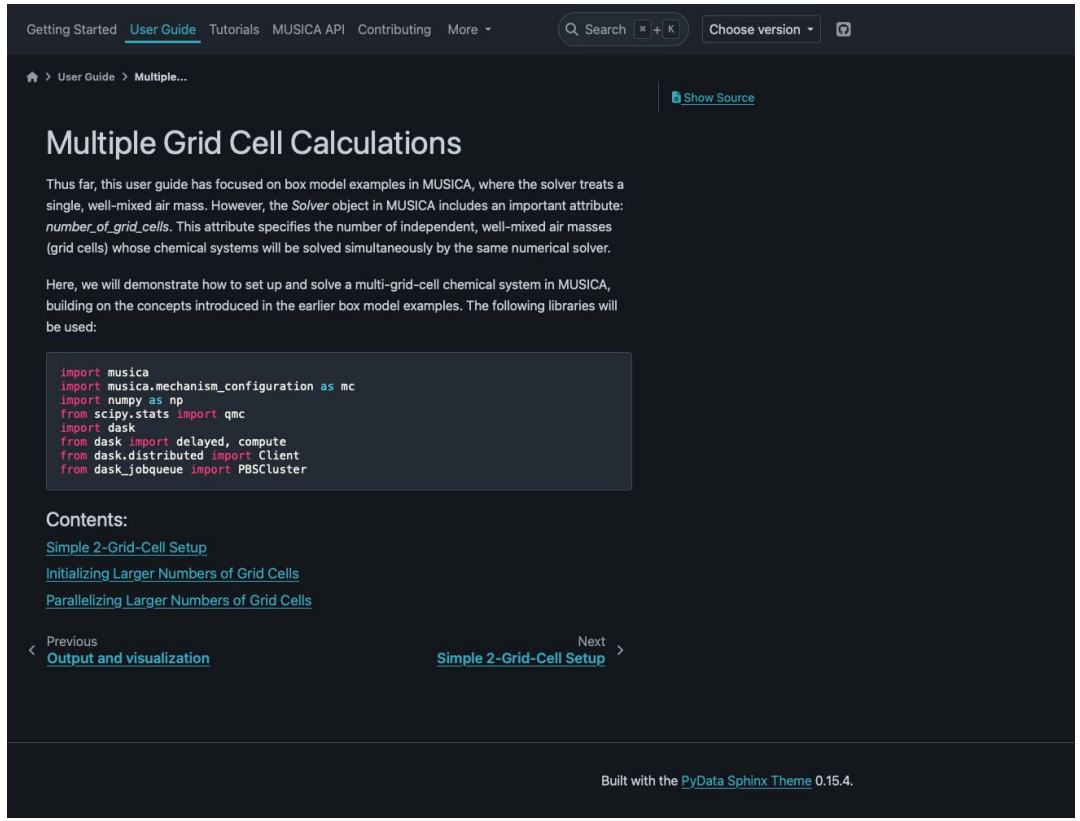


view in browser **http://bit.ly/3laCAXC**



view in browser **http://bit.ly/3GkBeZK**

MUSICA Details: Multiple Grid Cells and Hypercube Sampling

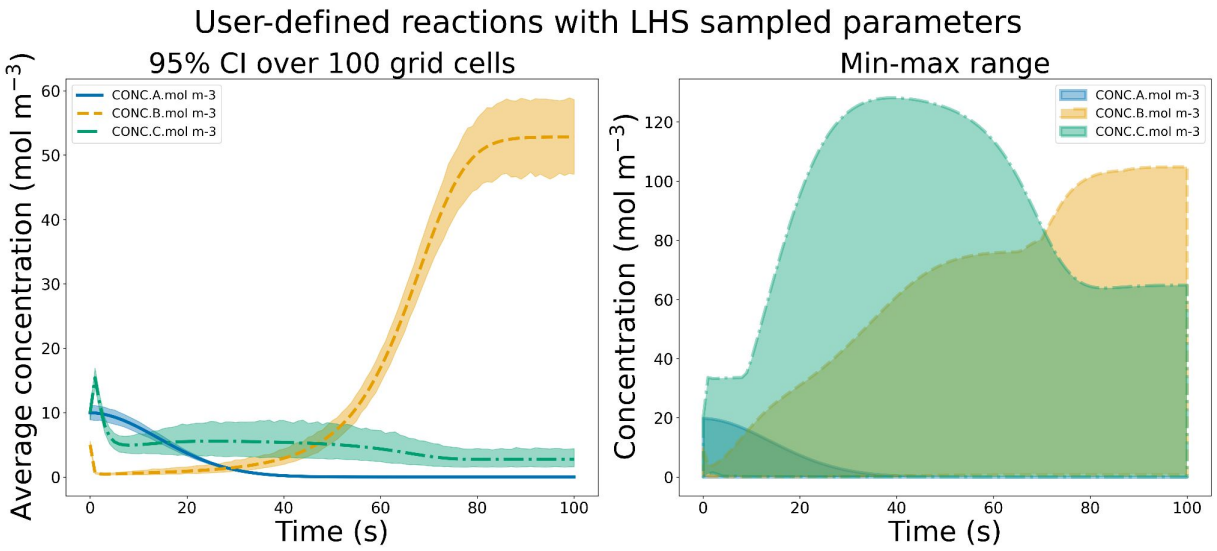


view in browser **http://bit.ly/4lJDkl7**

- Defining each of the grid cells both manually and through Latin Hypercube Sampling (LHS)
- Running the solver through every grid cell
- Plotting the concentration averages, ranges, and confidence intervals of all the grid cells

interactive notebook **https://bit.ly/3TMvJ9k**

MUSICA Details: User-Defined Reactions

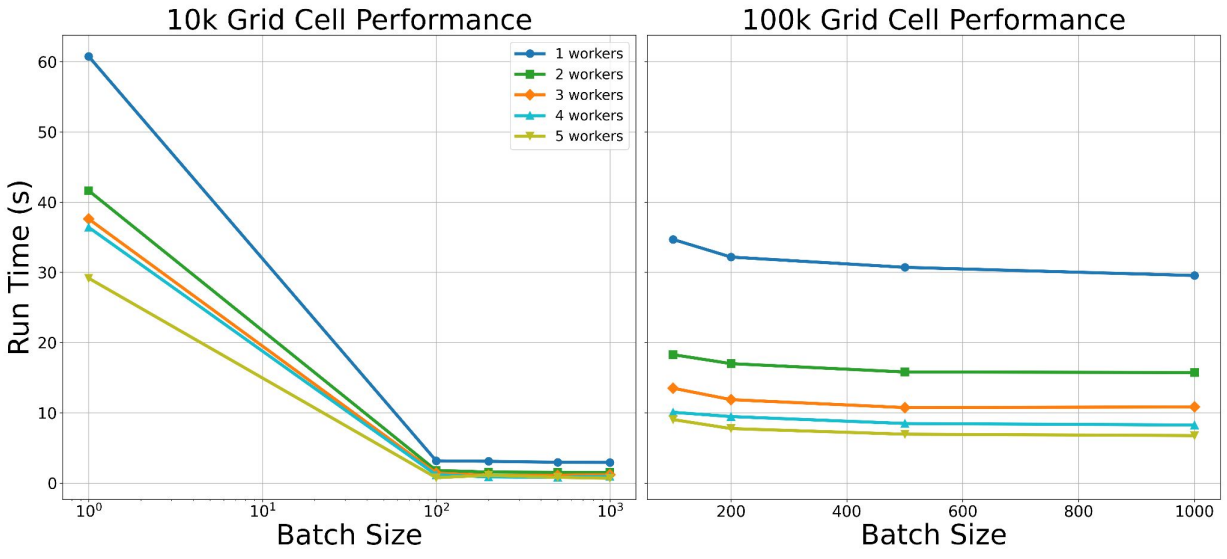


- The initial conditions and reaction rates are both randomized with LHS
  - Leads to a high variability between grid cells

- Defining the UserDefined, Emission, and FirstOrderLoss reaction classes
- Setting the user defined rate parameters
- Solving the system and visualize the concentrations between all the grid cells

interactive notebook **https://bit.ly/46h0nPy**

MUSICA Details: Multiple Grid Cells and Parallelization with Dask

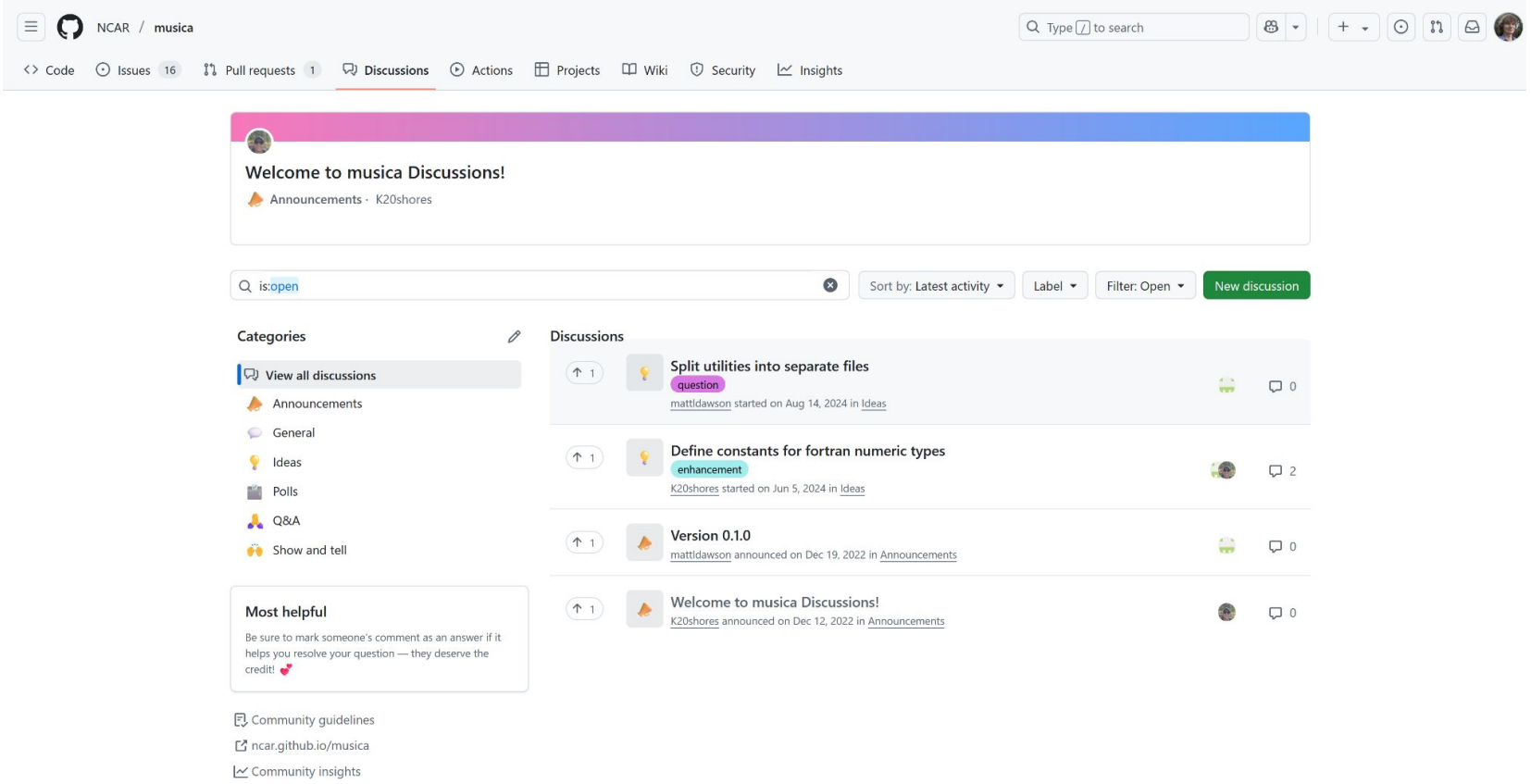


- Scaling tests performed on Derecho
- Parallelization recommended for > 10k grid cells
  - at least 4 workers
  - batch size > 100 cells at a time

- Parallelizing the solving of multiple-grid cells
- Local and PBSCluster implementations
- Batches grid cells due to low individual cost
  - 20x performance improvement optimizing 10k grid cells

interactive notebook **https://bit.ly/4lJBWiA**

Community Forum



MUSICA and MusicBox encourage the use of the community forums for:

- Q&A
- Discussions

Impacts

- Scientists at the **Indian Institute of Tropical Meteorology** are utilizing the tutorials to help students integrate MUSICA into their work
- Improved docs encouraged ACOM scientists to adopt MUSICA/MusicBox for their own projects
- Laid the groundwork for building a machine learning emulator, a key MUSICA goal

Summary and Acknowledgements

MUSICA and MusicBox systems have been enhanced with:

- Detailed documentation & interactive tutorial notebooks
- Additional performance features for multi-grid cell calculations
- Community forum

We gratefully acknowledge the support of:

- Our mentors Jian Sun, Kyle Shores, and Matthew Dawson
- Virginia Do and Jessica Wang
- CISL

This material is based upon work supported by the U.S. National Science Foundation National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1755088. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. National Science Foundation.