CMAS Center presents:

The CMAS Webinar Series

"Use of NOAA's Global Forecast System Data in the Cloud for Community Air Quality Modeling"

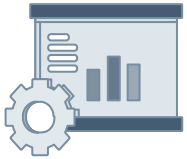
Dr. Patrick Campbell

Research Associate Professor, Associate Director of SESS, and NOAA Affiliate

Webinar Tools

Chat with hosts feature: Zoom/technical issues

Q & A feature: Questions for the speaker



Poll feature: Encourage participation from all

Use of NOAA's Global Forecast System Data in the Cloud for Community Air Quality Modeling

Patrick Campbell Weifeng (Rick) Jiang Zachary Moon
Youhua Tang Sonny Zinn

2024 CMAS Webinar
Cloud Computing
June 25, 2024 @ 1:00 p.m. EDT

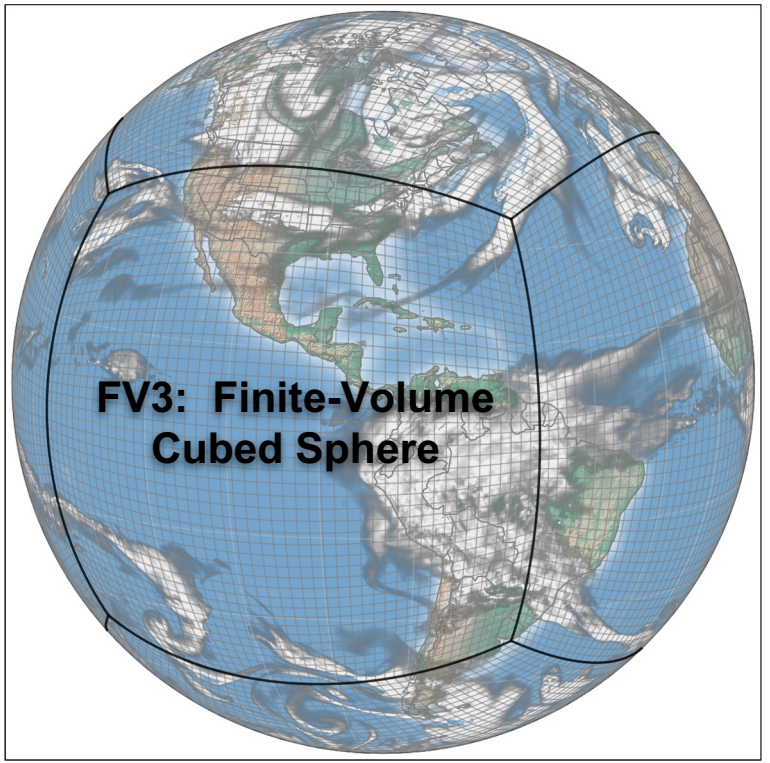
Outline

1. Background on NOAA's GFSv16 and NACC
2. AWS-HPC environment and NACC in the cloud
3. Development of the NACC-Cloud interface for the community
4. Examples of current applications of NACC-Cloud
5. Conclusions and path forward

The Global Forecast System Version 16 (GFSv16)

NOAA's Environmental Modeling Center Operational GFS Version 16

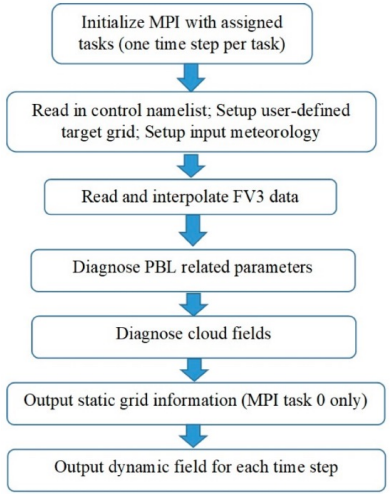
- Uses a FV3-based dynamical core
- Vertical structure of 127 layers
- Extensive model top of 80 km
- First model layer thickness ~ 20 m
- Relatively high spatial resolution of ~13 km
- Advanced physical parameterizations
- Uses the Global Data Assimilation System
- Includes NetCDF forecast output history files
- Provides hourly output out to 72-hr forecast
- Used in the current operational AQF at NOAA



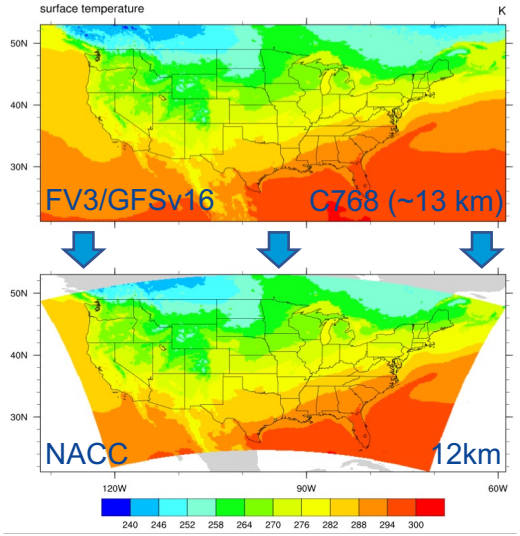
Campbell et al., *GMD* (2022)

The NOAA-EPA Atmosphere Chemistry Coupler

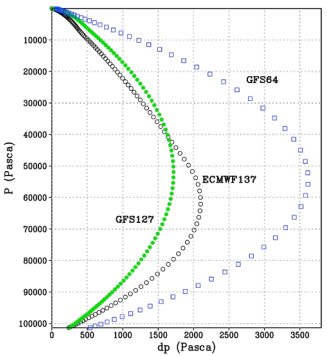
- The **N**OAA-EPA **A**tmosphere-**C**hemistry **C**oupler (NACC) (i.e., *"knack": meaning an acquired skill*) is adapted from EPA's Meteorology-Chemistry Interface Processor (MCIP) version 5 (Otte and Pleim, 2010).

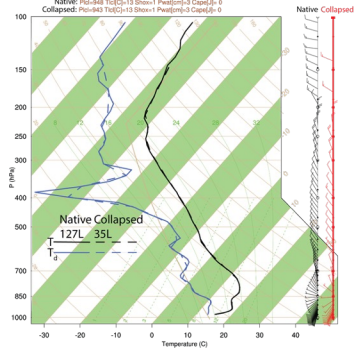


Directly Interpolated Fields



Redefined Vertical Structure





GFSv16 has higher vertical resolution with 127L and thinner 1st layer (20 m thick) than previous GFS.

NACC uses either native or collapsed GFSv16 layers, and there is good agreement in structure.

- NACC is available to the greater scientific community at: https://github.com/noaa-oar-arl/NACC.

Campbell et al., *GMD* (2022) Tang et al., GMD (2022) Campbell et al., *Atmosphere* (2023)

Outline

1. Background on NOAA's GFSv16 and NACC
2. AWS-HPC environment and NACC in the cloud
3. Development of the NACC-Cloud interface for the community
4. Examples of current applications of NACC-Cloud
5. Conclusions and path forward

Why bring NOAA's GFSv16 and NACC into the Cloud?

- Obtaining data, preprocessing tools, and running gridded meteorological models (e.g. WRF) for CMAQ regional applications across the globe can be very time and resource intensive.

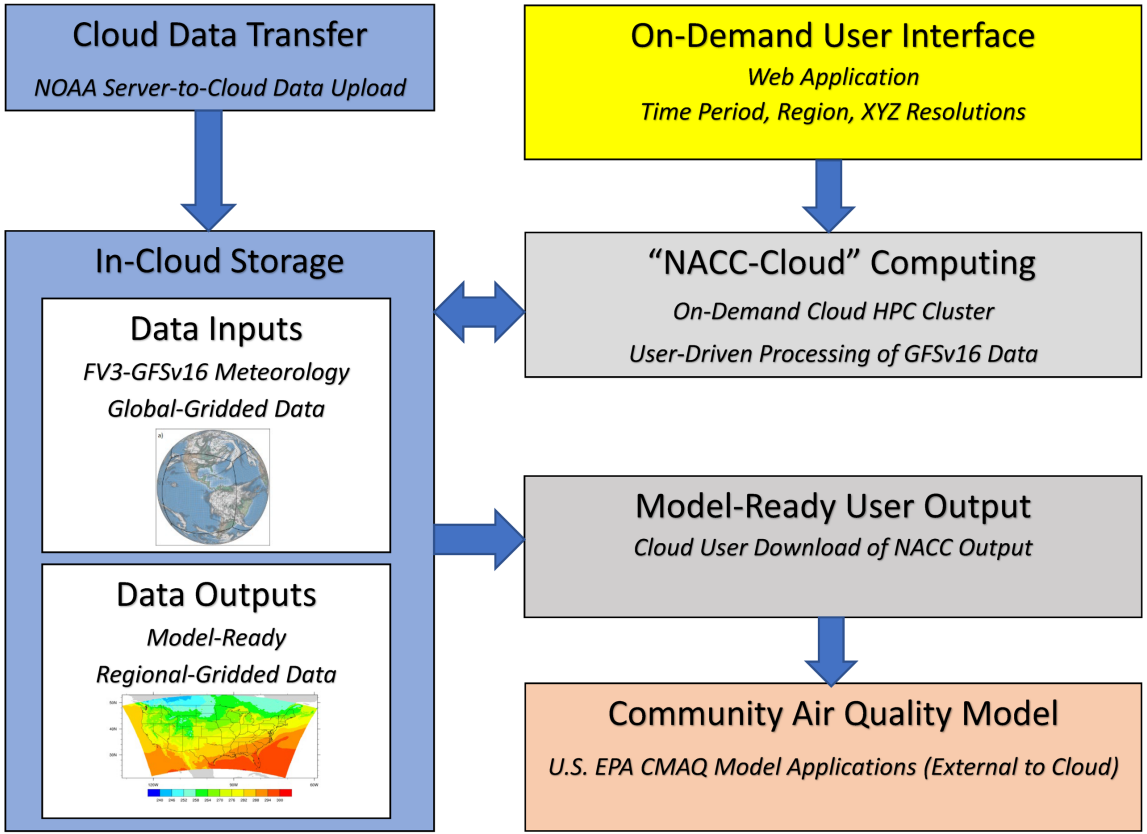
- Use of NOAA's operational GFSv16 (gridded NetCDF and hourly) input data, while robust, can be large and cumbersome to process for the scientific community (i.e., data intensive).

- Cloud computing and storage platforms are desirable as they are highly-customizable, on-demand, and much more scalable than traditional local servers…**User-Ready CMAQ Inputs!**

Main Objective:
Develop cloud interface to NOAA GFS-driven CMAQ applications for any regional domain globally…

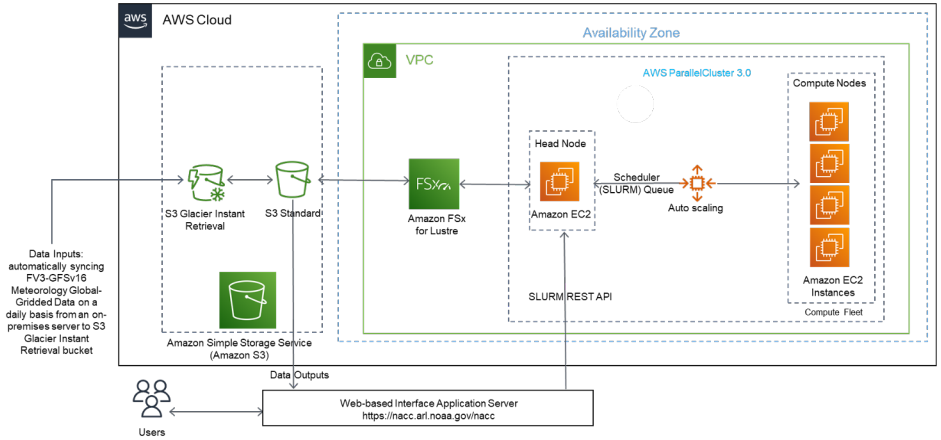
"NACC-Cloud"

Methods: Cloud-Based Data Transfer and Processing



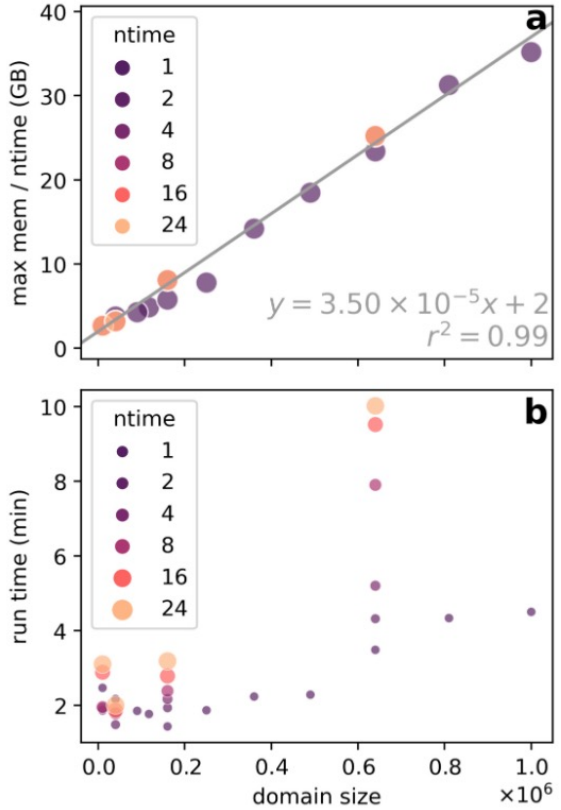
Methods: Streamlined AWS-Based Development

AWS S3 and HPC Configurations



Component	Option	Notes
Head Node—c5n.2xlarge	Reserved	4 cores, 21 GB RAM, up to 25 Gbps network bandwidth
Compute Nodes—r6i.32xlarge × 4	On Demand	(1) each node: 96 cores, 1 TB RAM, 50 Gbps network bandwidth (2) assuming: 4 nodes, 4 h per day
File system—Amazon® FSx for Lustre	-	1200 GB
Data input storage—Amazon® S3 Glacier Instant Retrieval	-	(1) GFSv16 data files: increase—200 GB/day, 6 TB/month, 72 TB/year (2) Monthly cost—should be accumulated from previous months
Data output storage—Amazon® S3 Standard	-	Output for one run (one-month data as input): 3.5 GB, 35 GB/day, 1 TB/month
End User Data download	-	Output for one run (one-month data as input): 3.5 GB, 35 GB/day, 1 TB/month
Other resources	-	VPC, EBS, Elastic IP,
AWS Support	-	10% of monthly AWS usage.

Scalability of NACC-Cloud



Outline

1. Background on NOAA's GFSv16 and NACC
2. AWS-HPC environment and NACC in the cloud
3. Development of the NACC-Cloud interface for the community
4. Examples of current applications of NACC-Cloud
5. Conclusions and path forward

Methods: Development of a Web-Based User Interface

- Web app based on the next-gen ARL READY framework (using the Java Spring ecosystem)
  - **R**eal-time **E**nvironmental **A**pplications and **D**isplay s**Y**stem: https://www.ready.noaa.gov/index.php
- NACC namelist and job script generated using Python based on web app form input, communicated via an API endpoint of the web app
- Slurm job submission via Slurm REST API using a generated Python client library
- A cron job syncs generated outputs from the Amazon FSx for Lustre file system to an Amazon S3 bucket, where they can be downloaded
  - S3 URLs are presented to the user in the web app interface
  - Will add user email notification that job is done and S3 URLs are ready for pickup

Results: NACC-Cloud Web Demo

NOAA-EPA Atmosphere Chemistry Coupler

https://nacc.arl.noaa.gov/nacc/

Date range:

Start date (UTC) 2022-07-22 | start hour 12

Ending date (UTC) 2022-07-23 | ending hour 12

Projection:

GDTYP 2

P_ALP 33.0 deg | P_BET 45.0 deg | P_GAM -97.0 deg

XCENT -97.0 deg | YCENT 40.0 deg

Domain:

XORIG -2508000.0 m | YORIG -1716000.0 m

XCELL 12000.0 m | YCELL 12000.0 m

NCOLS 442 | NROWS 265

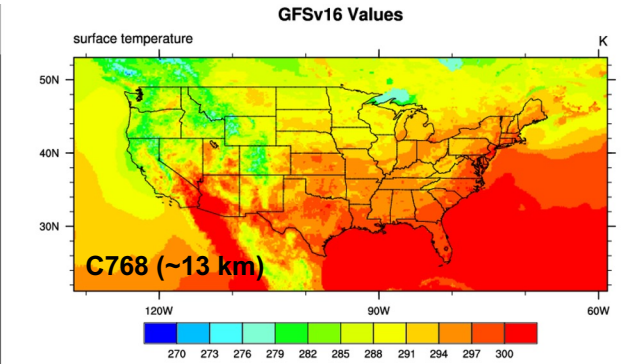
CTMLAYS 1.000000, 0.995253, 0.990479, 0.985679, 0.980781, 0.975782, 0.970684, 0.960187, 0.954689, 0.936895, 0.930397, 0.908404, 0.888811, 0.

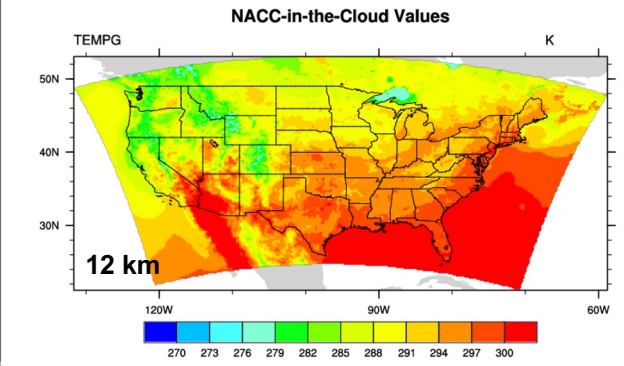
Restore default values | Start

For Lambert conformal (GDTYP=2), P_ALP and P_BET are the true latitudes of the projection cone, and P_GAM is the central meridian of the projection. XCENT and YCENT are the reference longitude and latitude for the domain. For polar stereographic (GDTYP=6), P_ALP is 1.0 for North Polar and -1.0 for South Polar, while P_BET is the secant latitude (latitude of true scale). As with GDTYP=2, P_GAM is the central meridian and XCENT and YCENT are the reference longitude and latitude for the domain. See the I/O API grid doc for more information. Note that the current minimum start date/time is **March 23, 2021 at 12 UTC**.

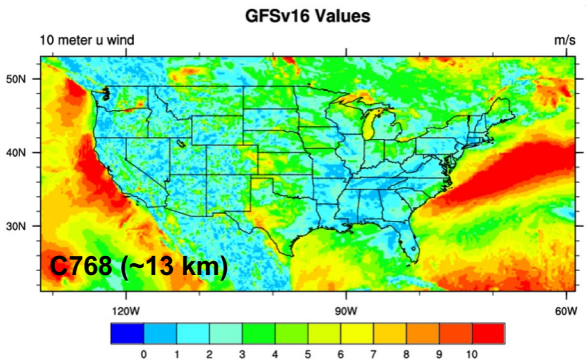
Results: GFSv16 and NACC-Cloud Comparison

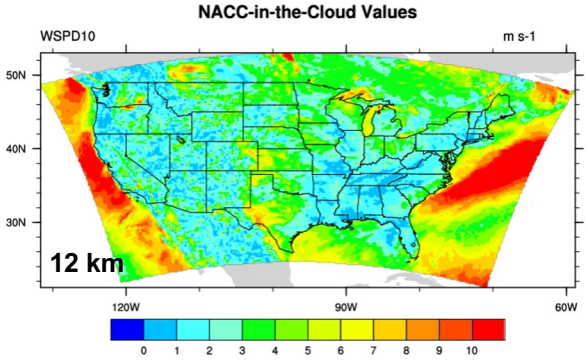
Surface Temperature



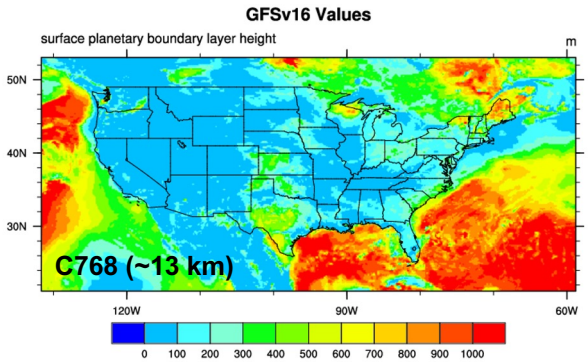


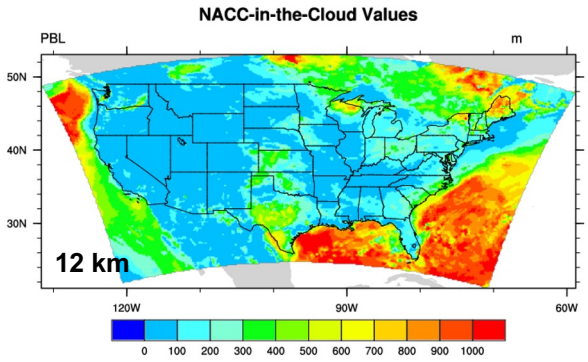
10-m Wind Speed





PBL Height

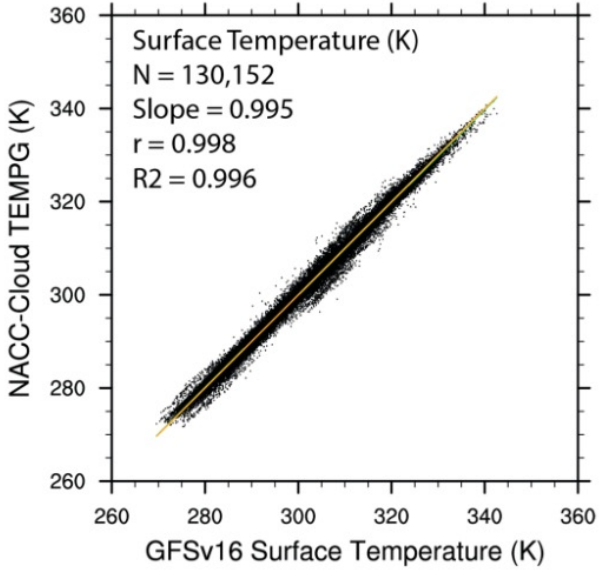




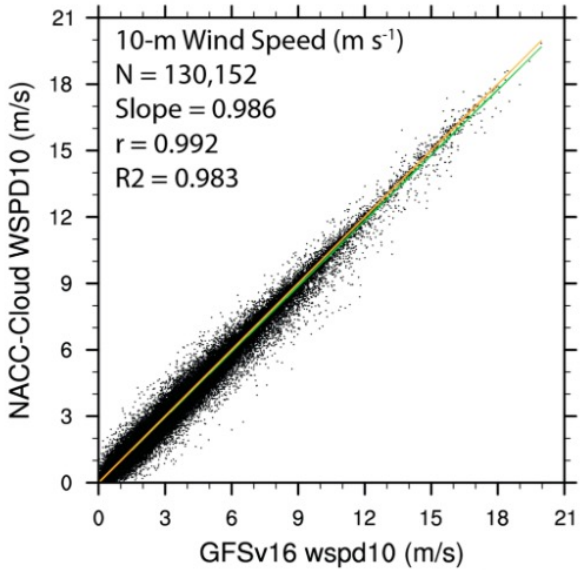
Excellent spatial agreement with GFSv16 inputs and NACC-Cloud outputs.

Results: GFSv16 and NACC-Cloud Comparison

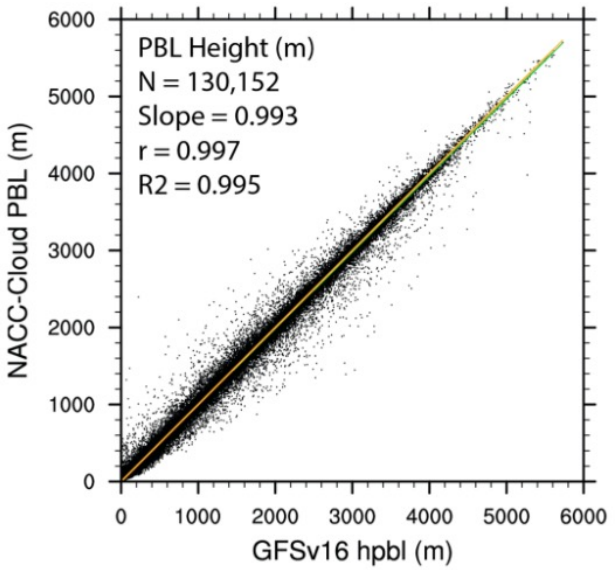
Surface Temperature



10-Meter Wind Speed



Boundary Layer Height



Provides confidence in using the NACC-Cloud system for user GFS-driven CMAQ applications.

Outline

1. Background on NOAA's GFSv16 and NACC
2. AWS-HPC environment and NACC in the cloud
3. Development of the NACC-Cloud interface for the community
4. Examples of current applications of NACC-Cloud
5. Conclusions and path forward

Example users and applications of NACC-Cloud

NACC-Cloud users from different U.S. states and countries around the world.

Some examples include:

- U.S. EPA: Quasi-Real Time Modeling (QRT)
- U.S. EPA: Expedited Modeling of Burn Events Results (EMBER)
- Georgia EPD: NACC-Cloud Application for State-Level AQF

Outline

1. Background on NOAA's GFSv16 and NACC
2. AWS-HPC environment and NACC in the cloud
3. Development of the NACC-Cloud interface for the community
4. Examples of current applications of NACC-Cloud
5. Conclusions and path forward

Conclusions

- NACC-Cloud has been developed with a functional web-based user interface: https://nacc.arl.noaa.gov/nacc/ and Campbell et al., *Atmosphere* (2023)

- NACC-Cloud is available for GFS-driven CMAQ applications for any regional domain globally for the entire operational GFSv16 data record (March 23, 2021 - Current Day, updated daily at ~ 1800 UTC).

- Please contact me for further questions and/or collaboration ideas for using NACC-Cloud for your applications (Patrick.C.Campbell@noaa.gov).

Final Notes

- Need to explore additional funding sources to continue NACC-Cloud on AWS.
- Using cloud storage and HPC resources can appear costly up front.
- Overall costs of cloud applications (e.g., NACC-Cloud) *can be much less* than costs to maintain dedicated onsite HPC servers with necessary support to run research- or operational-grade NWP models.
- Cloud resources are widely available, easy to set up, on-demand, and scalable.

Where to access NACC-Cloud?

Website

https://nacc.arl.noaa.gov/nacc/

Requests

https://forms.gle/jUWFKLyY6WGKySkv6

Feedback

https://forms.gle/ETwKS2qzoHndaruJ9

CMAS Conference Call for Abstracts

Abstract deadline:
July 1st, 2024

Visit cmascenter.org

CMAS Conference Call for Abstracts

Session Topics

- Air Quality, Climate and Energy
- Air Quality Studies that focus on Environmental Justice
- Cloud Computing
- Emissions Inventories, Models, and Processes
- Leveraging Community Knowledge through CRACMM
- Machine Learning and Reduced Form Models: Developments and Applications
- Model Development
- Modeling to Support Exposure and Health Studies and Community-scale Applications
- Multiscale Model Applications and Evaluations
- Regulatory Modeling and SIP Applications
- Remote Sensing/Sensor Technology and Measurements Studies
- Wildfire Emissions and Air Quality