

CMAQ Data Available on the Cloud through Amazon's Open Data Program

Fahim Sidi¹, Kristen Foley¹, George Pouliot¹, Alison Eyth², Liz Adams³, Saravanan Arunachalam³ ¹U.S. EPA Office of Research and Development, Research Triangle Park, NC | ²U.S. EPA Office of Air Quality Planning and Standards, Research Triangle Park, NC | ³Institute for the Environment, The University of North Carolina at Chapel Hill, NC

Advantages of Data Storage on the Cloud

- CMAS hosts a Data Warehouse for the larger user community containing popular model inputs and outputs
- Some of these data are on Google Drive and are being transitioned to the Amazon Open Data Program, part of Amazon Sustainability Data Initiative
- Advantages of Data Storage on the Cloud:
 - No egress fees to download/upload and host data as part of Amazon's Open Data Program
 - Fast transfer speeds (upload/download)
 - No limit or throttle for obtaining data
 - Ability to run CMAQ directly on the cloud since inputs are already on the cloud
 - Ability to share data widely and create reproducible workflows for running CMAQ on the cloud

CMAS Data Warehouse	i S J	
Description	Resources on AWS	
CMAS Data Warehouse on AWS collects and disseminates meteorology, emissions and air quality model input and output for Community Multiscale Air Quality (CMAQ) Model Applications. This dataset is available as part of the AWS Open Data Program, therefore egress fees are not charged to either the host or the person downloading the data. This	Description CMAQ Release Benchmark Data for Easy Download	
	Resource type S3 Bucket	
55 bucket is maintained as a public service by the University of North Carolina's CMAS Center, the US EPA's Office of Research and Development, and the US EPA's Office of	Amazon Resource Name (ARN)	
Air and Radiation. Metadata and DOIs for datasets included in the CMAS Data	arn:aws:s3:::cmaq-release-benchmark-data-for-ea	asy-download
Warehouse are available from the CMAS Dataverse site:	AWS Region	
nttps://dataverse.unc.edu/dataverse/cmascenter	us-east-1	
Update Frequency	AWS CLI Access (No AWS account required)	
New data is added as soon as it is available.	aws s3 lsno-sign-request s3://cmaq-release-b	benchmark-dat
License	easy-download/	
There are no restrictions on the use of this data. US EPA License (https://pasteur.epa.gov/license/sciencehub-license.html)	Browse Bucket	
Documentation		
https://dataverse.unc.edu/dataverse/cmascenter	Description	
Managed By	CMAQ CONUS-2 Benchmark Data	
Managed by	Resource type	
	S3 Bucket	
CHAS	Amazon Resource Name (ARN)	
OWAG	arn:aws:s3:::cmas-cmaq-conus2-benchmark	

Accessing AWS Data for Cloud Computing

- Download Data from AWS Registry (https://registry.opendata.aws/cmas data-warehouse/) using AWS command line interface
- Tutorials have been developed to run CMAQ on AWS and Microsoft Azure which includes instructions to access data directly on the cloud:
- AWS: <u>https://cyclecloud-</u> cmaq.readthedocs.io/en/latest
- Microsoft Azure: https://cyclecloudcmag.readthedocs.io/en/latest



documentation

Create Single VM

3. Performance and Cos

Post-process and QA

6. Logout and Delete ParallelCluster

Additional Resources

ontribute to this Tutorial

Training, research, and

onsulting. GenAl and HW/SV

Co-Design, Custom and on-site

At by RhimAtt 🔹 🚺

8. Future Work

Create a Parallel Cluster and un CMAQv5.4

4. Developer Guide to install and run CMAQv5.4 on Single VM or Parallel Cluster

MAO on AWS Tutorial	
inne on nwo futorial	

under continuous development Previous version is available here:

0 0

Community Multiscale Air Quality Model

The Community Multiscale Air Quality (CMAQ) modeling system an active open-source development project of the U.S. EPA. The CMAQ system is a Linux-based suite of models that requires significant computational resources and specific system configurations to run. CMAQ combines current knowledge in atmospheric science and air quality modeling, multi-processor computing techniques, and an open source framework to deliver fast, technically sound estimates of ozone, particulates, toxics and acid

 For additional background on CMAQ please visit the U.S. EPA CMAQ Website. CMAQ is a community modeling effort that is supported by the Community Modeling and Analysis System (CMAS) Center at the University of North Caroline at Chapel Hill.

Tutorial Overview

This document provides tutorials on how to use Amazon Web Service (AWS) high performance computing services. AWS's Elastic Compute Cloud (EC2) allows you to create and run a single Linux virtual machine, or Single VM, in the cloud. Another AWS resource, ParallelCluster, is a cluster management tool that helps you to deploy and manage many VMs in the cloud. ParallelCluster is well suited for large modeling applications as it automatically sets up the required compute resources. scheduler, and shared file system

The following tutorials will walk you through running a CMAQ benchmark case on both a Single VM and on the more advanced ParallelCluster. A benchmark case is provided with pre-installed software allowing you to jump right into running CMAQ and post-processing model output for analysis and visualization. The Developers Guide provided in Chapter 4 describes how to do the software installation process yourself in order to set up simulations tailored to your own applications.

The tutorials are aimed at users with cloud computing experience that are already familiar with AWS. For those with no cloud computing experience we recommend signing up for a free AWS Educate account as it is open to any individual, regardless of where they are in their education, technical experience, or career journey. If you are a new user to AWS, you can sign up for a free tier account. There are also lowcost tutorials available to learn Parallel Cluster from AWS: Parallel Cluster Tutorial and AWS Workshops on HPC Computing

See talk & workshop by Liz Adams during Cloud Computing session to learn more!

U.S. Environmental Protection Agency Office of Research and Development

- with scaling (WWLLNs)
- this dataset

EPA's Air Quality TimE Series (EQUATES) Project Data

CMAQ 2018 Modeling Platform

- mechanism

CMAQ Benchmark Data

- Domain
- of CMAQv5.4

CMAQ Input Data Available on AWS

CMAQ WWLLNs Lightning Data

 CMAQ-ready gridded hourly observed lightning flash density data for the 12 km CONUS domain from the World Wide Lightning Location Network (WWLLN)

• Used by CMAQ to estimate lightning emissions inline; see <u>chapter 6 of CMAQ Users Guide</u> for additional details on how to configure CMAQ to use

• Temporal Coverage: Jan 1, 2016 – Dec 31, 2022

• Link to Data: https://cmas-wwllnlightning.s3.amazonaws.com/index.html

· CMAQ-ready gridded hourly emissions, meteorology and air quality modeling inputs for the 12 km CONUS domain & 108 North Hemispheric Domain

Additional Documentation: <u>www.epa.gov/cmag/equates</u>

• Temporal Coverage: Jan 1, 2017 – Dec 31, 2017

• Link to data: https://cmasequates.s3.amazonaws.com/index.html

· CMAQ-ready gridded hourly emissions, meteorology and air quality modeling inputs for the 12 km CONUS domain

• Inputs available to run a CMAQ simulation with CB6R5 or the Community Regional Atmospheric Chemistry Multiphase Mechanism (CRACMM) version 1.0

 Includes inputs to an inline version of Model of Emissions of Gases and Aerosols from Nature (MEGAN) in CMAQ to estimate BVOC and SOIL NO emissions inline

• Temporal Coverage: Jan 1, 2018 – Dec 31, 2018

 Link to data: https://cmas-cmag-modeling-platform-2018.s3.amazonaws.com/index.html

• CMAQ-ready gridded hourly emissions, meteorology and air quality modeling inputs for the 12 km Northeast

Benchmark dataset used to verify successful installation

• Temporal Coverage: July 1, 2018 – July 2, 2018

 Link to data: https://cmascmaq.s3.amazonaws.com/index.html





main size: 299x459 | Max = 3.8 at (103, 141 Mean: 0.023 | Median:



Surface Gridded NOx Emissions no BEIS, RWC 0.750 moles/s

July 1,2018 16:00:00 Min= 0.000 at (53,1), Max= 6.926 at (65,66)

SMOKE Emissions Platform Data Available on AWS

National Emissions Inventory (NEI) 2019 Modeling Platform

- CMAQ-ready gridded hourly emissions from SMOKE for the 12 km CONUS domain
- 2019 emissions based on 2017 NEI released in 2020 including sector specific updates to represent emissions in 2019
- Temporal Coverage: Dec 22, 2018 Dec 31, 2019
- Supporting Documentation: https://www.epa.gov/air-emissionsmodeling/2019-emissions-modeling-platform
- Link to data: https://2019platform.s3.amazonaws.com/index .html

NEI 2016v3 Modeling Platform

- CMAQ & CAMx-ready gridded hourly emissions from SMOKE for the CONUS domain at various resolutions
- 2016 emissions version 3 (2016v3), an update to 2016 version 2 which incorporates updates made in response to comments by stakeholders, improved methods, some corrections, and refinements to projection factors
- Temporal Coverage: 2016, 2023, 2026
- Supporting Documentation: https://www.epa.gov/air-emissionsmodeling/2016v3-platformLink to data
- Link to data: https://2016v3platform.s3.amazonaws.com/index html

AMET Observation Data Available on AWS

- Observational data formatted for the Atmospheric Model Evaluation Tool (AMET) available from various networks including AQS, CASTNET, IMPROVE, MDN, NADP, SEARCH, CSN, NAPS, ESRL, TOAR
- Temporal Coverage: 2000 2019
- Link to data: https://cmasamet.s3.amazonaws.com/index.html

- & 2017-2018 CRACMM
- 2020 Emissions Modeling Platform based on 2020 NEI





Datasets coming to AWS Soon

MPAS-CMAQ 120 km uniform global mesh inputs for 2014-2018 CB6R5M

MPAS-CMAQ 92-25 km variable resolution mesh inputs for 2016 CB6R5M





