NWS FIELD EVALUATION PLAN

# Global Ensemble Forecast System v12

Official evaluation web page: https://www.emc.ncep.noaa.gov/users/meg/gefsv12

## Project Managers

Evaluation Plan Manager: Jason Levit
Field Evaluation Manager: Geoff Manikin
Model Project Lead: Yuejian Zhu
Waves/Ocean Component Lead: Henrique Alves
Aerosol Component Leads: Jeff McQueen, Partha Bhattacharjee
Subseasonal Evaluation Lead: Matthew Rosencrans
Project Manager: Vijay Tallapragada

## Important Dates

Retrospective forecasts complete: December 31, 2019
Reforecasts complete: January 31, 2020
MEG Evaluation Kickoff webinar: February 27, 2020
Field Evaluation: March 2, 2020 - April 10, 2020
MEG Evaluation Briefing: April 23, 2020
Field Recommendations due: April 27, 2020
NCEP Director Briefing: May 5, 2020 (tentative)
Code delivery to NCO: May 15, 2020
NCO 30-Day IT Test begins: July 27, 2020 (tentative)
NCO IT Briefing: August 31, 2020 (tentative)
Implementation date: September 7, 2020 (tentative)

## Evaluation Summary:

The GEFS v12 Evaluation will consist of four primary components: Atmospheric Week 1, Atmospheric Weeks 2-4, Waves Weeks 1-2, and an Aerosol 5-day component. The 2-year retrospective runs are initialized using the latest hybrid 4DEnsVar analysis and EnKF analysis, and are cycled, mirroring the initialized methodology that is used in real-time operations. The 30-year reforecast runs are initialized at 00Z using Climate Forecast System analyses for 1989-1999, and from 2000-2018, a hybrid FV3GFS/EnKF reanalysis system, developed by ESRL/PSD.

Atmospheric Week 1 Retrospectives:
Main retrospectives: June 1, 2017 - November 30, 2019; 00Z Cycle, 31 members
Extra hurricane retrospectives: June 1, 2019 - Sep. 30, 2017; 2018; 2019; 12Z Cycles, 31 members
Evaluation customers: National Weather Service regions and selected service Centers

Atmospheric Week 2 Retrospectives:
Main retrospectives: December 1 2018 - November 30 2019; 00Z Cycle, 31 members
Evaluation customers: Climate Prediction Center

Atmospheric Week 2-4 Reforecasts:
Main reforecasts: 1989-2019, 00Z Cycle, 5 members, to 16 days
Each Wednesday of forecast: 11 members to 35 days
Evaluation customers: Climate Prediction Center

Ocean Waves Week 1-2 Retrospectives:
Main retrospectives: Dec 1, 2018 - Nov 30, 2019; 00Z Cycle, 31 Members, to 16 days
Evaluation customers: OPC, NHC, MAG/NCO, FNMOC/Navy, CMC/EEEC

Aerosol 5-day Retrospectives:
Main retrospectives: Mar 1, 2019 - May 15, 2020, 00 Z Cycle, run to 120 hours, 1 member
Evaluation customers:

| **Name** | **Agency** |
|---|---|
| Craig Long | NCEP/CPC |
| Bob Grumbine | NCEP/EMC |
| NCEP WPC | WPC |
| NCEP AWC | AWC |
| NCEP NHC | NHS |
| Shobha Kondragunta | NESDIS/STAR |
| Brad Pierce | NESDIS/CIMMS |
| Pius Lee, Barbara Stunder | NOAA/ARL |
| Jeff Reid | DOD/Navy/NRL |
| Arlindo DaSilva, Mian Chen, Steve Pawson | NASA/GSFC |
| Ed Hyer | DOD/Navy/NRL/ICAP rep |
| Barron Henderson | U.S. EPA |
| Scott Epstein | South Coast California AQMD |
| Michael Geigart | CT DEP |

| | |
|---|---|
| Dorothy Koch, Jose Delgado | NWS/STI |
| Mike Staudenmeier | Acting NWS/WR SSD chief |
| Gregory Patrick | NWS/SR SSD Chief |
| Ken Johnson | NWS/ER SSD Chief |
| Bruce Smith | NWS/CR SSD Chief |
| Ernesto Rodriguez | WFO San Juan |
| Pablo Santos or current SOO | WFO Miami |
| Larry Horwitz, Rusty Benson | NOAA/GFDL |
| Sarah Lu, Dave Edwards | NCAR |
| John Kerekes | U.S. Department of State |
| Dana Carlis | OAR/OWAQ |

## Evaluation Overview:

Atmospheric Week 1:
Version 12 of the Global Ensemble Forecast System (GEFS) is planned to be the first upgrade to the system in five years, and it will become the second system in NCEP operations to have the FV3 dynamical core. GEFSv12 will be validated during a short period in March 2020. EMC's Verification, Post-Processing, and Product Generation (VPPPG) branch will lead the evaluation of the Atmospheric Weeks 1 component, although the effort will involve the larger community of users. Due to resource constraints, no real-time parallel system will be run; the evaluation will be based entirely on retrospective runs, covering the recent 2.5-year period. Statistical measures will be computed, and a series of representative cases will be assessed.

A full set of verification statistics will be generated for GEFSv12 using the legacy VSDB statistical verification package, while a separate set of statistics will be generated using Model Evaluation Tools (METplus) software. A comparison of output from the two streams will be performed. If the METplus statistics are found to be satisfactory, those results and images will be used to populate web displays.

The primary comparison statistics will be GEFSv12 vs GEFSv11 to help assess improvement of the new system relative to what is currently running in operations. A full list of currently available metrics from the GEFS is listed in Appendix 1, and will be used as the starting point for verification efforts.

The MEG will select a set of cases from the GEFSv12 retrospective experiments for subjective evaluation. All images will need to be completed prior to the start of the evaluation period. The cases must cover a diverse range of GEFS user needs including tropical cyclones, winter storms, significant QPF events, large-scale severe weather outbreaks, pattern transitions, and low skill events. A list of candidate cases that the MEG has identified is contained in Appendix 2.

The MEG is targeting around 30 case studies for the evaluation, but that list may change at evaluation time. For each selected case, the most relevant cycles will be selected, and graphics comparing multiple GEFSv12 retrospective cycles with operational GEFSv11 runs will be generated for a relevant domain. Basic parameters including the ensemble means of 500 hPa geopotential heights, mean sea level pressure, QPF, 850 hPa winds and temperatures, 10m winds, precipitable water, and 250 hPa winds will be plotted. Snow accumulations will be plotted for winter cases, and surface-based convective available potential energy and 2m dew points will be displayed for severe weather cases. Means and spreads will be displayed, and the MEG plans to compute and display some basic probability fields (i.e. QPF > 1, 2”/24h, 2m and 850 hPa T < 0° C, CAPE > 2000, 4000 J/kg…..). There are plans to display low centers and generate spaghetti plots, but these images require downloading all ensemble members, so they will be generated sparingly. For the same reason, generating comparison station plume images is not feasible.

Assessments will be based on performance statistics as well as the forecast maps from high-impact cases.  It is expected that all of the NWS regions will participate in the evaluation, as well as the National Hurricane Center, Weather Prediction Center, Ocean Prediction Center, and Storm Prediction Center. Participation by the Space Weather Prediction and Aviation Weather Centers will be optional. The Climate Prediction Center will focus on the weeks 2-4 evaluation. Each center will be expected to provide a formal assessment at a briefing given to the NCEP Director at the conclusion of the evaluation period. The MEG will be responsible for gathering the field assessments and will give their own independent assessment.

MEG meetings, held each Thursday at 11:30 AM Eastern Time, have been excellent forums for informing customers and stakeholders about evaluation procedures and sharing evaluation findings with the field. The MEG will plan to present an overview of the evaluation process and availability of statistics/graphics at a kickoff webinar approximately one to two weeks prior to the start of the evaluation period. The MEG will then give their overview findings the week after the evaluation period ends, and a final webinar covering recommendations from the field will be presented the week after that. These webinars are open to all, and presentations are recorded and placed along with the slides in an online google folder. Presentations will also be made available on the evaluation website.

Atmospheric Weeks 2-4:

The evaluation of the Atmospheric component for weeks 2-4 will be conducted by the Climate Prediction Center (CPC), and the results will be collected by EMC's Model Evaluation Group prior to the NCEP Director briefing. A list of metrics that will be provided by CPC is listed in Appendix 3.

CPC will be comparing GEFSv12 output in a few manners. For stratospheric component to CPC operations, CPC will compare the GEFSv12 retrospective runs to operational runs of GEFSv11 valid over the same period archived at CPC. For tropospheric circulation and surface variables that we have in the GEFSv10 reforecast dataset archived at CPC, CPC will process the reforecast dataset to obtain the relevant statistics, then use the retrospective period as real-time forecasts, and obtain relevant metrics. For the surface variables of 2-meter temperature and precipitation, CPC will process the reforecast dataset to obtain the relevant statistics, then use the retrospective period as real-time forecasts, and obtain relevant metrics. Since CPC never adopted the GEFSv11, this comparison will be against the GEFSv10 (GEFS legacy).

Waves Weeks 1-2:

EMC's Wave Modeling Group will be conducting the evaluation for the Waves component of the GEFS v12, with results collected by EMC's Model Evaluation Group prior to the briefing to the NCEP Director.

Validation data will be provided to our main customers including the Ocean Prediction Center (OPC), National Hurricane Center (NHC), US Navy and Environment Canada. Focus will be on general skill of probabilistic wave forecasts relative to NDBC buoy data covering all validation parameters listed below, and altimeter measurements of wave height and surface wind speeds. Verification of wave forecasts will be made relative to the existing operational products at NCEP and US Navy (FNMOC), including impacts to the multi-center wave ensemble products, focusing on changes to predictability of both the individual NCEP ensemble data and the joint NCEP-FNMOC products. Validation and verification data for evaluation will include bulk statistics by month, and yearly means, as well as case studies to be selected in consultation with stakeholders.

Aerosol 5-day:

The NEMS coupled app (FV3GFS-CHEM) includes two components: FV3GFS V15 and GSDCHEM. GSDCHEM is a NUOPC-based chemistry component developed to replace the current NEMS GFS Aerosol Component (NGAC at 1x1°, Wang, et al. 2018) GSDCHEM includes the WRF-Chem (Grell, et al. 2005) chem_driver with updates for consistency with the NASA Goddard Operational Chemistry and Aerosol Radiation andTransport (GOCART; Chin, et al., 2007) version. The chemistry and aerosol modules used for FV3GFS-CHEM include simple sulfur chemistry, hydrophobic and hydrophilic black and organic carbon, and a 5-bin sea salt module. Additionally, included is the FENGSHA (Dong, et al. 2016) 5-bin dust module, wildfires modeling using Fire Radiative Power (FRP) data from MODIS measurements and the NESDIS Global Biomass Burning Emissions Product (GBBEPx; Zhang, et al., 2012). Plume rise modeling is done with a 1d cloud model (Grell & Freitas, 2014), and, optionally, volcanic ash emissions are also included. Tracers are transported by the dynamics as well as the GFS physics (GFS PBL and Simple Arakowa Shubert (SAS) deep and shallow convection parameterization). Subgrid scale wet scavenging is done inside the two SAS routines.

The system will be run at C384L64 resolution (~25 km) as a member of the GEFS but with GOCART simple aerosol chemistry (19 species) run to 120 forecast hours four times per day. FV3GFS-Chem currently requires 40 nodes to run 5 days in 37 cpu minutes on the Dell Phase III systems.

Technically, coupling occurs two-way, as mixing ratios of chemical tracers are exchanged between FV3GFS and GSDCHEM at each coupling step to be advected by FV3 dynamical core. However, at this point coupling is considered to be only one way in this milestone from a scientific standpoint, since feedback to the meteorology is not yet activated.

At each coupling time step, a complete set of fields is provided by FV3GFS to GSDCHEM, which includes them in chemistry computations and returns updated mixing ratios for the chemical tracers to FV3GFS. Tracer concentrations and some diagnostic chemical output are included in FV3GFS history files. Optionally, the results of GSDCHEM computations can also be written to Fortran unformatted sequential files (for debugging), one file per tile of the cubed sphere grid.

All 2D and 3D fields exported by FV3GFS are initialized using baseline input data provided for regression testing on Theia, Cray and Dell (fv3_control). FV3 data structures (IPD_Data) corresponding to these fields are shown in Table 2. 19 chemical tracers are defined in the FV3GFS input field_table file with a spatially constant non-zero value at the surface. These tracers are also added to the diag_table file to be included in FV3GFS dynamics history files.

GEFS-Aerosols retrospectives will be run from March 2019 and ending May 2020. Evaluations against ATOM-1 2016 field experiment will have already been performed by OAR. Output data will be 3 hourly and run with 1x/day cycling and require UPP grib2 files and special diagnostic files. Evaluation will follow previous NGAC protocols (Bhattacharjee, et al. 2018). The following fields will be evaluated by comparing model outputs to

- MODIS satellite AOD 1° gridded product
- VIIRS satellite AOD 0.25° degree gridded product
- AERONET AOD especially stations near dust and smoke sources
- International Centers for Aerosol Prediction (ICAP) ensemble forecast - 1° total and dust AOD
- NASA GEOS-5/MERRA-II gridded total AOD and speciated analyses (PM2.5, PM10, SO4, dust, OC, BC)
- Global PM2.5 and PM10 surface measurements
- Monthly compared Calipso aerosol profiles

Daily and monthly averaged comparisons to above observations/analyses of

- Gridded AOD RMSE,BIas and correlation on global and regional maps
- diurnal and daily time series at AERONET sites

An annual budget calculation of species will also be performed and compared to typical budgets as provided by NASA for previous NGAC upgrades.

Specific attention will be paid to

- Emissions
  - o Evaluate GBBEPx wildfire smoke emission with plume rise for both wildfires and agricultural burns
- Evaluate sea salt predictions especially in tropics
- Evaluate SO4 predictions especially over East Asia and Eastern U.S.
- detailed evaluation during high aerosol concentration episodes of :
  - biomass burning
  - o dust
  - o anthropogenic SO4
  - o over ocean predictions

Each run will require 6 gb/day for 3 hourly output. 18 nodes will be used per run with a 1 hour wall clock.

Remaining Evaluation work for GEFS-Aerosols

- Run two cold start runs for July and Dec. 2019. One with 300 and one with 450 sec. Present preliminary evaL Li Pan/ Partha: **DT=300 sec decided upon**
- Implement warm start option into GEFS workflow and test. **Feb 18, 2020**
- Run one year of retrospectives (March 1, 2019 - Feb. 29, 2020 ): **March 18, 2020.**
- Complete retro eval, finalize web page for Partha's retro graphics (daily/monthly avg) access and provide to users for two month evaluation. **March 25, 2020**

GEFS-Aerosols Output

Outputs will be in grib2 and include the same fields currently output by NGAC but at a finer 0.25 degree resolution. These fields will include:

| Table 2. Aerosol name | unit | **Domain** |
|---|---|---|
| 3D fields | | |
| DUST1_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| DUST2_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| DUST3_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| DUST4_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |

| | | |
|---|---|---|
| DUST5_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| SEASALT2_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| SEASALT3_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| SEASALT4_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| SEASALT5_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| BCPHILIC_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| BCPHOBIC_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| OCPHILIC_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| OCPHOBIC_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| SO4_ON_HYBRID_LVL | ug/m3 | 1 hybrid level |
| 2D fields | | |
| AER_OPT_DEP_at550 (total) | | entire atmosphere |
| DUST_AER_OPT_DEP_at550 | | entire atmosphere |
| SEASALT_AER_OPT_DEP_at550 | | entire atmosphere |
| SULFATE_AER_OPT_DEP_at550 | | entire atmosphere |
| ORGANIC_CARBON_AER_OPT_DEP_at550 | | entire atmosphere |
| BLACK_CARBON_AER_OPT_DEP_at550 | | entire atmosphere |
| DUST25_SFC_MASS_CON (dust pm2.5) | ug/m3 | 1 hybrid level |
| SEAS25_SFC_MASS_CON (sea salt pm2.5) | ug/m3 | 1 hybrid level |
| PM10_SFC_MASS_CON | ug/m3 | 1 hybrid level |
| PM25_SFC_MASS_CON | ug/m3 | 1 hybrid level |
| PM10_COL_MASS_DEN | kg/m2 | entire atmosphere |
| PM25_COL_MASS_DEN | kg/m2 | entire atmosphere |
| DUST_COL_MASS_DEN (PM2.5) | kg/m2 | entire atmosphere |
| SEAS_COL_MASS_DEN (PM10) | kg/m2 | entire atmosphere |
| BC_COL_MASS_DEN | kg/m2 | entire atmosphere |
| OC_COL_MASS_DEN | kg/m2 | entire atmosphere |
| SULF_COL_MASS_DEN | kg/m2 | entire atmosphere |

In addition, an option to output a diagnostic file is available but will not be used operationally. FV3GFS-GSDchem also outputs other chemical tracers that are needed for reinitializing the next forecast.

## Appendix 1: Atmospheric Week 1 Evaluation Metrics

| METRIC | LEVEL(s) | VERIFICATION | VERIFICATION DATA |
|---|---|---|---|

| | | METHOD | |
|---|---|---|---|
| RMSE, mean absolute error, and ensemble spread of T | 2m, 850 hPa | grid-to-grid | GFS analyses |
| RMSE, mean absolute error, and ensemble spread of Z | 500 hPa, 1000 hPa | grid-to-grid | GFS analyses |
| RMSE, mean absolute error, and ensemble spread of U, V | 10m, 250 hPa, 850 hPa | grid-to-grid | GFS analyses |
| ROC curve, Economic Values, and Ranked Probability Skill Score for Z | 500 hPa, 1000 hPa | grid-to-grid | GFS analyses |
| ROC Curve, Economic Values, and Ranked Probability Skill Score for T | 2m, 850 hPa | grid-to-grid | GFS analyses |
| ROC Curve, Economic Values, and Ranked Probability Skill Score for U, V | 10m, 250 hPa, 850 hPa | grid-to-grid | GFS analyses |
| Brier Skill Score, CRP Score, CRP Skill Score, and Anomaly Correlation for Z | 500 hPa, 1000 hPa | grid-to-grid | GFS analyses |
| Brier Skill Score, CRP Score, CRP Skill Score, and Anomaly Correlation for T | 2m, 850 hPa | grid-to-grid | GFS analyses |
| Brier Skill Score, CRP Score, CRP Skill Score, and Anomaly Correlation for U, V | 10m, 250 hPa, 850 hPa | grid-to-grid | GFS analyses |
| ETS, TSS, and Bias for Precipitation (at days 1, | Surface | grid-to-grid | CCPA |

| | | | |
|---|---|---|---|
| 2, 3, 5, 8) | | | |
| RMSE, Spread, Mean/Absolute Error, and CRP Score for Precipitation Time Series | Surface | grid-to-grid | CCPA |
| Precipitation Reliability Diagrams | Surface | grid-to-grid | CCPA |
| Tropical Cyclone Mean Track Errors | Surface | track-to-track | NHC Best Track Data |

## Appendix 2: Atmospheric Week 1 Evaluation Case Studies

TROPICAL CYCLONES

| | | |
|---|---|---|
| Harvey (2017) | Irma (2017) | Maria (2017) |
| Nate (2017) | Noru (2017) | Alberto (2018) |
| Florence (2018) | Lane (2018) | Michael (2018) |
| Walaka (2018) | Barry (2019) | Dorian (2019) |

WINTER WEATHER

| | | |
|---|---|---|
| Southeast Snowstorm (December 2017) | East Coast Bomb (January 2018) | West Coast Atmos River (March 2018) |
| Mid-Atlantic Wind Storm (March 2018) | Arctic Outbreak (January 2019) | Central US Bomb (March 2019) |
| Arctic Outbreak (2017) | April Alaska Cyclone (2018) | November 2018 Mid-Atlantic Snowstorm |

QPF

| | | |
|---|---|---|
| TX/OK Bust (Dec 2018) | Summer Nor'easter (Jul 2017) | TC Imelda (2019) |
| Will contact Alaska Region for some more cases | | |

SEVERE WEATHER

| | | |
|---|---|---|
| Late May pattern (2019) | February Dixie Outbreak | May 2-3 Neg Tilt Trough |

| | | |
|---|---|---|
| | (2018) | (2018) |
| Early November Outbreak (2017) | | |

LOW SKILL PERIODS

| | | |
|---|---|---|
| General Low Skill Period (Dec 2018-Feb 2019) | Missed Gulf Cutoff (Feb 2018) | Atlantic Pattern Issues (Nov 2017) |
| Atlantic Pattern Issues (Aug 2017) | Low Skill (June 2017) | |

WEEK 2-4 PERIOD

| | | |
|---|---|---|
| Flash Drought June 2017 | Winter cold snaps (2019) | CA Wind Events in Nov 2019 |
| Optional: Alaska Heat waves of 2018, 2019 | | |

## Appendix 3: Atmospheric Week 2-4 Evaluation Metrics

| METRIC | LEVEL(s) | PERIOD(S) | Domain(s) | VERIFICATION METHOD | VERIFICATION DATA |
|---|---|---|---|---|---|
| AC | 500mb heights | 6-10 day | NH | Grid-to-grid | R2 |
| AC | 500mb heights | 8-14 day | NH | Grid-to-grid | R2 |
| AC | 500mb heights | Weeks 3-4 | NH | Grid-to-grid | R2 |
| AC die off curves | MJO | Through 35 days | Global | Timeseries | WH timeseries from CADB |
| | Temperature | 0-384 hour | NH/SH high lat | Grid-to-grid / Timeseries | GEFS Analysis |
| | Zonal Wind | 0-384 hour | NH high lat | Grid-to-grid / Timeseries | GEFS Analysis |

| | | | | | |
|---|---|---|---|---|---|
| | Total Ozone | 0-384 hour | NH/SH high lat | Grid-to-grid / Timeseries | GEFS Analysis |
| | SPFH | 0-384 hour | NH high lat | Grid-to-grid / Timeseries | GEFS Analysis |
| | U wind | 0-384 hour | NH high lat | Grid-to-grid / Timeseries | GEFS Analysis |
| HSS (RPSS - maybe) | 2m temp | 6-10 day | ERF Domain | Grid-to-grid | CPC Temp Analysis |
| HSS (RPSS - maybe) | 2m temp | 8-14 day | ERF Domain | Grid-to-grid | CPC Temp Analysis |
| HSS (RPSS - maybe) | 2m temp | Weeks 3-4 | ERF Domain | Grid-to-grid | CPC Temp Analysis |
| HSS (RPSS - maybe) | Precip | 6-10 day | ERF Domain | Grid-to-grid | CPC Precip Analysis |
| HSS (RPSS - maybe) | Precip | 8-14 day | ERF Domain | Grid-to-grid | CPC Precip Analysis |
| HSS (RPSS - maybe) | Precip | Weeks 3-4 | ERF Domain | Grid-to-grid | CPC Precip Analysis |
| Threat score | TC formation | Weeks 3-4 | ATL and EPAC | Timeseries / Point data | Best Tracks |

## Appendix 4: Ocean Waves Week 1-2 Evaluation Metrics

| METRIC | LEVEL(s) | VERIFICATION METHOD | VERIFICATION DATA |
|---|---|---|---|
| Bias, RMSE, SI, Correlation | Significant wave height (Hs), Peak wave periods (Tp), Surface wind speed (U10), Mean wave direction (ThetaM), Mean zero-crossing wave period (T02) | Station-based (buoys, coastal and deep water), grid-to-grid with zonal focus | NDBC Buoys, Altimeters, operational wave models (NCEP, US Navy) |

| | | | |
|---|---|---|---|
| CRPS, Spread-skill relationship, Talagrand diagrams, Threshold exceedance reliability, Brier skill score, Hit/miss rate | Significant wave height (Hs), Peak wave periods (Tp), Surface wind speed (U10), Mean wave direction (ThetaM), Mean zero-crossing wave period (T02) | Station-based (buoys, coastal and deep water), grid-to-grid with zonal focus | NDBC Buoys, Altimeters, operational wave models (NCEP, US Navy) |

## Appendix 5: Chemistry 5-day Evaluation Metrics

| **METRIC** | **LEVEL(s)** | **VERIFICATION METHOD** | **VERIFICATION DATA** |
|---|---|---|---|
| Bias, RMSE, SI, Correlation | Aerosol Optical Depth (AOD, total column) | Grid-to-grid | ICAP, GEOS-5 analyses, MODIS and VIIRS 25 km gridded fields |
| Bias, RMSE, SI, Correlation | AOD | Grid-to-obs | Global AERONET surface lidar network |
| Bias, RMSE, SI, Correlation | Organic Carbon (OC), Dust, Sea salt, Sulfate | Grid-to-grid | GEOS-5, MERRA-II analysis |
| Bias, RMSE, SI, Correlation | Surface PM2.5 | Grid-to-obs | Global surface air quality network |