CMAQ inverse modeling, WRF downscaling, and MPAS simulations on the cloud: Challenges and best practices

Matthew Alvarado, Nicholas Heath, Chase Calkins, Rebecca Adams-Selin, John Henderson, Elizabeth Bettencourt, Richard Pernak, and Amy McVey

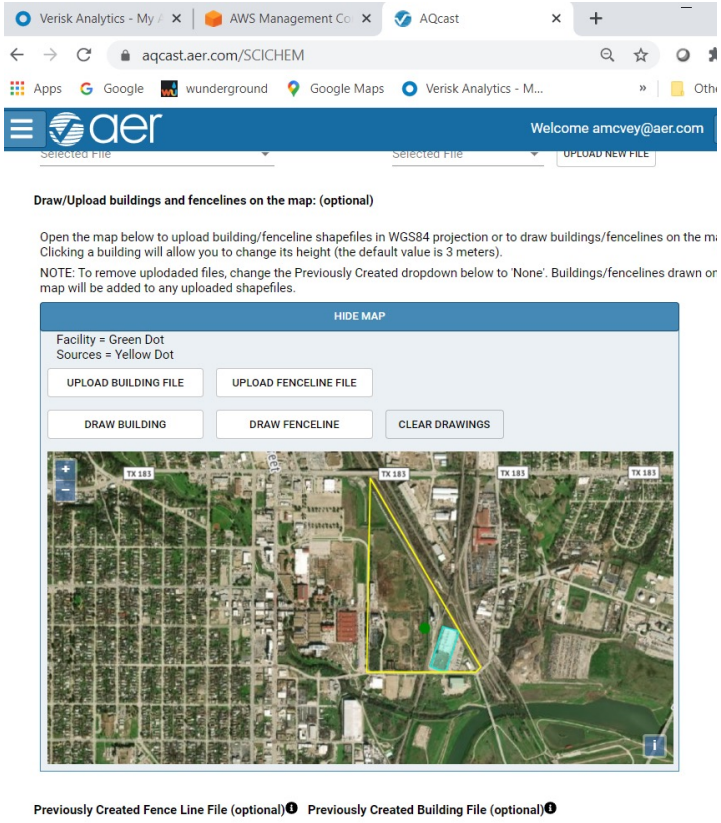
20th Annual CMAS Conference

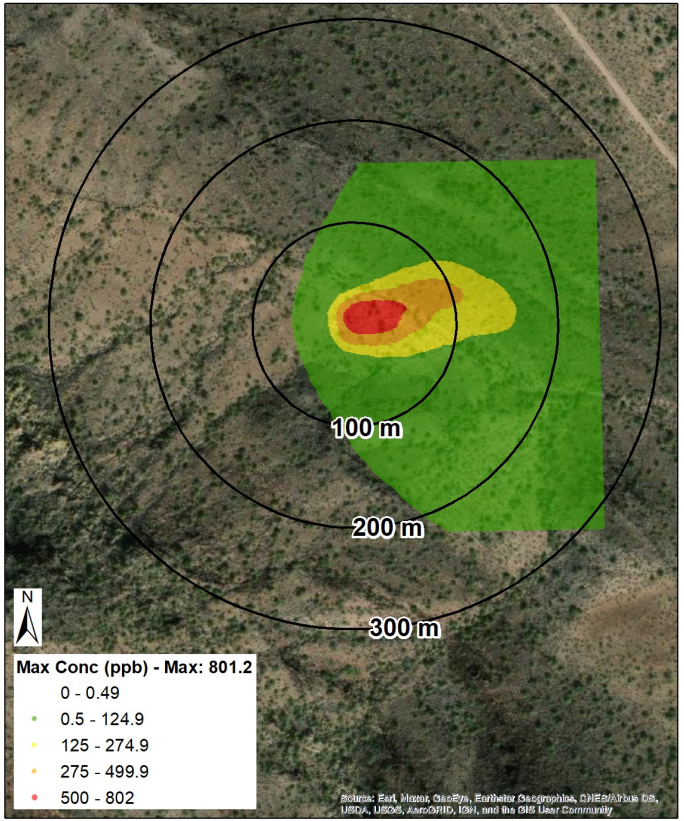
October 20, 2021

SERVE | ADD VALUE | INNOVATE

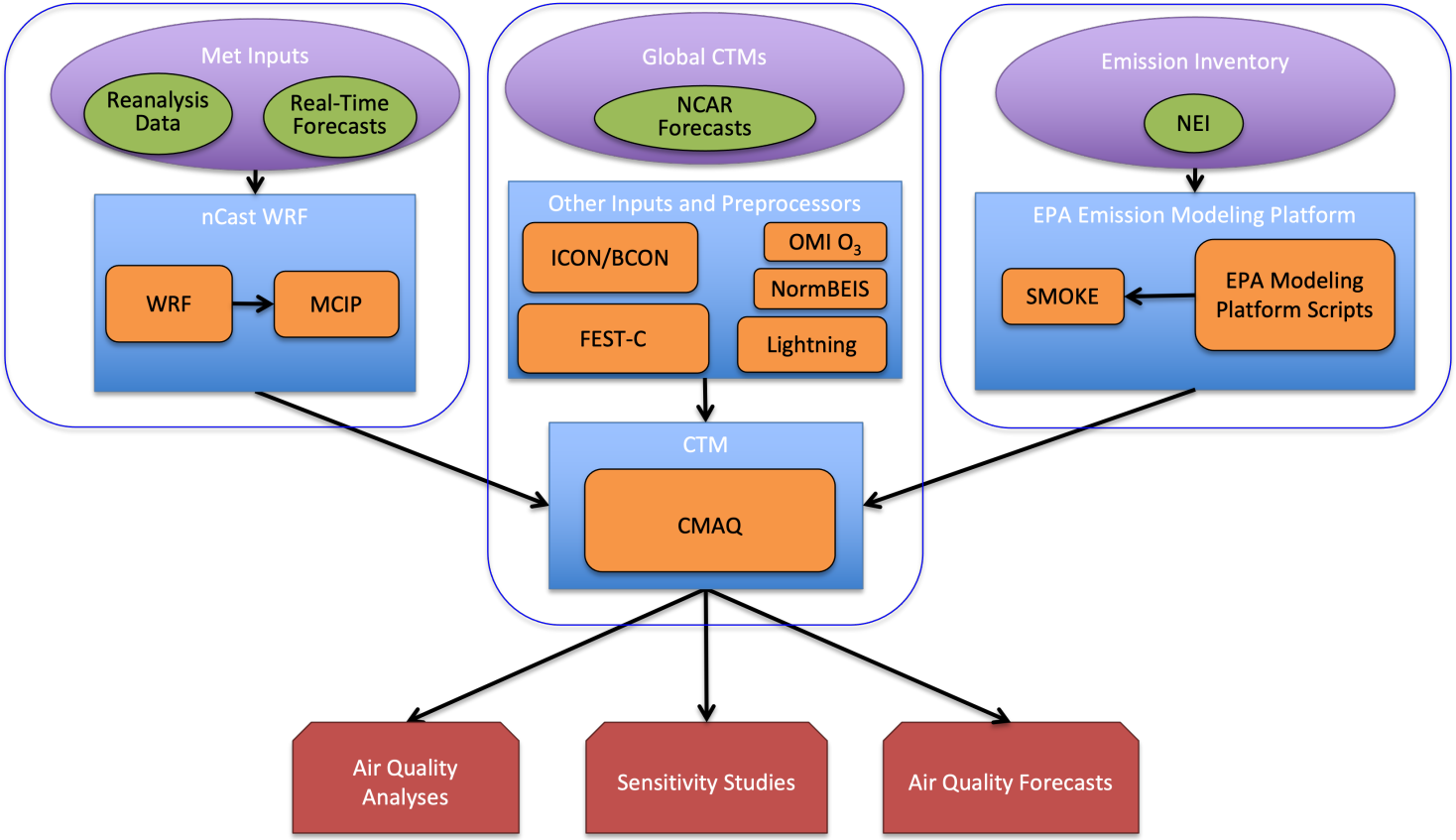
Our History on the AWS Cloud

AQcast: AERMOD and SCICHEM

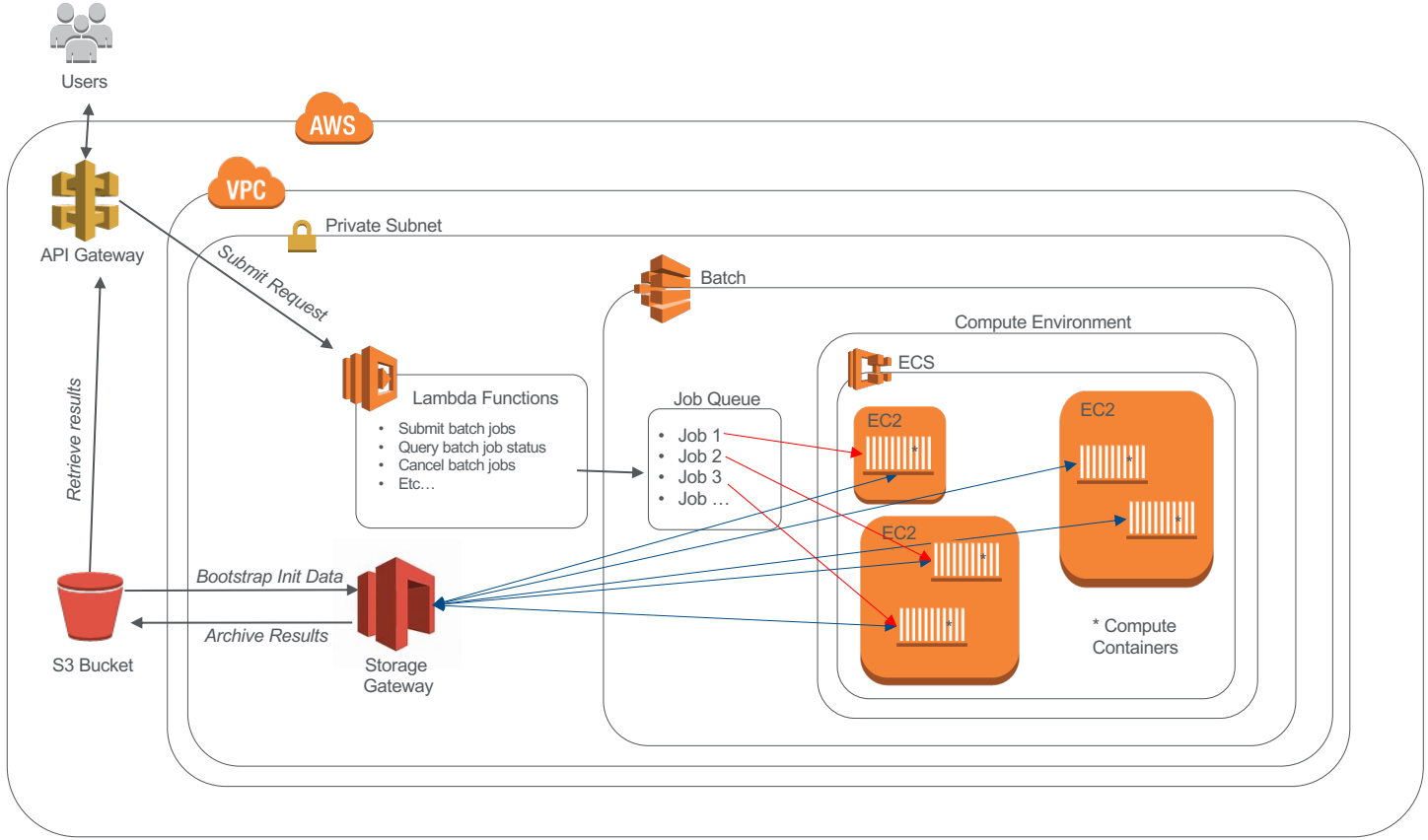




WRF, SMOKE, CMAQ, and CAMx on the cloud at AER

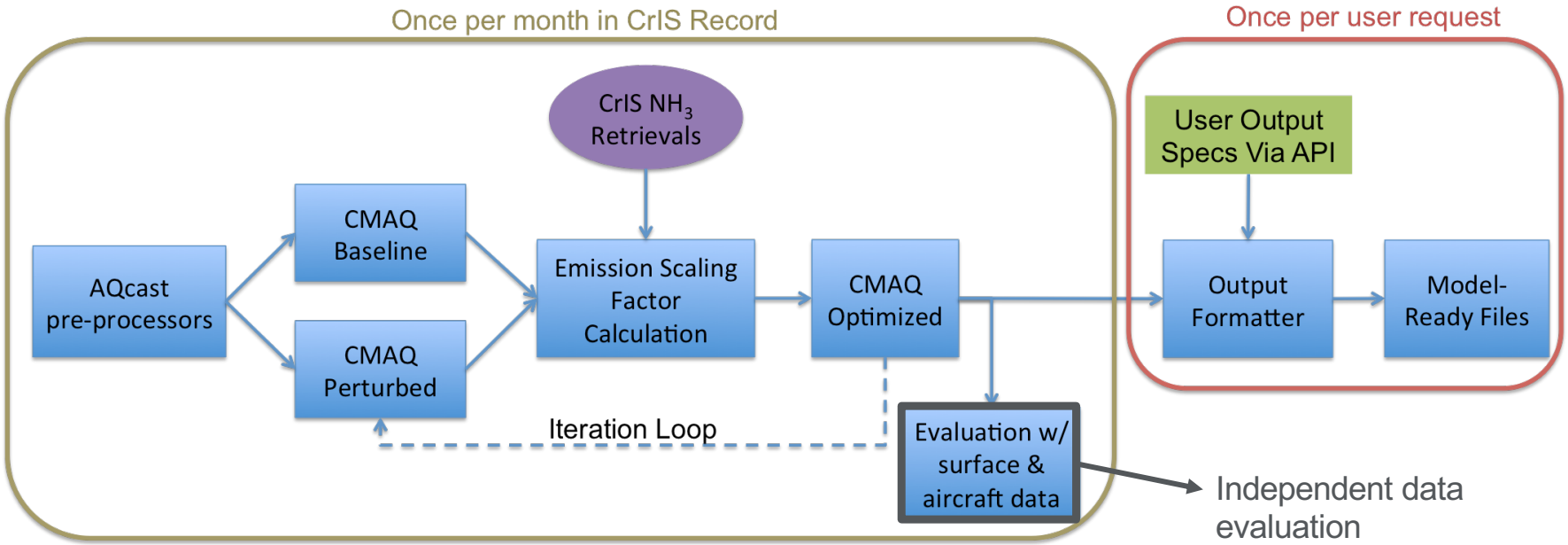


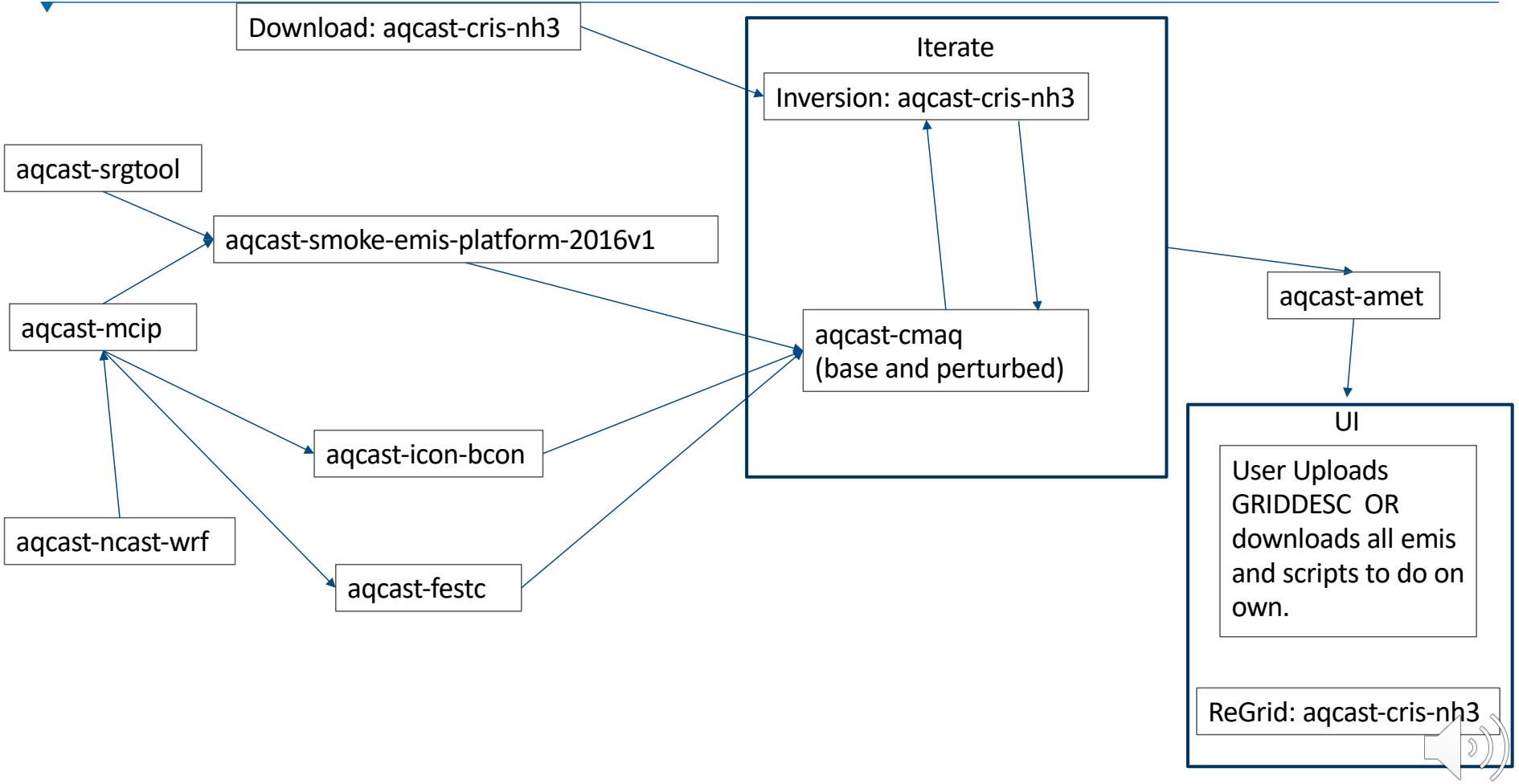
AQcast Production Architecture



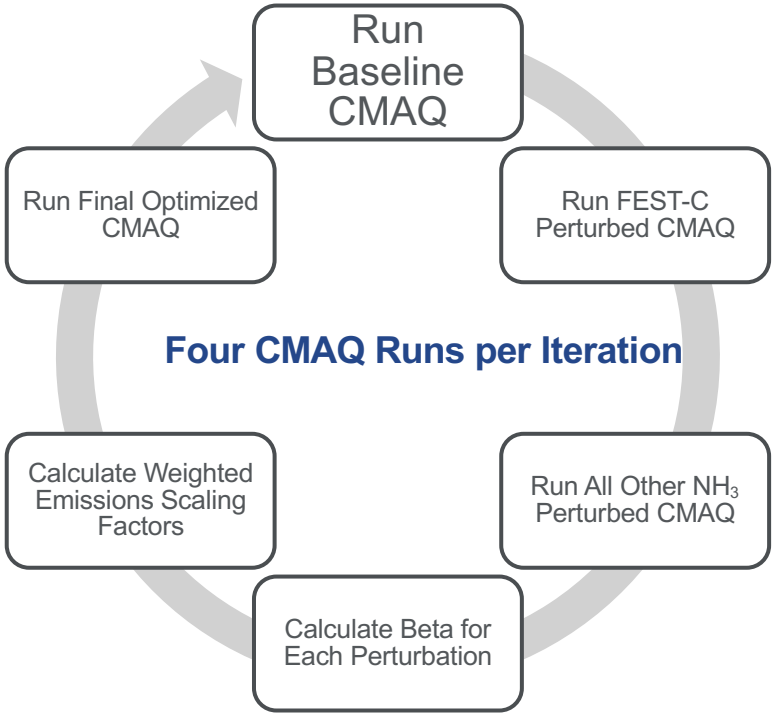
Specific Projects

Inverse modeling of NH_3 using CMAQ



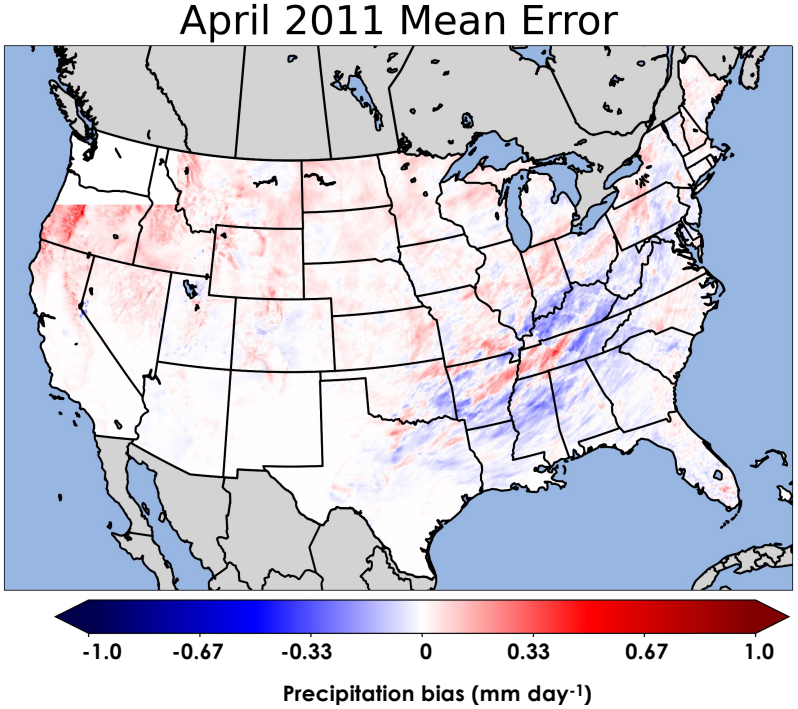


Latest Updates: Calculation of Final NH_3 Emissions using bidirectional flux



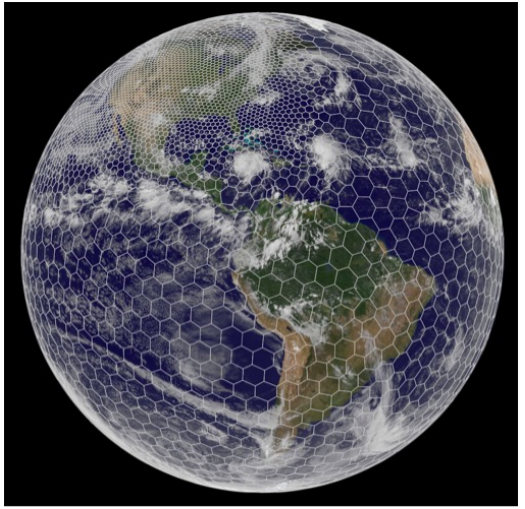
WRF Dynamic Downscaling

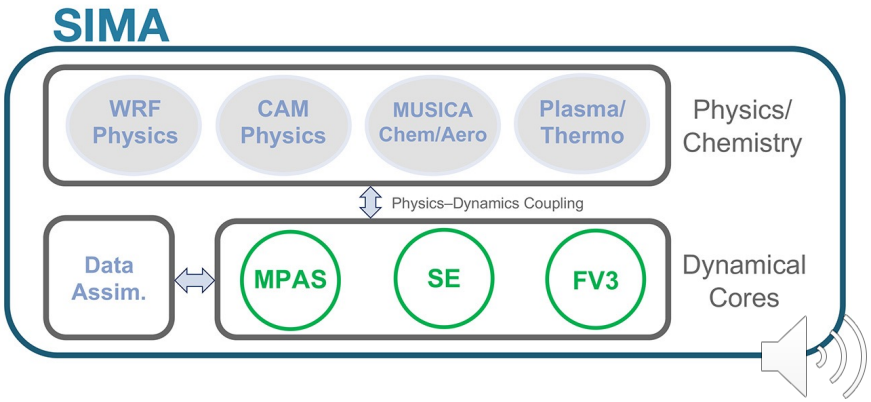
- Goal: downscale current global reanalyses and future climate projections for a wide variety of customers (renewable energy, insurance, etc.)
- Currently running WRF over CONUS at 4 km for 20 years with ERA5 to validate capability
  - 30-hr simulation, 1200 x 800 grid points
- Using AWS Batch, instance c5n.18xlarge
  - Not the fastest, but the most cost–effective
  - Using SPOT pricing to save costs
  - Use FSX for Luster instead of EFS as it's more cost efficient
- WRF and WPS are run as separate Batch jobs
  - Increases efficiency within the compute environments and removes bottlenecks in i/o.
- Additional container to move the data to cheaper s3 bucket before post-processing is done.



MPAS Installation and Testing

- Developed “nCast-Global” to expand our WRF infrastructure to MPAS
- nCast-Global automatically downloads needed input data and simplifies configuration options using a high-level XML control file
- Successfully tested MPAS on a single instance on AWS
- Scaling up has proven challenging
  - ParallelCluster instructions violated company security rules
  - Lots of MPAS development still needed to make it a “user-friendly” as WRF
  - SIMA to the rescue?

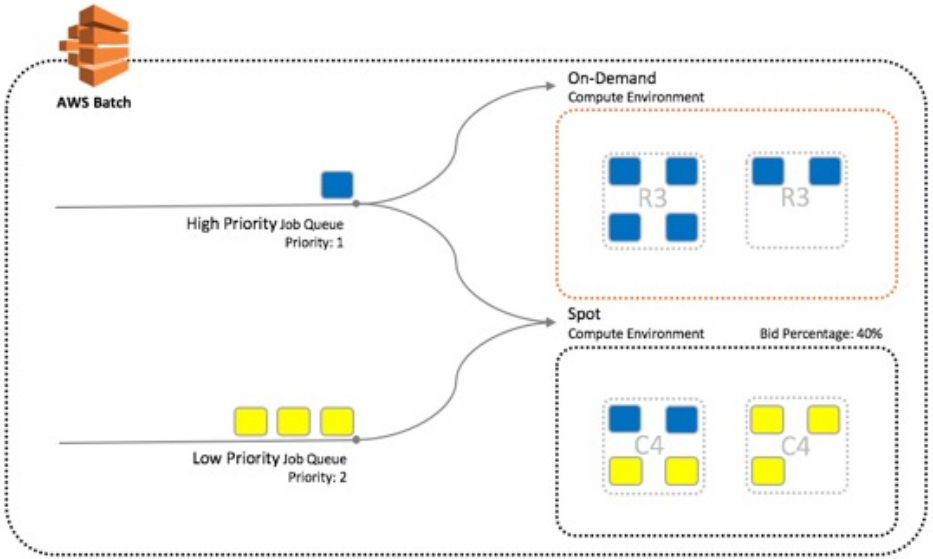




Testing New Techniques

Multi-node Parallelization on AWS Batch

- AWS Batch is a free service that automatically scales your compute resources
  - Removes danger of multiple clusters that need their own set up, security maintenance, etc.
- Configuring Batch to use multiple nodes in parallel (MNP) required additional work
  - Elastic Fabric Adapter (network device for HPC)
  - FSx for Lustre file system
- Rick Pernak built infrastructure to run WRF v4.0.2 on an arbitrarily large number of cores with a simple Python script
  - 30-hour test run over 15 EC2 instances and 1080 vCPUs (processes) in less than 2 hours
- Can now include any other model on AWS that has a Docker image



Graviton Instances

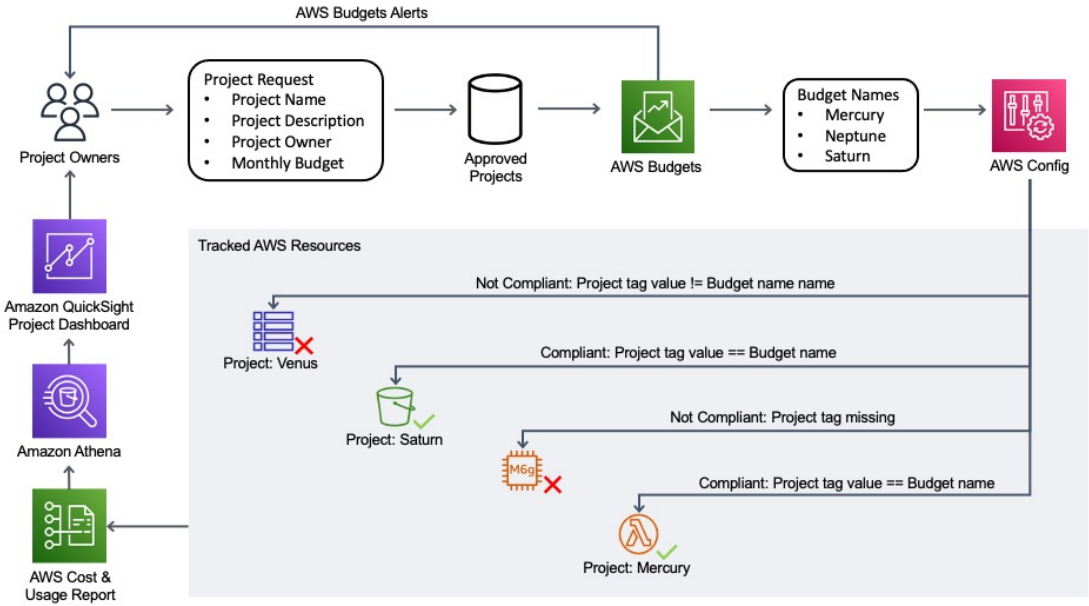
- In principle, Graviton Instances are better for HPC
  - "Up to **40%** better price/performance versus comparable current generation x86-based network optimized instances
- **We compared WRF runs on two instances (with our CONUS 4km downscaling configuration)**
  - Graviton c6gn.16xlarge:
    - 4.32 ¢/core-hour, 64 cores, 4 GiB RAM per core (EBS Only), 100 Gbps bandwidth
  - EC2 Intel c5.18xlarge:
    - 4.25 ¢/core-hour, 72 cores, 4 GiB RAM per core (EBS Only), 25 Gbps bandwidth
- We could not use Intel compiler on Graviton
- **Graviton test run took 16% longer**, most likely due to compiler change (gofrtran) and fewer processors
- Now exploring the use of ARM compilers to improve performance

This is Galvatron, not Graviton, but if you say Galvatron your AWS folks will figure it out.

Administration and Training

Project tagging, cost tracking, and configuration rule enforcement

- **Project tracking solution**
  - A project creation request process.
  - An AWS Budget for each project.
  - An AWS Config Custom Rule that enforces the above
  - An Amazon QuickSight dashboard that uses data from an AWS Cost and Usage Report queried via Amazon Athena to give a per-resource breakdown of cost for each project.
- **AWS Config rules**
  - AWS Config service scans of resources to see if they pass pre-defined rules.
  - Config rules have the option to be auto-remediated to manage security and data safety as projects and access grows

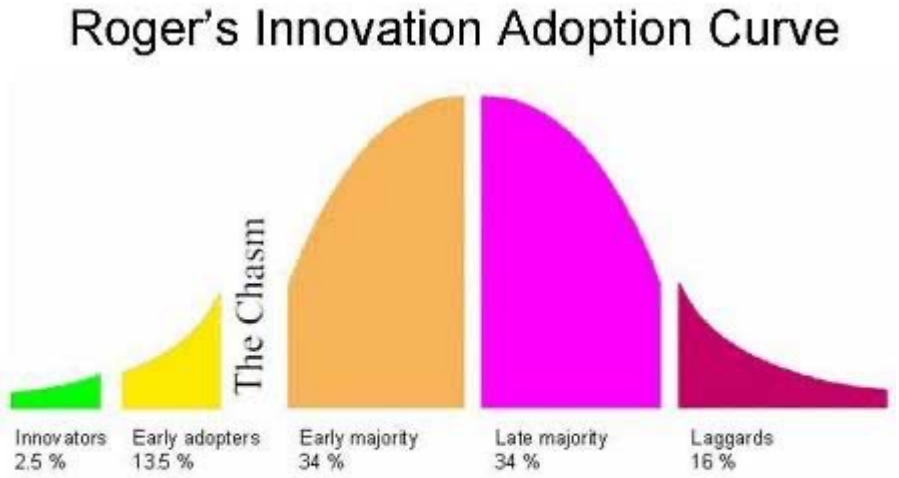


Common processing and development infrastructure

- **Amazon Workspaces are a great way to get comfortable with the cloud**
  - Works like a Linux desktop, but with nigh-infinite storage and compute power
- **Build your infrastructure *once* to accommodate all your users and projects**
  - AQcast was one of the first AER projects to have the batch infrastructure and ci/cd in relation to docker repository and ECR.
  - It has since been generalized into a sandbox where testing and development can be done without every scientist needing to learn how to setup Batch
  - Sandbox environment size is limited to avoid runaway costs, but a duplicate without these limits is easy to create once initial testing is done.
  - Infrastructure is written and deployed in Terraform and kept in a version controlled gitlab project. This makes updates and copying easy.

Getting your *entire* workforce ready for the cloud

- **Supporting Your Innovators**
  - Time and money to try new services and approaches
  - Don't ask for miracles
- **Empowering Early Adopters**
  - Learn@Lunch to share successes and get support
  - Cloud Academy and A Cloud Guru formal training
  - Teams Wiki pages and chats for general questions
- **Addressing Fears of the Majority**
  - Build infrastructure for them
  - Go "back to basics" in Learn@Lunch frequently
  - Have Early Adopters provide support
  - Amy is Amazing!
- **Dragging the Laggards Kicking and Screaming**
  - Mandatory AWS training and
  - Shutting down local resources



aer™

A Verisk Business