NOAA
OMD

OMD OFFICE NOTE 2

https://doi.org/10.25923/jm9h-1e59

Development Towards an Initial Version of the Global Constituent Data Assimilation System (GCDAS)

Andrew Tangborn, Barry Baker, Cory R. Martin,
and Yaping Wang

US DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Weather Service
Office of Modeling and Development
Riverdale Park, MD
July, 2026

Development Towards an Initial Version of the Global Constituent Data Assimilation System (GCDAS)

Andrew Tangborn [1, 2] (https://orcid.org/0000-0003-0617-7387)

Barry Baker [1] (https://orcid.org/0000-0002-6431-2391)

Cory R. Martin [1] (https://orcid.org/0000-0002-1546-9755)

Yaping Wang [1,2] (https://orcid.org/0000-0003-4645-2136)

[1] NOAA National Weather Service, Office of Modeling and Development, Riverdale Park, MD, USA (https://ror.org/01pbt4y78).

[2] SAIC at NOAA NWS OMD, Riverdale Park, MD, USA

Office of Modeling and Development Office Note 2
July, 2026

Abstract

This Office Note documents the configuration and performance of the aerosol data assimilation component of the Global Chemistry and Aerosols Forecast System (GCAFS), the Global Constituent Data Assimilation System (GCDAS). The primary emphasis is on the specification and implementation of background error statistics, covariance modeling, and observation assimilation within the Joint Effort for Data assimilation Integration (JEDI) framework. A brief description of the forecast model configuration is provided for context; however, the focus of this document is on the aerosol data assimilation methodology rather than on the meteorological data assimilation system or the model configuration and coupling.

This page is intentionally left blank

Table of Contents

1. Introduction

As part of the Unified Forecast System (UFS) [Morgan et al. 2024], NOAA is building a coupled Earth system modeling system across spatial and temporal scales which includes atmospheric composition (initially aerosols) components within the atmospheric prediction system. The Global Chemistry and Aerosols Forecast System (GCAFS) is NOAA's operational global aerosol forecasting system developed at the National Weather Service's (NWS) Environmental Modeling Center (EMC). GCAFS provides global analyses and forecasts of atmospheric aerosols in support of air quality monitoring, visibility assessment, and evaluation of aerosol radiative impacts. A key component of GCAFS is the aerosol data assimilation (DA) system, named the Global Constituent Data Assimilation System, or GCDAS, which integrates model forecasts with satellite observations to produce gridded aerosol analyses.

The meteorological component of GCAFS is based on the Finite-Volume Cubed-Sphere (FV3) [Lin and Rood, 1997] dynamical core and is coupled to the Goddard Chemistry Aerosol Radiation and Transport (GOCART) model for tropospheric aerosols [Chin et al., 2002] through the UFS . Due to the comparatively long computational cost of aerosol forecasts, GCAFS is implemented as a separate system, allowing for different model configuration, forecast lead times, and initialization cadences from the medium-range weather model guidance, in order to best serve the needs of all users while balancing computational limitations.

Meteorological analyses are produced operationally using NOAA's Gridpoint Statistical Interpolation (GSI) data assimilation algorithm as part of the Global Forecast System (GFS)'s Global Data Assimilation System (GDAS) [Kleist et al., 2009]. Following completion of the operational meteorological analysis, an offline analysis is performed for GCAFS by interpolating the GDAS analysis to the GCAFS resolution, computing analysis increments as the difference between the regridded analysis and the previous short-term forecast from the cycling GCDAS. These increments are then applied to the next GCDAS and GCAFS forecasts to constrain the meteorology consistently with the GDAS, albeit at a lower resolution.

The aerosol DA system in GCAFS is implemented using the Joint Effort for Data assimilation Integration (JEDI) infrastructure. This is a multi-year effort between NOAA, other US agencies, the UK MetOffice, and led by the Joint Center for Satellite Data Assimilation (JCSDA). The Object-Oriented Prediction System (OOPS) provides the variational assimilation framework and control variable formulation, the Unified Forward Operator (UFO) manages observation operators and quality control, and the System Agnostic Background Error Representation (SABER) supplies background error covariance modeling. The FV3-JEDI interface is used to couple the aerosol forecast model to the DA system, and the Community Radiative Transfer Model (CRTM) is employed to simulate aerosol optical depth (AOD) observations from satellite instruments.

Background error specification is a critical element of aerosol data assimilation due to the strong spatial and temporal variability, and species dependence of aerosol distributions. In GCAFS, background error variances are derived from climatological diagnostics and partitioned by aerosol species. Spatial error correlations are represented using diffusion-based operators with configurable horizontal and vertical length scales. Additional tunable parameters control standard deviations, smoothing, and localization to ensure physically realistic aerosol analysis increments.

The objectives of this Office Note are to (1) document the aerosol data assimilation configuration used in the GCAFS version 1 system, (2) describe the methodology used to estimate and apply background error covariances, (3) summarize the observation datasets and assimilation settings, and (4) evaluate aerosol analysis performance using independent observations and external reanalyses. This documentation is intended to support system evaluation, reproducibility, and future development of the operational global constituent data assimilation capability.

2. Forecast Model

2.1 Model Components

The Unified Forecast System (UFS) [UFS Weather Model Users Guide], a years-long effort by NOAA and its partners, aims to reduce redundancy and share components across applications of various spatial and temporal scales. A fully coupled Earth system framework for prediction, the UFS contains components for the atmosphere, ocean, sea-ice, waves, space weather, atmospheric composition, and more, allowing for configurations of one or more of these components combined together to create specific forecasting systems for a variety of applications. The following subsection will provide an overview of the two UFS components leveraged in GCAFSv1, the atmosphere and atmospheric composition.

The Finite-Volume Cubed-Sphere (FV3) dynamical core [Lin and Rood, 1997] is used as the basis for the atmospheric component of GCAFS coupled to the rest of the UFS through the Earth System Modeling Framework (ESMF) based National Unified Operational Prediction Capability (NUOPC) layer. The Common Community Physics Package (CCPP) is used to include the additional components to make up the atmospheric forecast model including microphysics, gravity wave drag, and the land surface model, with the schemes used listed in the following subsection.

Coupled to FV3 through the NUOPC layer is the second generation of the National Aeronautics and Space Administration (NASA) Goddard Chemistry Aerosol and Radiation Transport (GOCART) model [Collow et al., 2024], which simulates the evolution of atmospheric aerosols through emissions, transport, and removal processes. The model represents multiple aerosol species, including sulfate, black carbon, organic carbon, dust, and sea salt. Aerosol mass mixing ratios are prognostic variables and are transported using the meteorological wind fields provided through the coupling to FV3, ensuring consistency with the meteorological model numerics. Removal processes include dry deposition and wet deposition, with the latter parameterized using the supplied meteorological fields.

For emissions, the FENGSHA dust emissions scheme [Zhang et al., 2022] is used, which emits dust based off of the meteorology and land model parameters in the coupled GCAFS system. Smoke emissions are provided from the Blended Global Biomass Burning Emissions Product [GBBEPx, Zhang et al., 2014]. Anthropogenic emissions are provided through the 2024 release of the Community Emissions Data System (CEDS).

Additional infrastructure leveraged includes the NOAA Emission and Exchange Unified System (NEXUS) to preprocess input emissions prescribed from external datasets which are then applied as source terms during the aerosol forecast integration, the Unified Post Processor (UPP) to produce products for dissemination, and a comprehensive workflow (global-workflow) to facilitate a self-cycled system. The components described above can be run in a number of configurations as part of the global-workflow superstructure, with choices for resolution, physics parameterizations, and more, left to the decision of the user. In the following subsection below, the specific configuration for GCAFS/GCDASv1 is outlined to use as a reference.

2.2 Model Configuration

For the initial version of GCAFS and GCDAS, the model resolution was selected to be C384L127 which is approximately 25 km resolution horizontally and 127 model layers with a model top of around 80km. This matches the horizontal resolution of the previously operational GEFS-Aerosol member and also matches the effective horizontal and vertical resolution of the operational GDAS analysis. Through the CCPP, GCAFS and GCDASv1 use the same physics configuration as targeted for version 17 of GFS and GDAS [Bengtsson and Han, 2024]. This includes the Thompson Aerosol-Aware Cloud Microphysics Scheme, the NoahMP Land Surface Model, and the RRTMG Radiation Scheme. For more information, please reference the GFS_v17_p8_ugwpv1 CCPP suite here: https://dtcenter.ucar.edu/GMTB/v7.0.0/sci_doc/_g_f_s_v17_p8_ugwpv1_page.html.

The GCDAS cycling system runs four times per day with a nine-hour forecast initialized from an analysis valid at 00, 06, 12, and 18 UTC. This analysis includes a 3D-Variational with First Guess at Appropriate Time (3DVarFGAT) for the aerosol fields (more information in the next section) as well as an offline analysis for the atmospheric and surface fields. For this offline analysis, the operational GDAS analysis valid at the appropriate time is regridded from the C768L127 equivalent Gaussian grid to the C384L127 grid. Then the GCDAS six-hour forecast is subtracted from the regridded GDAS analysis and is used to create an analysis increment which is passed to the FV3 model after a warm start. An offline surface analysis is created leveraging inputs preprocessed for the GDAS and includes sea surface temperatures analyzed through the GDAS analysis. At 00 and 12 UTC a 120-hour forecast is initialized from the same analysis as the appropriate GCDAS forecast to create the GCAFS free-forecast that is disseminated to users. Figure 1 shows a schematic of the cycling system.

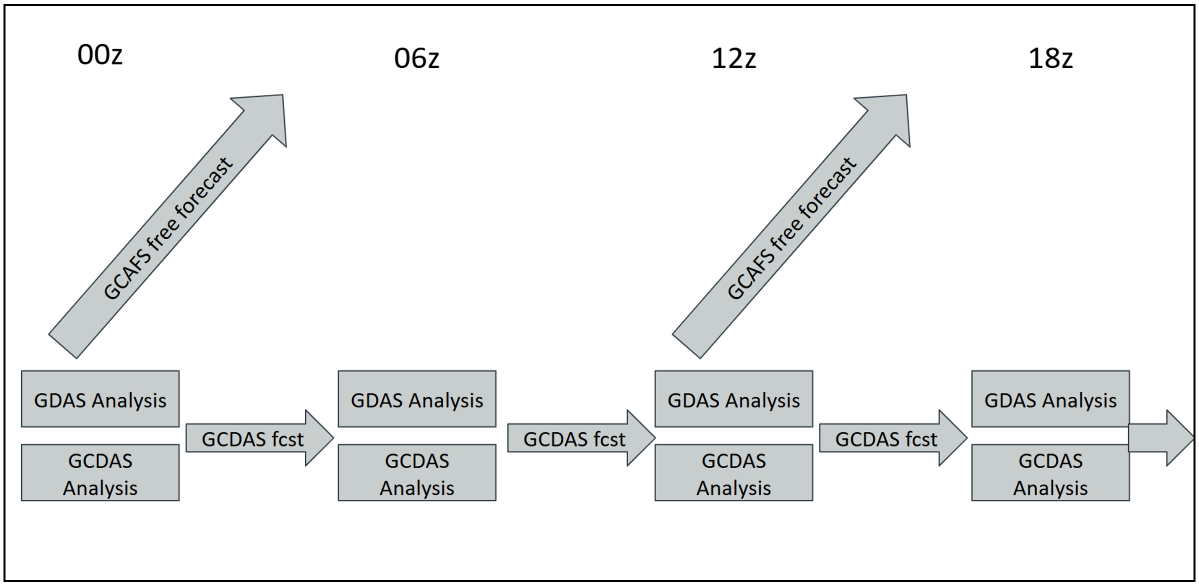


Fig 1: Diagram of cycling configuration for GCDAS and GCAFS.

2.3 Aerosol State Variables

The aerosol data assimilation system operates on aerosol mass mixing ratio fields for each simulated species. These variables are defined on the same horizontal and vertical grid as the meteorological forcing fields. No direct coupling between aerosol and meteorological prognostic variables is applied within the aerosol DA system; meteorological fields are treated as externally specified and are not updated by aerosol fields.

The separation of aerosol and meteorological state variables simplifies the aerosol DA formulation and allows the background error covariance model to be tailored specifically to aerosol processes. This design choice is consistent with the current GCAFS version 1 configuration and provides a flexible framework for future updates.

2.4 Relevance to Data Assimilation

The forecast model configuration described above directly influences the design of the aerosol data assimilation system. The offline meteorological forcing constrains the temporal evolution of aerosols and limits the role of DA increments to modifying aerosol mass fields only. The use of a bulk aerosol model motivates species-dependent background error variances, while the smoothness of the transport operator informs the choice of spatial correlation length scales.

These considerations are reflected in the background error covariance modeling and assimilation configuration described in Sections 3 and 4.

3. Aerosol Data Assimilation Framework

3.1 Overview

The aerosol data assimilation system in GCAFS is implemented using the Joint Effort for Data assimilation Integration (JEDI) framework. JEDI provides a modular, community-developed infrastructure for variational data assimilation, enabling flexible specification of control variables, observation operators, and background error covariance models [Huang et al. 2022]. In GCAFS, JEDI is configured to perform variational aerosol data assimilation using a static background error covariance formulation tailored to aerosol processes.

The aerosol DA system operates independently of the operational meteorological data assimilation, which is performed using NOAA's Gridpoint Statistical Interpolation (GSI) system. Meteorological fields are externally analyzed and are not modified by aerosol analysis increments. The aerosol DA system updates only aerosol mass mixing ratio fields for each simulated species.

3.2 Variational Formulation

Aerosol data assimilation in GCAFS uses the 3-Dimensional Variational-First Guess at AppropriateTime (3DVar-FGAT) DA algorithm. This improves the assimilation accuracy because O-Fs are computed at the correct time through the interpolation of the model variables to the correct observation time. The analysis is obtained by minimizing a cost function of the form

$$J(\mathbf{x}) = 1/2\ (\mathbf{x} - \mathbf{x}b)^{T}\ \mathbf{B}^{-1}\ (\mathbf{x} - \mathbf{x}b) + 1/2\ (y - H(\mathbf{x}))^{T}\ \mathbf{R}^{-1}\ (\mathbf{y} - H(\mathbf{x}))$$

where $\mathbf{x}$ is the aerosol state vector, $\mathbf{x}b$ is the background (forecast) state, y is the observation vector, H is the nonlinear observation operator, $\mathbf{B}$ is the background error covariance matrix, and $\mathbf{R}$ is the observation error covariance matrix.

The control variables consist of aerosol mass mixing ratio increments for each species. The analysis increment is applied to the background aerosol fields to produce the analyzed aerosol state used to initialize the subsequent forecast.

3.3 JEDI Components

The aerosol DA system in GCAFS uses several core JEDI components:

- **Object-Oriented Prediction System (OOPS):**
  Provides the variational minimization algorithm, state vector definition, and incremental formulation.

- **Interface for Observation Data Access (IODA):**
  Provides the infrastructure for input and output of observation space data, metadata, and diagnostics.

- **Unified Forward Operator (UFO):**
  Implements aerosol observation operators, quality control, and observation preprocessing.

- **System Agnostic Background Error Representation (SABER):**
  Supplies the background error covariance model, including variance specification and spatial correlation operators.

- **FV3-JEDI Interface:**
  Connects the aerosol forecast model state to the JEDI framework and handles state transformations and grid mappings.

- **Community Radiative Transfer Model (CRTM):**
  Simulates aerosol optical depth (AOD) from model aerosol fields for comparison with satellite retrievals.

This modular structure allows aerosol-specific configurations to be developed independently of the meteorological DA system.

For more information on the JEDI components and details on how to use JEDI, please see their online documentation at:
https://jointcenterforsatellitedataassimilation-jedi-docs.readthedocs-hosted.com/en/latest/

3.4 VIIRS Observations

The observations assimilated in GCDASv1 are satellite-derived aerosol optical depth (AOD) retrievals from the Visible Infrared Imaging Radiometer Suite (VIIRS) from three polar orbiting satellites (Suomp-NPP, NOAA-20 and NOAA-21) [Zhang et al. 2016]. VIIRS has 16 moderate resolution bands (M1-16), of which M1, 2, 3, 5 and 11 are used for retrieving AOD over land and ocean at 550 nm with a resolution of 750x750 m at nadir and a swath of about 3040 km. These observations provide column-integrated information on aerosol loading and are sensitive to the vertically integrated aerosol mass distribution. VIIRS AOD observation retrievals are provided with uncertainties obtained by comparisons with AERONET observations.

3.4.1 Observation operator

Aerosol lookup tables (LUTs) in the Community Radiative Transfer Model (CRTM) are pre-calculated datasets that store aerosol optical properties—such as extinction coefficients and single scattering albedo, used for efficient radiative transfer simulations. These tables allow the CRTM to simulate the impact of aerosols on satellite radiance measurements by looking up optical properties based on particle size and concentration. This aerosol LUTs approach with the CRTM is exercised in the UFO which connects the GOCART aerosol variables to the CRTM LUTs bins (see more in subsequent sections).

A key part of the UFO is a number of generic quality control (QC) procedures that can be implemented through YAML configuration. For details on all QC filters available, please see the JCSDA JEDI online documentation for UFO Quality Control (https://jointcenterforsatellitedataassimilation-jedi-docs.readthedocs-hosted.com/en/latest/inside/jedi-components/ufo/qcfilters/index.html).

Quality control procedures include screening based on observation uncertainty, spatial representativeness, and consistency with the background state. Observation errors are assumed to be uncorrelated, resulting in a diagonal observation error covariance matrix **R**.

Because AOD observations do not directly constrain the vertical distribution of aerosols, the vertical structure of analysis increments is determined primarily by the background error covariance model. This does lead to an ill-posed problem in the minimization, as a single 2D observation type is used to adjust over a dozen 3D control variables.

3.5 Aerosol control vector

While the O-Fs are computed from the total AOD, the control vector is the individual aerosol species, with units of kg/kg. The GOCART sea salt bins are

- Bin 1: (radius 0.03 - 0.1 um)
- Bin 2: (0.1 - 0.5)
- Bin 3 (0.5 - 1.5)
- Bin 4 (1.5 - 2.0)
- Bin 5 (5.0 - 10.0)

and bins 2-5 correspond most closely to CRTM bins 1-4. The 5 CRTM dust bins match the corresponding gocart dust bins.

The 14 CRTM bins are

- hydrophobic black carbon
- hydrophilic black carbon
- hydrophobic organic carbon
- hydrophilic organic carbon
- sulfate
- dust 001 (radius 0.1 - 1.0 um)
- dust 002 (1.0 - 1.8)
- dust 003 (1.8 - 3.0)
- dust 004 (3.0 - 6.0)
- dust 005 (6.0 - 10.0)
- sea salt 001 (radius 0.1 – 0.5 um)
- sea salt 002 (0.5 – 1.5)
- sea salt 003 (1.5 - 5.0)
- sea salt 004 (5.0 – 10.0 um)

This means that GOCART sea salt Bin 1 is not included in the assimilation. Its contribution to the total AOD is negligible so this has little or no impact on the aerosol forecast.

3.6 Background Error Variance Estimation using local grid points

A good flow dependent error variance estimate is essential to getting the spatial distribution of the aerosol analysis increments. Computational constraints don't allow for the use of an ensemble of aerosol forecasts, so we are using a scheme that estimates the variance from nearby grid points, which is based on the Space Adaptive Forecast error Estimation (SAFE) [Keppene et al., 2014], method to estimate uncertainty from a single aerosol forecast. This approach uses aerosol values at nearby grid points (both vertically and horizontally) to compute variance for each species (Figure 2). We use horizontal and vertical smoothing of the variance so as to avoid small scale variations that could lead to a growing instability in one or more of the aerosol species.

The variance for each aerosol species at node (i,j,k) is computed using

$$Var_{i,j,k} = \ < (X - \ < X >)^2 >$$

Where <> represents the ensemble mean of neighboring grid points and X is an aerosol mass mixing ratio. Note that we are only computing this at the center of the analysis window without incorporating any time varying information in the estimation. Future work may include extending the variance from 3D to 4D to incorporate variability throughout the assimilation window.

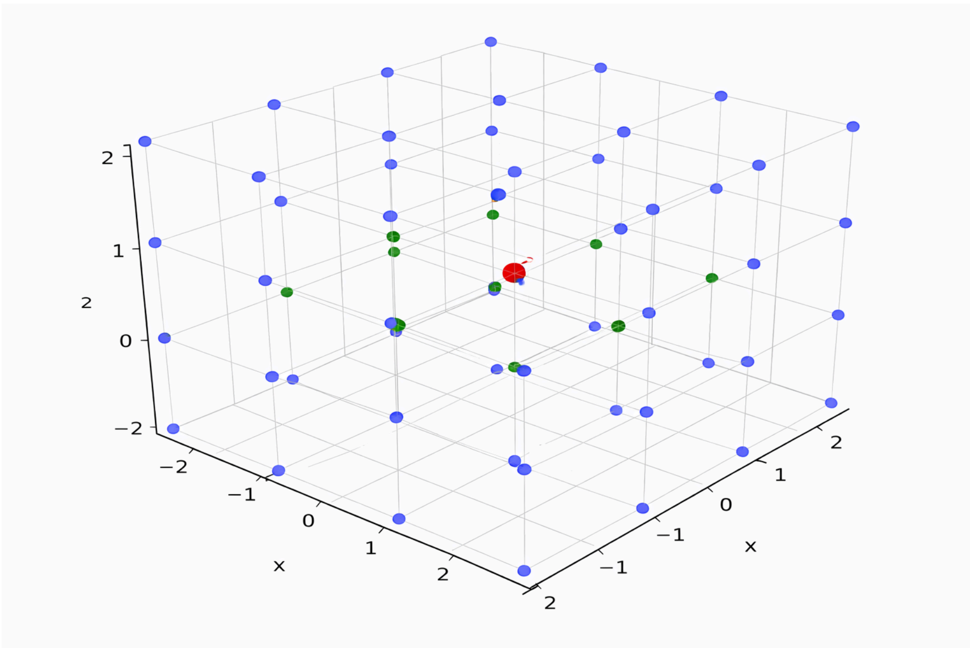


Figure 2: 3D Computational Grid. A representation of the 3D mesh mesh used for error estimation, highlighting the central gridpoint (red) and its neighboring nodes (green) utilized for calculating the local error variance.

The smoothing procedure includes configurable parameters such as:

- Number of neighboring grid points used in averaging
- Horizontal smoothing length scale
- Treatment of grid boundaries and halo points

These parameters are selected to balance noise reduction with preservation of physically meaningful spatial gradients. Excessive smoothing can lead to overly diffuse analysis increments, while insufficient smoothing can introduce unrealistic small-scale variability.

4. Rescaling of Background Error Standard Deviations

As with the ensemble Kalman filter, SAFE estimates of background errors might need to be rescaled (inflated). Reasons for rescaling possibly include increments that are too small (leading to little DA impact) or too large (adding too much noise to the analysis).

These scaling factors can be adjusted by constants globally for each aerosol species by editing the yaml:

global-workflow/parm/gdas/jcb-gdas/algorithm/aero/aero_gen_bmatrix_rescale_default.yaml.j2

And they can be varied spatially by uncommenting the lines:

{% set rescaling_factor_file = aero_rescaling_factor_file|default(aero_gen_bmatrix_rescale_default)%}

{% include rescaling_factor_file %}

In

[global-workflow/parm/gdas/jcb-gdas/algorithm/aero/aero_gen_bmatrix_diagb.yaml.j2]

And then the tile files of the form:
[global-workflow/parm/gdas/jcb-gdas/algorithm/aero/aero_gen_bmatrix_diagb.yaml.j2]

rescale.fv_tracer.res.tile1.nc, etc need to be changed, which are stored in global-workflow/fix/gdas/aero/clim_b.

Testing of the forecast system has shown that rescaling values of 1.0, or no rescaling, is close to optimal for all of the species. This may change as the model and assimilation system are refined in the future.

5. Spatial Error Correlation Modeling Using Diffusion Operators

5.1 Overview

In the GCAFS aerosol data assimilation system, spatial correlations in background error are modeled using diffusion operators implemented within the SABER component of the JEDI system. This approach provides a flexible method for representing horizontal and vertical error correlations for aerosol variables, while avoiding the explicit construction of large covariance matrices.

The diffusion method defines spatial correlations implicitly by applying smoothing operators to the background aerosol fields. When combined with the background error standard deviation fields described in Section 4, the diffusion operators determine the spatial extent and structure of the aerosol analysis increments.

5.2 Diffusion-Based Correlation Formulation

The diffusion-based correlation model represents spatial correlations by solving a discrete diffusion equation applied to the control variable increments [Derber and Rosati, 1989], [Egbert, et. al., 1994], . Repeated application of the diffusion operator results in spatial spreading of information from observation locations, with the effective correlation length determined by user-specified diffusion parameters.

In this formulation, the background error covariance matrix is factorized into an amplitude component, provided by the standard deviation fields, and a correlation component, represented by the diffusion operator. The resulting covariance model is positive definite by construction and is well suited for large-scale operational applications.

5.3 Horizontal and Vertical Correlations

Separate diffusion operators are applied in the horizontal and vertical directions. Horizontal diffusion controls the lateral spreading of analysis increments and is characterized by user-defined horizontal length scales. These length scales represent the typical spatial extent over which aerosol forecast errors are correlated and are chosen based on physical considerations such as aerosol lifetime and transport characteristics.

Vertical diffusion controls the coupling of analysis increments between model levels. Because aerosol optical depth observations provide primarily column-integrated information, error correlations in the vertical play a critical role in distributing analysis increments throughout the column. These diffusion parameters are selected to produce smooth, physically realistic structures in height without introducing excessive spreading.

The error correlation is computed from the solution to the diffusion equation on the cubed sphere grid so that the covariance can be modeled as the solution C to

$$(I - \ell^2 \nabla^2)\, C = X'$$

Where ℓ is the imposed correlation length scale, X’ is the three dimensional aerosol field (minus the mean) and η is a generated white noise field.

5.4 Species-Dependent Configuration

Diffusion-based error correlations are configured independently for each aerosol species. This allows the spatial correlation length scales to differ among species, reflecting differences in source mechanisms, transport behavior, and atmospheric residence times. For example, coarse-mode aerosols such as dust and sea salt typically exhibit larger horizontal correlation scales than fine-mode aerosols such as sulfate.

No explicit cross-species error correlations are included in the current GCAFS configuration. Correlations are applied only within each aerosol species. Future work may include not only cross-species correlations between aerosol fields but also correlations with meteorological variables.

5.5 Numerical Implementation and Parameters

The diffusion operators are implemented on the same grid as the aerosol forecast model and use configurable numerical parameters, including:

- Number of diffusion iterations (200)
- Horizontal and vertical diffusion coefficients
- Treatment of domain boundaries and halo points

These parameters are specified through external YAML configuration files, including the aero_diffusionparm.yaml.j2 template. This design allows correlation length scales and numerical behavior to be adjusted without modifying source code and facilitates sensitivity testing during system development.

5.6 Hybrid Covariance

GCAFS currently has the ability to make use of a static error variance estimate, but this is currently turned off by setting its weight to zero. Further work, including constructing an improved static background covariance model, is needed to make this feature fully functional. The static B matrix that was initially evaluated was constructed using the Background error on Unstructured Mesh Package (BUMP), which is a part of SABER, one of the components of the JEDI data assimilation system [Jung, et al. 2024].

5.7 Practical Considerations

The choice of diffusion parameters represents a balance between physical realism and computational efficiency. Excessively large correlation length scales can lead to overly smooth analysis increments and reduced sensitivity to localized observations, while overly short length scales can result in noisy analyses. In addition, longer correlation lengths increase the computational time needed for the diffusion operator. Parameter values used in GCAFS v.1 were selected based on empirical testing and evaluation against independent observations.

6 Generating background errors for aerosol DA

Task **gcdas_aeroanlgenb** contains:

aero_interpbkg Interpolate background to analysis resolution (skipped in GCAFSv1 because the resolutions are equal).
global-workflow/sorc/gdas.cd/sorc/da-utils/src/regridStates

aero_diagb Compute background error variance using the SAFE algorithm:
global-workflow/sorc/gdas.cd/utils/chem

aero_diffusion Compute background error correlation by solving diffusion equation:
global-workflow/sorc/gdas.cd/sorc/saber/src/saber/diffusion

The tasks are found in

global-workflow/ush/python/pygfs/task/aero_bmatrix.py

6.1 Configure Files for variance and diffusion correlation settings

The file config.aeroanlgenb contains the values of:

Aero_diagb_n_neighbors: Number of neighbors used in calculating stddev from the background

Aero_diagb_n_halo: Number of halo points (grid points outside the processor grid points) used in the stddev calculation.

Aero_diagb_smooth_horiz_iter: Number of horizontal smoothing iterations

Aero_diagb_smooth_vert_iter: Number of vertical smoothing iterations

Aero_diffusion_iter: number of diffusion iterations. A larger number results in max correlation closer to 1.0, but requires more computational time.

Aero_diffusion_horiz_len: horizontal correlation length scale in km.

Aero_diffusion_fixed_val: vertical correlation length scale in number of model levels.

Aero_diagb_weight: A hybrid approach (not currently used) is possible when this is less than 1.0.

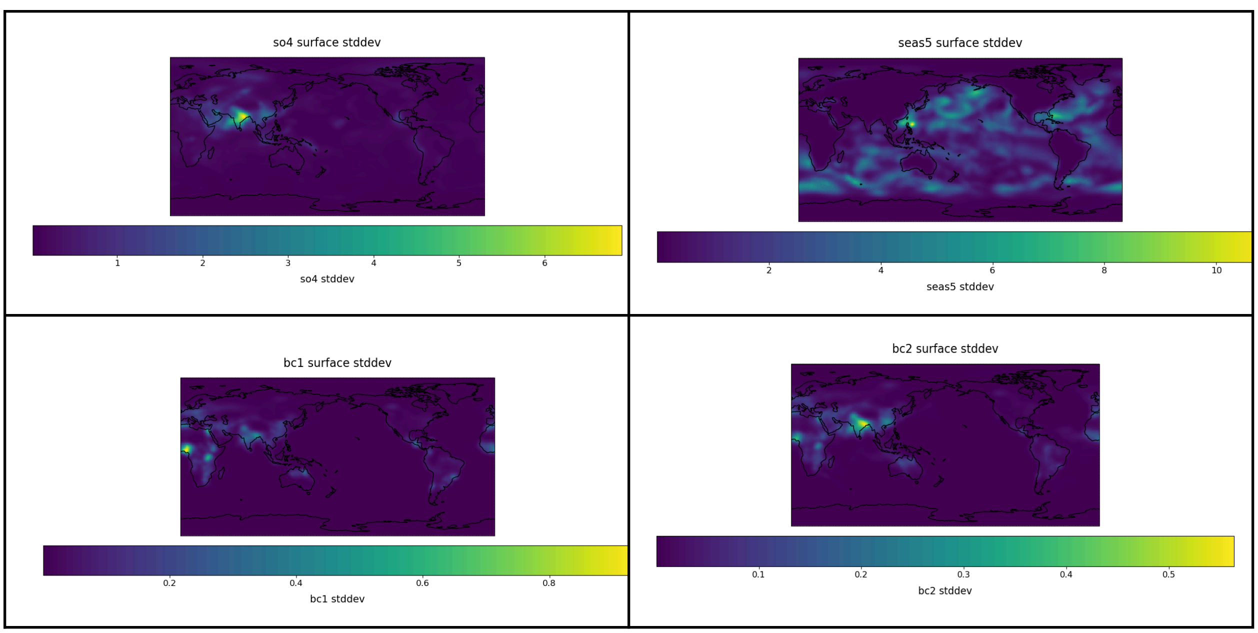
Aero_staticb_rescaling_factor: Scaling factor for the staticb background values. Only used when Aero_diagb_weight is less than 1.0.

The species specific rescaling values are stored in netcdf files:

{{ BERROR_DATA_DIR }}/rescale.fv_tracer.res.tile1.nc, etc

6.2 Examples of aerosol background error standard deviations

Some of the aerosol species generated by gcdas_aeroanlgenbbe gcdas_aeroanlgenb are shown in Figure 3, for the lowest layer of the atmosphere.



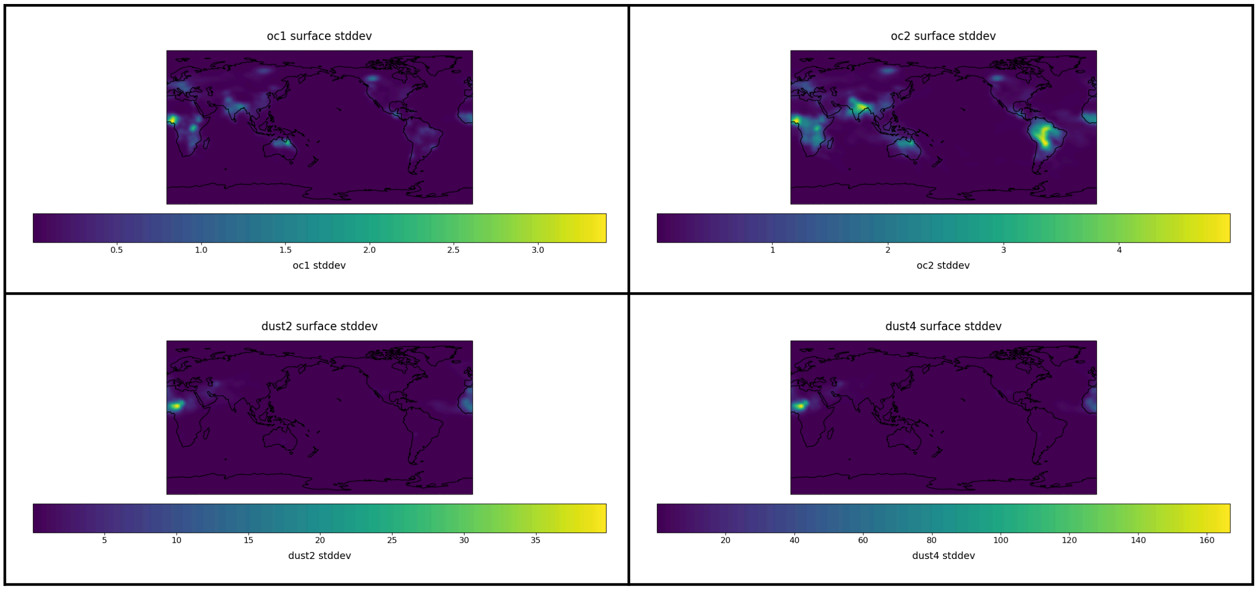


Figure 3. Examples of background Stddev for several aerosol species, starting with top left (SO4, Seas2, BC1, BC2, OC1, OC2, Dust2 and Dust4) at the lowest layer of the atmosphere.

7 Running Aerosol DA

7.1 DA settings

Observations:

- High-quality obs within [-3h, +3h] time window thinned or superobbed approximately to C384 resolution.
- Domain check: Limit obs to 60°S–60°N latitude range.
- Bounds Checks: obs values are within [0, 5].
- Background Checks: Inflate observation error by 3 times if the departure between observation and background guess is larger than 3 times the size observation error.

Minimization:

- nouter = 1 (one outer loop).
- ninner = 35 (35 inner loops)

7.2 Prep, initialize, run and finalize aerosol 3dvar

Tasks:

- gcdas_prep — gets obs from obsforge
- gcdas_aeroanlinit — prepares aeroanl rundir and copies files
- gcdas_aeroanlvar — runs aero DA
- gcdas_aeroanlfinal — copies analysis files from rundir after adding increment to restart.

8. Cycling and Validation

8.1 Operational parameters and settings

Model resolution is C384 with 127 vertical levels. VIIRS AOD retrievals from NOAA-NPP, N20 and N21 are assimilated every 6 hours into GCDAS to produce the aerosol analysis at 00, 06, 12 and 18Z.

The GCAFS 5-day forecasts are launched at 00 and 12Z, as shown in Figure 1.

8.2 Cycling and Tuning Experiments

Cycling and tuning experiments involve comparisons with reanalyses (Copernicus Atmosphere Monitoring Service or CAMS [Blake et al., 2025] and the Modern Era Retrospective Reanalysis 2 or MERRA2 [Buchard et al., 2017], satellite obs (VIIRS, MODIS) and ground based data (AERONET [Holben, 1998]). A cycling experiment with a cold start on 15 December 2023 is running through 31 December 2024 for analysis. While the comparisons with observation are more important for validation, only the reanalyses contain information on individual aerosol species and therefore are critical to understand the sources of error in the total AOD forecasts.

8.2.1 Tuning of background errors using comparisons with CAMS and MERRA2.

Tuning of the background error covariances is done for individual species. And since only CAMS and MERRA2 have global estimates of aerosol species, they are the best data sets for comparing with the aerosol forecasts. Reanalyses contain information from both the model and observations, weighted by their relative uncertainty. The tuning is therefore not a substitute for validation using independent observations.

The CAMS aerosol reanalysis is produced by the European Centre for Medium Range Weather Forecasts (ECMWF). It provides global, multi-decadal atmospheric composition fields (aerosols + reactive gases) using a coupled chemistry-aerosol forecast + data assimilation system.

The CAMS reanalysis is based on the ECMWF Integrated Forecast System (IFS) with interactive chemistry and aerosol modules. It uses harmonized global emissions inventories using consistent and comparable methodologies. These are Anthropogenic (e.g., EDGAR-type inventories), Biomass burning (GFAS fire emissions), Biogenic emissions, Sea salt (wind driven), Dust (wind + soil source parameterizations). CAMS assimilates AOD observations from

MODIS (Moderate Resolution Imaging Spectroradiometer) on Terra/Aqua, PMAp (Metop-B) and VIIRS (NPP, N20 and N21).

The following plots show the weekly averaged RMS differences between 6-hour GCDAS aerosol forecasts and CAMS and MERRA2 reanalyses to demonstrate how the tuning process is done. Figure 4 shows two experiments in which the background error rescaling coefficients are rs=1 for all species except for sulfate which uses either rs=1 or rs=0.1.

The plots below for sulfate, black carbon (hydrophilic + hydrophobic), organic carbon (hydrophilic + hydrophobic) , dust (bins 1-5), sea salt (bins 1-5) and total AOD demonstrate how the species interact through the background errors. In each plot the blue lines are the RMS differences between the 6-hr forecast and CAMS (solid line) for sulfate rs=1, and MERRA2 (dashed line). The red lines are the differences when sulfate rs=0.1. The sulfate differences with the reanalyses are somewhat higher when sulfate rs=1 compared to rs=0.1. The change in sulfate rs also affects the OC comparisons, and the sulfate rs=1 is slightly lower than rs=0.1 towards the end of the experiment. This is because sulfate and organic carbon have similar sources from biomass burning and tend to overlap more than other species. On the other hand, the dust and sea salt comparisons with reanalyses are largely unaffected by the change in sulfate rescaling, due to the different origin and location of these aerosol species. The total AOD is also affected, and we see a slight reduction in the RMS differences with sulfate rs=1.

Data assimilation traditionally cannot improve forecast bias since one of the basic assumptions in all DA schemes is that forecasts are unbiased. This can be seen here as well in the mean differences with CAMS and MERRA2 in Figure 5. Generally the biases are much smaller than the RMS differences, and while not all of the species show mean differences that are reduced by the DA, the total AOD bias is reduced.

Both the background errors from SAFE and the observation errors tend to assign higher uncertainty to larger AOD values. This means that smaller AOD values carry greater weight in the data assimilation and therefore larger increments when the observed AOD is smaller than forecast AOD. This makes low bias harder to overcome and ultimately a somewhat low bias in the analysis fields. This could be overcome by carefully imposing a floor on background error, though it can create its own complications.

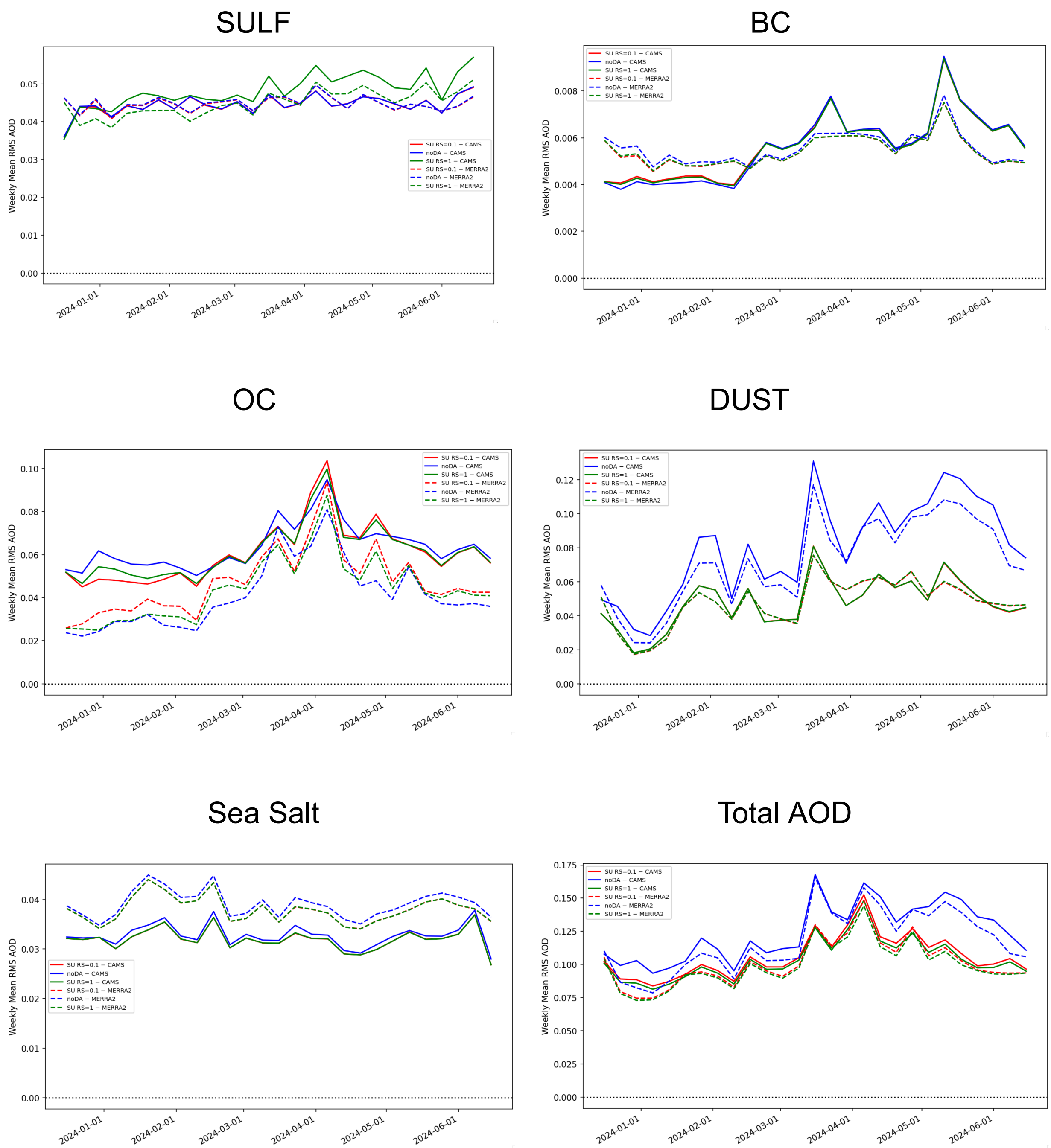
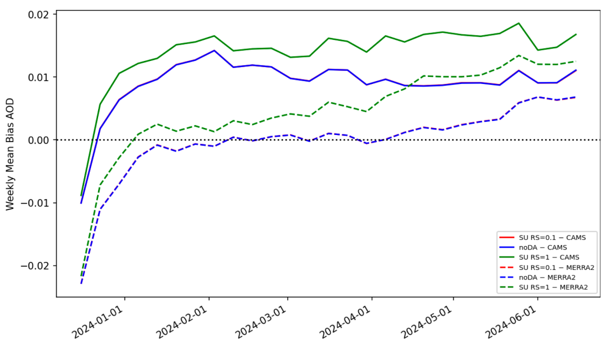
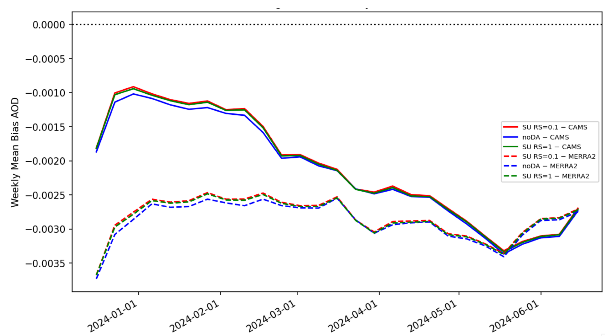


Figure 4 Weekly mean RMS differences between 6-hour GCAFS forecasts and CAMS (solid line) and MERRA2 (dashed line) for no DA (blue), DA with Sulf rescaling rs=0.1 (red) and rs=1.0 (green) for each of the aggregated species and total AOD.

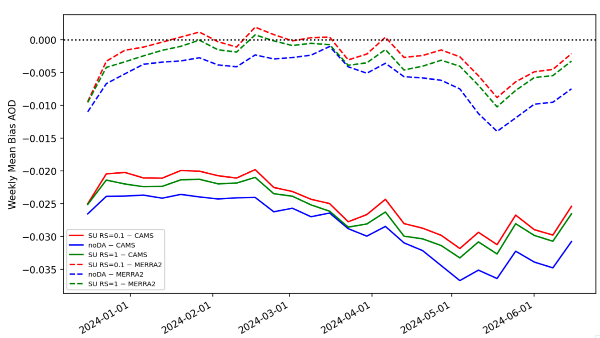
SULF



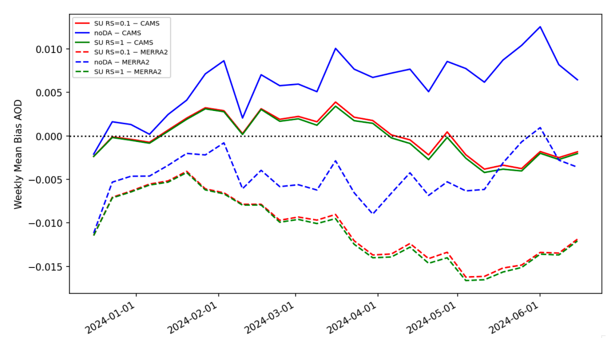
BC



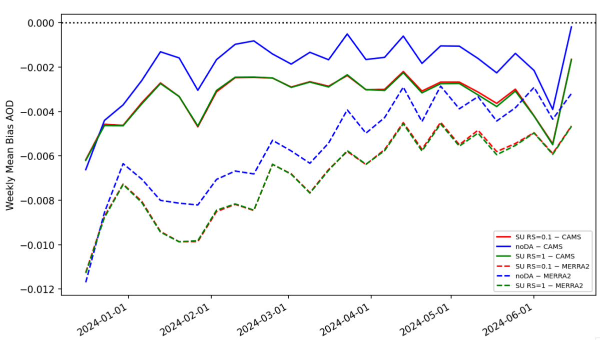
OC



DUST



Sea Salt



Total AOD

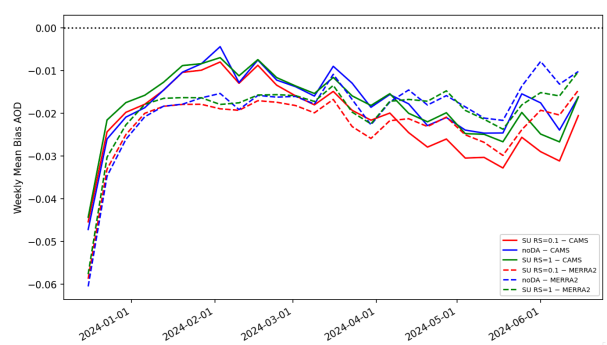


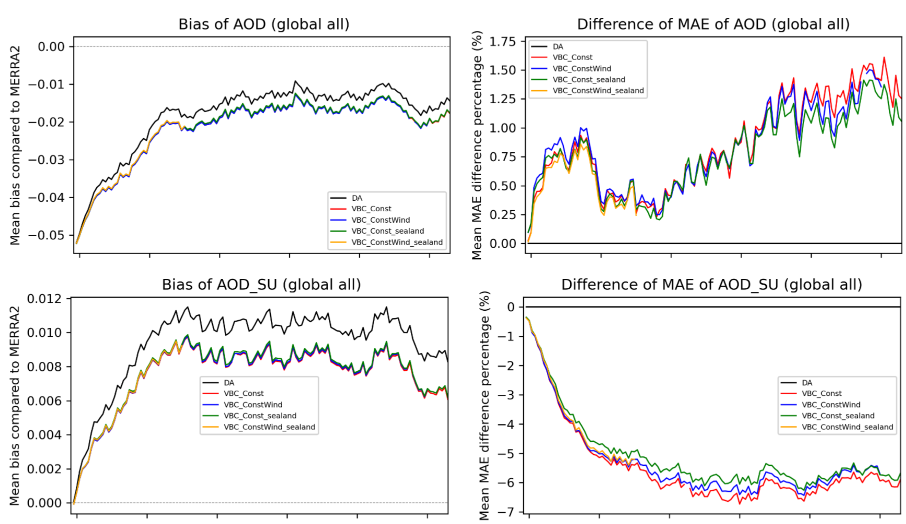
Figure 5 GCAFS forecast weekly mean differences with CAMS (solid line) and MERRA2 (dashed line), for no DA (blue), DA with Sulf rescaling rs=0.1 (red) and DA with Sulf rescaling rs=1.0 (blue).

8.2.2 Comparisons with observations

Extensive validation through comparisons with observational data has been done, including AERONET and MODIS, in addition to the VIIRS O-F statistics. The details of these are included in a manuscript (Wang et al) to be submitted.

8.3 Variational bias correction

The GCAFS framework supports Variational Bias Correction (VarBC), a standard method for mitigating systematic observation errors. Given the mild positive bias observed in VIIRS NPP data relative to VIIRS N20 and N21, we evaluated the impact of VarBC with several bias predictors—including a land/sea sensitive constant term, surface wind speed, and their combinations. After implementing the VarBC procedure to the NPP observations, results indicate a neutral impact on total AOD and individual aerosol species at the global scale (Figure 6). An exception was observed for sulfate, which showed a 5% error reduction. Due to these subtle impacts, and the non-trivial effort needed to support updating the coefficients in the operational workflow, VarBC is not enabled by default in the current version of GCAFS.



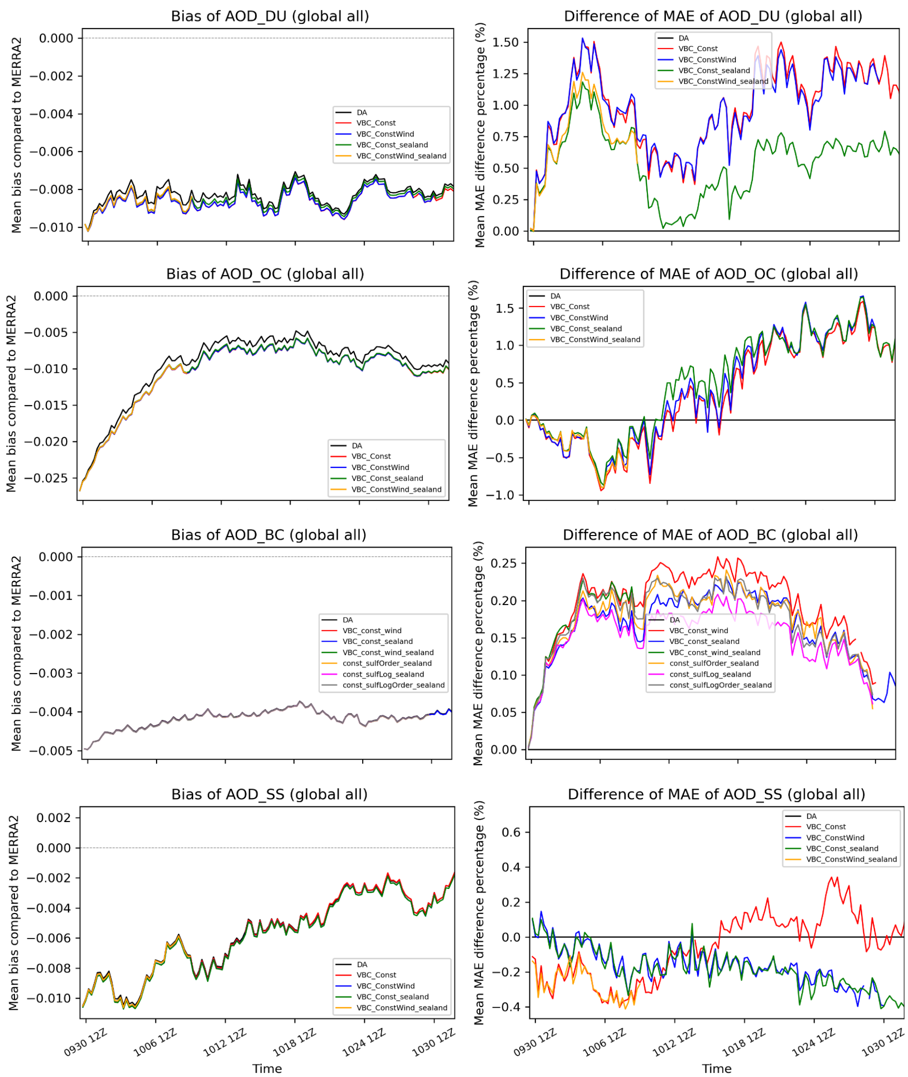


Figure 6. Global mean bias and mean absolute error (MAE) of 6-hour total AOD and individual aerosol species forecasts. The MAE is expressed as a relative change between the VarBC and noVarBC experiments. Results are validated against MERRA-2 reanalysis data.

9. Future Plans

GCAFS v.1 is the first step in building a general air quality forecast system at NOAA. The initial implementation here focuses on just aerosols but the longer term plans include a comprehensive aerosol and chemistry model with assimilation of a number of observations. The next steps include developing medium-complexity gas-phase chemistry and assimilation of in situ particulate matter, geostationary AOD, and trace gas retrievals from polar orbiting satellites.

Eventually, the high-resolution regional Air Quality Model (AQM) will be merged with GCAFS and a multi-scale, varying-complexity constituent prediction system will be made operational. .

The minimization currently uses a single outer loop, as existing limitations within JEDI do not properly update the backgrounds for subsequent loops when using 3DVar-FGAT. Once this capability is incorporated, using multiple outer loops should enable further improvements to the analysis.

While the varBC scheme showed little impact in the current version of GCAFS, it’s possible that this could change with some of the improvements proposed in this section. We will revisit this scheme in the future.

10. Acknowledgements

We have several entities and people that we would like to acknowledge for support of this work. First and foremost, a majority of the development and evaluation work was performed on the NOAA Research and Development HPC System (RDHPCS). Much of this work was funded through the NOAA Weather Program Office Joint Technology Transfer Initiative (JTTI) with the remainder coming from the UFS R2O program. Cycling system workflow and library infrastructure is essential to the success of this system so we would like to acknowledge several colleagues including (but not limited to) Travis Elless, Kate Friedman, David Huber, Walter Kolczynski, Hang Lei, Rahul Mahajan, and Alex Richert. Prior groundwork from Bo Huang, Sarah Lu, Mariusz Pagowski and Shih-Wei Wei helped contribute to the success of this initial operational system. The JEDI framework is at the core of this assimilation system and thus we would like to acknowledge everyone who has contributed to JEDI, and while this is a large group, in particular we would like to acknowledge current and former JCSDA core team members including Maryam Abdi-Oskouei, Jerome Barre, Cheng Dang, Francois Hebert, Steve Herbener, Ben Johnson, Ben Ruston, Travis Sluka, Greg Thompson, Yannick Tremolet, and Steve Vahl. We would like to acknowledge all of the people who have worked on the UFS model, with a particular focus on those involved in the implementation and evaluation of GOCART coupled to FV3, which include (but are not limited to) Alicia Bentley, Partha Bhattacharjee, Anning Cheng, Bing Fu, Ho-Chun Huang, Jeff McQueen, Wen Meng, Raffaele Montuoro, Li Pan, Fanglin Yang, and Kate Zhang. Thank you to members of NWS Central Operations (NCO) to assist in the data ingest, dissemination, and implementation of this system including Justin Cooke, Margaret Curtis, Carlos Dias, and Steven Earle. Finally, we would like to acknowledge the entire EMC DA group, and in particular Hyundeok Choi, Iliana Genkova, Dan Holdaway, David New, Anna Shlyaeva, Russ Treadon, and Guillaume Vernieres for their help, support, and contributions to this system.

11. References

Bengtsson, L., and J. Han, 2024: Updates to NOAA's Unified Forecast System's Cumulus Convection Parameterization Scheme between GFSv16 and GFSv17. Wea. Forecasting, 39, 1559–1570, https://doi.org/10.1175/WAF-D-23-0232.1.

L. Blake, A. Arola, A. Benedictow, Y. Bennouna, I. Bouarar. Cuevas, Q. Errera, H.J. Eskes, J. Griesfeller, L. Ilic, J. Kapsomenakis, B. Langerock, C. W. Y. Li, E, A. Mortier, I. Pison, M. Pitkänen, A. Richter, A. Schoenhardt, M. Schulz, J. Tarniewicz, A. Tsikerdekis, T. Warneke, C. Zerefos, Validation report for the CAMS global reanalyses of aerosol and reactive trace gases: 2003-2024, Copernicus Atmosphere Monitoring Service (CAMS) report, CAMS2_82_2023SC2_D82.4.2.1-2025.pdf, April 2025, doi:10.24380/vv0t-8tcg

Buchard, V., C.A. Randles, A.M. Da Silva, A. Darmenov, P.R. Colarco, R. Govindaraju, R. Ferrare, J. Hair, A.J. Beyersdorf, L.D. Ziemva, an d H. Yu, The MERRA-2 Aerosol Reanalysis, 1980−Onward. Part II: Evaluation and Case Studies," Journal of Climate, 30(17), 6851−6872.

Collow, A. B., Colarco, P. R., da Silva, A. M., Buchard, V., Bian, H., Chin, M., Das, S., Govindaraju, R., Kim, D., and Aquila, V.: Benchmarking GOCART-2G in the Goddard Earth Observing System (GEOS), Geosci. Model Dev., 17, 1443–1468, https://doi.org/10.5194/gmd-17-1443-2024, 2024.

Derber, J. and A. Rosati, A Global Oceanic Data Assimilation System, J. Phys. Ocean., V.19, N.9, 1333–1347, 1989.

Egbert, G. D., Bennett, A. F. and Foreman, M. G. G.,: Topex/Poseidon tides estimated using a global inverse model. J. Geophys. Res., 99, 24,821–24,852, 1994.

Holben, B.,T, F. Eck, I al Slutsker, D. Tanre, J.P. Buis, A. Setzer, E. Vermote, J.A. Reagan, Y.J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, A., Smirnov: AERONET—A federated instrument network and data archive for aerosol characterization. Remote Sens.Environ., 66, 1–16, doi:10.1016/S0034-4257(98)00031-5., 1998

Huang, B., M.Pagowski, S. Trahan, C.R. Martin, A. Tangborn, S. Kondragunta, D.T. Kleist, JEDI-based three-dimensional Ensemble-Variational Data Assimilation System for global aerosol forecasting at NCEP, J. Adv. Mod. Earth Sys., Volume 15, 2022.

Jung, B.J., B. Ménétrier, C. Snyder, Z. Liu, J.J. Guerrette, J. Ban, I. Hernández Baños, Y.G. Yu, and W.C. Skamarock, Three-dimensional variational assimilation with a multivariate background error covariance for the Model for Prediction Across Scales – Atmosphere with the Joint Effort for Data assimilation Integration (JEDI-MPAS 2.0.0-beta), Volume 17 Geooscientific Model Development, 17, 2024.

Keppenne, C.L., Rienecker, M.M. , Kovach, R.M., and Vernieres, G.,
Background Error Covariance Estimation using Information from a Single Model Trajectory with Application to Ocean Data Assimilation into the GEOS-5 Coupled Model, NASA Technical Report Series on Global Modeling and Data Assimilation,
Volume 34, 2014.

Kleist, D., J.R. Carley, A. Collard, E. Liu, S. Liu, C.R. Martin, C. Thomas, R. Treadon, G. Vernieres, Data Assimilation Strategy and Development Plan for NCEP's Environmental Modeling Center, Office Note 517 https://doi.org/10.25923/kqra-6z66, 2024.

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. Wea. Forecasting, 24, 1691–1705, https://doi.org/10.1175/2009WAF2222201.1.

Lin, S.-J., and R. B. Rood, 1997: An explicit fux-form semi-Lagrangian shallow-water model on the sphere. Quarterly Journal of the Royal Meteorological Society, 123 (544), 2477–2498.

Morgan, Michael et al. "NOAA Modeling Strategy : strategic plan 2024-2033" , 2024, https://doi.org/10.25923/qggz-jb43

UFS Weather Model Users Guide,
https://ufs-weather-model.readthedocs.io/en/develop/Introduction.html

Zhang H., S. Kondragunta, I. Laszlo, H. Liu, L.A. Remer, J.F. Huang, S. Superczynski, P. Ciren, An enhanced VIIRS aerosol optical thickness (AOT) retrieval algorithm over land using a global surface reflectance ratio database, J. Geophys. Res. - Atmos., https://doi.org/10.1002/2016JD024859, 2016.

Zhang, L., Montuoro, R., McKeen, S. A., Baker, B., Bhattacharjee, P. S., Grell, G. A., Henderson, J., Pan, L., Frost, G. J., McQueen, J., Saylor, R., Li, H., Ahmadov, R., Wang, J., Stajner, I., Kondragunta, S., Zhang, X., and Li, F.: Development and evaluation of the Aerosol Forecast Member in the National Center for Environment Prediction (NCEP)'s Global Ensemble Forecast System (GEFS-Aerosols v1), Geosci. Model Dev., 15, 5337–5369, https://doi.org/10.5194/gmd-15-5337-2022, 2022.

Zhang, X., Kondragunta, S., da Silva, A., Lu, S., Ding, H., Li, F., and Zhu, Y.: The blended global biomass burning emissions product from MODIS and geostationary satellites (GBBEPx), http://www.ospo.noaa.gov/Products/land/gbbepx/docs/GBBEPx_ATBD.pdf (last access: 1 June 2018), 2014.

This page is intentionally left blank

A. Revision history

The Table below documents revisions of this document.

Version	Date	Description
1.0	7/13/2026	Final approval for publication.