

## Recommendations for Evaluating the Performance of the IWDW-WAQS Photochemical Grid Model Platform

### Purpose:

This document recommends a series of model evaluation analyses to assist the IWDW-WAQS Technical Committee in determining the adequacy of the photochemical grid model (PGM) platform for use in NEPA and other air quality modeling applications. Along with the recommendations for the completion of a set of standard analyses for characterizing and evaluating the model platform, the document also provides an approach for summarizing and interpreting the results of the model performance evaluation (MPE) to determine the suitability of the model platform for a specific use. The approach outlined below does not replace the Air Quality Modeling Protocol. However, this document could be attached as an appendix or incorporated into the modeling protocol.

It is expected that this document will assist the IWDW-WAQS Technical Committee in identifying the resources (time and funding) that will be needed to complete future model platform evaluations. The document is based on EPA guidance on model performance evaluations (USEPA, 2007, 2014<sup>1</sup>) and is supplemented with analyses that are particularly relevant for air quality concerns in the western United States.

### Background:

The objective of a model performance evaluation (MPE) is to demonstrate that the base case model can simulate observed pollution concentrations during historical periods and thereby develop confidence that the model can reliably predict how future pollution levels will change in response to changes in emissions. There are multiple approaches for evaluating model performance (USEPA, 2014). This document focuses specifically on the analyses important for operational and diagnostic evaluations. Analyses to support dynamic and probabilistic evaluations can be found in EPA's air quality modeling guidance.

The goal of an operational evaluation is to determine how well the model replicates observed concentration of PM<sub>2.5</sub> components, ozone and their precursors, as well as deposition of specific nitrogen and sulfur species. While historically most of the evaluation efforts have focused on the operational evaluation, the MPE should also include some level of diagnostic evaluation where practical. Diagnostic evaluations are used to determine whether the model accurately represents the processes that affect pollutant concentrations. For instance, these evaluations assist in determining whether the model accurately reproduces the observed conditions for the right reasons. These processes within the model include emissions, meteorology (e.g., dispersion), chemical reactions, and deposition. For cases in which the model performs poorly in the operational evaluation, the diagnostic evaluation can provide insights into the causes of the poor performance and how to improve model performance. Comprehensive diagnostic testing of the model can also be a valuable component of the weight of evidence analysis used to determine when the model platform is sufficient for particular modeling applications.

### Organization of Document:

The analyses and work products recommended for completion as part of the IWDW-WAQS air quality modeling efforts have been grouped by each component of the PGM platform to help in being specific and for ease in distribution. Sections of the document include:

1. Assessment of boundary conditions
2. Assessment of emissions inventory and modeling results
3. Evaluation of meteorological modeling results
4. Evaluation of photochemical grid modeling results
5. Summarizing and interpreting results of platform MPE

---

<sup>1</sup> EPA's current model performance evaluation guidance was released in 2007. However, EPA released draft model performance evaluation guidance in December 2014 that will replace the current guidance upon public review and final approval.

Within each section or component of the modeling platform included in this document, analyses and work products are prioritized and categorized by operational evaluation or diagnostic evaluation. While not all of the analyses outlined in this document are required and may not be necessary to complete until model performance issues are identified in the operational evaluation, it is important to recognize these analyses and work products so that the components of the modeling platform can be configured to store and output information for instances when these analyses are needed. In addition, this document includes a section that recommends an approach for summarizing and interpreting the results of the MPEs to assist the IWDW-WAQS in documenting the level of the model performance and determining the adequacy of the model platform for use in NEPA and other air quality planning projects.

## ASSESSMENT OF BOUNDARY CONDITIONS

Previous regional modeling analyses over the IWDW-WAQS region have shown significant influence from model boundary conditions on simulated pollutant concentrations within the domain (WRAP, 2013). As a result, an initial qualitative assessment of the boundary conditions and their influence on simulated concentrations within the domain is recommended. This is intended to assist in determining how accurately the model platform simulates the transport of air pollution from outside the domain to locations within the domain through the use of boundary conditions derived from a global model (GM) simulation (e.g., GEOS-Chem, CAM-chem, AM3, hemispheric CMAQ, etc.). Two options for the initial assessment of the boundary conditions include:

- a. Evaluation of the Global Model: This assessment includes a review the MPE completed by the GM developers (if available), or performing an independent MPE to determine the adequacy of the GM performance. To the extent possible, an evaluation of this sort should consider performance aloft (where ozone transport is most important), as opposed to simple surface comparisons.
- b. Estimating Lateral Boundary Inflow using a PGM Inert Simulation: This assessment includes a model simulation using BC inputs without simulating emissions or chemistry. Relative to the other options, minimal resources would be needed to complete this evaluation. However, inert tracer runs have been shown to overestimate boundary ozone influence in some cases (Baker et al., 2015), so an evaluation of this type is not intended to provide a quantitative estimate of ozone lateral boundary inflow.

After the base case PGM development is completed, more quantitative approaches can be used to assess BC contributions. Three options for the quantitative assessment of the boundary conditions include:

- a. Estimating Lateral Boundary Inflow using a PGM Reactive Tracer Simulation: This assessment includes a model simulation tracking BC inputs with reactive tracers to account for atmospheric chemistry. This option is preferred over using unreactive tracers when quantitative estimates of lateral boundary ozone contribution are desired. Examples include CAMx OSAT/PSAT, and CMAQ ISAM.
- b. Estimating US Background Contribution with a PGM Chemistry Simulation: This evaluation includes a simulation with chemistry and no US anthropogenic emissions to assess US background levels of ozone and PM<sub>2.5</sub>.
- c. Evaluation of a Full PGM Simulation: This assessment includes the PGM MPE operational evaluation with full-chemistry and full-emissions. The analysis would focus on rural or remote sites that have little influence from US anthropogenic emissions.

When utilizing these options, plots of time-series of O<sub>3</sub> and speciated PM<sub>2.5</sub> at selected monitors are recommended for assessing the model performance. These types of plots will aid in identifying days on which the model has high background levels. The results will also be useful for comparison with subsequent model simulations with anthropogenic emissions to assess the contribution of anthropogenic emissions to O<sub>3</sub> and PM<sub>2.5</sub>.

If issues are identified as part of the operational evaluation, a more intensive evaluation of the ozone and PM<sub>2.5</sub> boundary contributions should be completed as part of a diagnostic evaluation. This evaluation will assist in determining whether days with large boundary contributions are a result of long-range/international transport, stratospheric intrusions, fires, or methane. Based on these results, a determination as to whether and what types of improvements in the global model are warranted. For model scenarios with questionable model performance on high boundary influence days, it may be useful to perform a diagnostic model simulation with boundary inputs derived from a different global model to see how the inputs would change the photochemical grid model performance.

## ASSESSMENT OF EMISSIONS INVENTORY AND MODELING RESULTS

Emissions inventories are compiled from a large number of different data sources, which can lead to many types of data problems. These problems can include errors in the values of emissions inventory data fields, such as the emissions amounts, incorrect codes, missing values that are required, or incorrect data such as facility names, stack heights, or latitude/longitude locations. Problems can also include errors that cause databases to not work, such as missing records in relational database tables, duplicate records, or incorrect formats. Therefore, it is important to perform quality assurance (QA) checks on the emissions inventories to identify and correct these errors before the data are used for the modeling activities. The purpose of QA is to ensure the development of a complete, accurate, and consistent emissions inventory. These QA activities generally include:

- Checking data codes: Check inventory data codes to ensure they are valid codes.
- Point versus nonpoint reconciliation: Ensure sources are identified properly, are not double-counted across point and nonpoint, and do not have missing pollutant emission rates, missing temporal allocations or missing spatial components.
- Rankings and percent differences techniques: Ranking emissions and calculating percent differences (with previous emissions estimates) can be useful to identify errors in emissions values or other numeric inventory data fields, or at least in finding records that need further investigation. This analysis must be done one pollutant at a time. Large or unexpected differences can be investigated further to make sure that the differences have some reasonable explanation.
- Graphical displays: Mapping emissions values for individual SCCs, groups of SCCs, county level emissions, and emissions differences among inventories can identify unexpected patterns in the data that could signal an error. These displays also identify pollutants that are expected from particular processes, and conversely, pollutants that are not expected from a particular process and need additional investigation.
- Comparisons among emissions inventories and regions: Comparing previous emissions, new (or proposed) emissions, region-by-region emissions, and the computed difference or percent difference side-by-side can reveal unexpected changes or patterns that could signal an error.

The QA checks should be done on a pollutant-by-pollutant basis for pollutants important to the given analysis. In particular, the assessment should include analyzing NO<sub>x</sub>, VOCs, NH<sub>3</sub>, CH<sub>4</sub>,<sup>2</sup> SO<sub>2</sub>, PM<sub>2.5</sub>, PM<sub>10</sub>, and PMC for the source categories outlined in Table 1. In addition to utilizing the automated QA checks built into EPA's emissions models, the recommended QA checks should be used to identify additional differences or outliers that may warrant further explorations and corrections. Examples of various recommended graphical and tabular displays are also included below.

### Recommended analyses:

1. Check inventory data codes to ensure they are valid codes.

---

<sup>2</sup> Note: Quality assurance checks for methane may depend on the completeness of the NEI for this constituent.

- a. SMOKE Sector (or other emissions processor) QA for each modeling sector:
  - i. Check SMOKE logs to make sure the correct inputs were used
  - ii. Check that all output files were created and are the correct size
  - iii. Use EPA's "log analyzer" to make sure there are no errors or unusual warnings
  - iv. Compare post-SMOKE annual totals with Emissions Modeling Framework (EMF) inventory totals; usual tolerance is 1.0% or less
- b. SMOKE Merge QA:
  - i. Check SMOKE logs to make sure the correct sectors and cases are used in the merge
  - ii. Calculate post-merge annual domain totals and compare to the sum of the sector annual reports, minus inline emissions totals
  - iii. Check that all output files were created and are the correct size
2. Ensure point and nonpoint sources are identified properly, are not double-counted across point and nonpoint, and do not have missing emissions.
3. Spatial plots with pie charts for each county of the various source category contributions by pollutant and region.
4. Bar plots of source category contributions by county by pollutant.
5. Tabulate the significant sources (SCC) contributing to each source category by pollutant and region.
6. Pie Chart of Spatial, Temporal, and Chemical Allocations by source category by pollutant.

Additional analyses:

1. Bar, spatial, and pie plots of VOC Reactivity Analysis (RTOG/TOG)<sup>3</sup> for various source category contributions by region.
2. Plots and Tables of spatial and temporal allocations and chemical speciation assignments.
3. Comparison to previous emissions modeling efforts to verify expected emissions changes by source category(s) and geographic area(s) of interest.
  - a. Bar plots of annual emissions contributions of source categories by pollutant and region.
  - b. Trends tables between years for various sectors and pollutants by region.
  - c. Bar plots of emissions differences of the contributions of various source categories by state and pollutant.
  - d. Bar plots of changes from the previous year to the current year (e.g., 2008 to 2011) NEI for various source categories and pollutants.
  - e. Side-by-side pie charts of source category contributions by pollutant and region.

**Table 1. Source categories for emissions assessment**

Source category	Category description
Point sources	Emissions sources that are attributed to specific facilities and emissions release points within those facilities. To the extent possible, CEMS data for NO <sub>x</sub> and SO <sub>2</sub> should be used when evaluating model performance.
Nonpoint sources	Also called "area sources" or sources of emissions that have not been inventoried as specific point sources.
On-road mobile sources	Emissions from on-road vehicles are the result of several emission processes, including the combustion of fuel while vehicles are starting, idling, and moving, evaporation of fuel from the fuel system and during refueling, and from brake wear and tire wear.
Non-road mobile	Emissions resulting from the use of fuel in a diverse collection of vehicles and

<sup>3</sup> This analysis could consider Maximum Incremental Reactivity (MIR) values developed by William P. L. Carter.

sources	equipment, and the EPA's NONROAD emissions model estimates of these emissions. Also includes emissions from Commercial Marine Vessels, Locomotives, and Aircraft.
Oil and Gas sources	Emissions resulting from oil and gas sources.
Biogenic sources	Subset of natural emissions sources, including vegetation (i.e., forests and agriculture) and microbial activity in the soil.
Geogenic and other natural sources	Emissions resulting from oil or natural gas seeps, soil wind erosion, sea salt spray, lightning, volcanoes and fumaroles (i.e., vapor or gas vents in a volcanic region).
Fires	Emissions from fires (i.e., agricultural, prescribed, or wildfire sources).
Canadian and Mexican emissions	Emissions from in-domain portions of Canada or Mexico

## EVALUATION OF METEOROLOGICAL MODELING RESULTS

The objective of this assessment should be to determine if the meteorological model output fields represent a reasonable approximation of the actual meteorology that occurred during the modeling period. This includes the identification of model biases and errors in order to allow for a downstream assessment of how the air quality modeling results may be impacted by the meteorological inputs. Therefore, we recommend statistical and graphical comparisons that are partitioned into meaningful subsets, such as by observational site, geographic sub-regions, and daily/monthly/seasonal periods, to allow for a more comprehensive assessment of the meteorological strengths and weaknesses. It may also be useful to break out model performance aloft, at the surface, during individual episodes (e.g., high ozone / PM<sub>2.5</sub> days), over the diurnal cycle, and as a function of synoptic regime. Wherever possible, the analyses should evaluate the model results against independent datasets or datasets not used in the nudging analysis to determine the model's ability to reproduce fields at locations with data not used in the data assimilation. In this case, one option could be to sequester some of the meteorological data for the performance evaluation.

### Recommended statistical analyses:

1. For temperature, mixing ratio, wind speed and direction, and precipitation, calculate the values of the following statistics for each day and for monthly averages:
  - a. Number of Data Points
  - b. Mean Model / Observations
  - c. Mean Bias / Error
  - d. Fractional Bias / Error
  - e. Correlation Coefficient
2. Where appropriate, calculate the same statistical outcomes (as above) for high pollution episodes for ozone and PM<sub>2.5</sub>. Depending on the nature and intent of the particular study, finer temporal and spatial aggregations may be appropriate if there are sufficient data to permit the calculation of meaningful statistics. For instance, metrics aggregated by day and monitor may be necessary to understand specific meteorological events.

### Recommended graphical analyses:

1. For temperature, mixing ratio, wind speed and direction, and precipitation, generate time-series plots of hourly modeled and observed data at individual monitor sites. These plots can be used to identify particular times and locations when/where the model performs well or poorly. It may also be useful to plot model results for multiple grid cells close to the monitor.
2. For temperature, mixing ratio, wind speed and direction, and precipitation, generate scatter plots of hourly modeled and observed data at individual monitor sites: Used to assess the model performance for a range of levels.

3. For temperature, mixing ratio, wind speed and direction, and precipitation, generate spatial plots of hourly modeled and observed data. Used to ensure that nudging is not producing “crop circles” in meteorological fields.
4. Vertical profiles of modeled and observed temperature, potential temperature, wind speed, and wind direction. These plots can be used to determine how well the model represents the atmospheric profiles and boundary layer heights.
5. Daily and/or monthly spatial plots of modeled and observed cloud cover and precipitation. These plots can be used to identify times/locations with good or poor performance, and where the model accurately simulates observed levels but misplaces the location of the high levels. Plots may also be a useful approach to determine days where the model may have problems predicting ozone because cloud cover impacts the photolysis rates and formation of ozone. An emerging dataset that could be considered for this comparison includes the Parameter-elevation Relationships on Independent Slopes Model (PRISM).
6. Daily and monthly spatial plots of modeled and observed albedo and snow depth for the winter season: Used to identify times with good or poor performance, and where the model accurately simulates observed levels but misplaces the location of the high levels.
7. Diurnal cycle plots of the model and observed planetary boundary layer height for time periods with measurements: Used to assess the model performance for the vertical structure of the atmosphere.

Additional graphical analyses:

1. For temperature, mixing ratio, wind speed and direction, and precipitation, generate Q-Q and Box and Whisker plots of hourly modeled and observed data at individual monitor sites. These analyses can be used to assess the dynamic range of the model.
2. Time series comparisons of modeled and observed shortwave downward radiation where available (i.e., from SURFRAD and ISIS networks).

If performance issues are identified as part of the operational meteorological evaluation (e.g., consistent temperature bias during periods with poor air quality, seasonal precipitation biases, etc.), additional diagnostic assessments may be needed. Analyses for diagnostic evaluations would include re-running the meteorological model with different configuration options based on issues identified during the operational evaluation. Testing various configuration options should be conducted in an iterative approach to better understand which options are controlling or improving the model performance. Similar graphical displays and/or statistical analyses outlined for the operational evaluation should be completed for the diagnostic evaluation. Alternative meteorological model configuration options that could be completed as part of the diagnostic evaluation include:

- Different meteorological models
- New releases of the meteorological model platform
- Cumulus parameterization schemes
- Planetary Boundary Layer schemes
- Analysis/Observational Nudging alterations
- Land Surface Model schemes
- Radiation schemes
- Microphysics schemes
- Analysis fields used in IC/BC
- Analysis of Nudging Configuration/Options

**EVALUATION OF PHOTOCHEMICAL MODELING RESULTS**

The goal of this portion of the platform assessment is to determine how well the air quality model replicates observed concentrations of ozone, ozone precursors, total PM<sub>2.5</sub>, individual PM<sub>2.5</sub> components, and deposition of key pollutant species. Statistical analyses should be calculated for pollutants and time periods that are most relevant for understanding the processes important for a given model application. The recommended statistical analyses and graphical displays are outlined below by pollutant.

Recommended statistical analyses for ozone and ozone precursors:

1. For hourly ozone and maximum 8-hour average (MDA8) ozone values, calculate the values of the following statistics for each day and for monthly averages:
  - a. Number of Data Points
  - b. Mean Model / Observations
  - c. Mean Bias / Error
  - d. Normalized Mean Bias / Error
  - e. Mean Fractional Bias / Error
  - f. Correlation Coefficient
2. For hourly ozone and MDA8 ozone values above certain thresholds (i.e., 60 ppb and other appropriate thresholds), calculate the values of the following statistics for each day and for monthly averages:
  - a. Number of Data Points
  - b. Mean Model / Observations
  - c. Mean Bias / Error
  - d. Normalized Mean Bias / Error
  - e. Mean Fractional Bias / Error
  - f. Correlation Coefficient
3. For hourly ozone precursors with observed data (e.g., NO<sub>x</sub>, NO<sub>y</sub>, VOC, speciated VOC, HNO<sub>3</sub>, NO, NO<sub>2</sub>, PAN, CO, CH<sub>4</sub>, SO<sub>2</sub>, and NH<sub>3</sub>), calculate the following statistics for each day and for monthly averages:
  - a. Number of Data Points
  - b. Mean Model / Observations
  - c. Mean Bias / Error
  - d. Normalized Mean Bias / Error
  - e. Mean Fractional Bias / Error
  - f. Correlation Coefficient
4. As appropriate, it may be valuable to aggregate certain regions (or even down to individual monitor) or time periods and evaluate model performance for those specific subsets using the same metrics shown above (e.g., seasonal averages, state-specific model performance, etc.).

Recommended graphical analyses for ozone:

1. Time-series plots of modeled and observed hourly and daily MDA8 ozone at individual monitor sites. These plots can be used to identify particular locations and periods where the model performs well or performs poorly.<sup>4</sup>
2. Scatter plots of modeled and observed hourly and MDA8 ozone at individual monitor sites. These plots can be used to assess model performance for a range of observed ozone levels. Again, it may be useful to

---

<sup>4</sup> Where there are concerns about fine-scale gradients affecting performance, it may also be useful to plot model results for multiple grid cells close to the monitor (e.g., best 9-cell match or shade ranged of predicted concentrations across 9 cells).

aggregate the modeled/observed pairs by geographic region (e.g., State) or time period (e.g., month/season) to assess whether there are important variations in performance across the domain.

3. Spatial plots of modeled and observed MDA8 ozone. These analyses can be used to assess performance on individual days (including days that may be particularly relevant to NAAQS issues).
4. Vertical profiles of modeled and observed ozone for locations and time periods with aloft measurements (e.g., ozonesondes). These plots can be used to determine how well the model represents ozone above the boundary layer and ozone transport.

#### Additional graphical analyses for ozone:

1. Box plots of the distribution of modeled hourly and MDA8 ozone bias. These plots can provide an overall perspective on the range of model bias within a simulation.
2. Q-Q plots of hourly and 8-hour ozone at individual monitor sites for time periods with measurements: Used to assess the dynamic range of the model.
3. Time-series plots of modeled and observed *hourly* ozone at individual monitor sites. Can be used to identify particular times of day or days of the week when the model performs well or poorly. Time periods with relatively poor performance can be flagged for additional investigation.
4. Spatial plots of modeled and observed hourly ozone. Can be used to identify specific hours (and locations within days) with good/poor performance.
5. Where appropriate, it may be valuable to generate graphical displays for subsets of the data most relevant to the application. For example, in cases involving winter ozone formation within a snow-covered basin, it may be valuable to link the air quality evaluation to the meteorological evaluation (e.g., assess model performance on days grouped by snow cover).

#### Recommended statistical analyses for total PM<sub>2.5</sub> and PM<sub>2.5</sub> components<sup>5</sup>:

1. It is important to evaluate each PM<sub>2.5</sub> component separately in order to understand whether the model is properly replicating the causes of high PM<sub>2.5</sub> episodes. For SO<sub>4</sub>, NO<sub>3</sub>, NH<sub>4</sub>, elemental carbon, organic carbon, and crustal elements, we recommend calculating the following statistics for each day and for monthly averages:
  - a. Number of Data Points
  - b. Mean Model / Observations
  - c. Mean Bias / Error
  - d. Normalized Mean Bias / Error
  - e. Mean Fractional Bias / Error
  - f. Correlation Coefficient
2. Again, as appropriate, it may be valuable to aggregate certain regions or time periods and evaluate model performance for those specific subsets using the same metrics shown above.

#### Recommended graphical analyses for total PM<sub>2.5</sub> and PM<sub>2.5</sub> components:

1. Time-series plots of modeled and observed daily PM<sub>2.5</sub> components and total PM<sub>2.5</sub> at individual monitor sites (e.g., SO<sub>4</sub>, NO<sub>3</sub>, NH<sub>4</sub>, elemental carbon, organic carbon, and crustal elements). Where available, ammonia (NH<sub>3</sub>) from available AMON sites should also be considered. These plots can be used to identify particular days and locations where the model performs well or poorly.

---

<sup>5</sup> An assessment of total PM<sub>2.5</sub> is only needed when monitors with speciated data are not available. In those cases, total PM<sub>2.5</sub> should be evaluated at individual monitoring sites at hourly and 24-hour averaging times.

2. Scatter plots of modeled and observed daily PM<sub>2.5</sub> components and total PM<sub>2.5</sub> at individual monitor sites. These plots can be used to assess model performance for a range of observed fine particulate levels. As always, it may be useful to aggregate the modeled/observed pairs by geographic region (e.g., State) or time period (e.g., month/season) to assess whether there are important variations in performance across the domain.
3. Spatial plots of modeled and observed daily PM<sub>2.5</sub> components and total PM<sub>2.5</sub>. These plots can be used to assess performance on days most relevant to the NAAQS.

Additional graphical analyses for total PM<sub>2.5</sub> and PM<sub>2.5</sub> components:

1. Time-series plots of modeled and observed hourly total PM<sub>2.5</sub> and PM<sub>2.5</sub> components at individual monitor sites for time periods with measurements. Can be used to identify particular times of day or days of the week when the model performs well or poorly.
2. Spatial plots of modeled and observed hourly total PM<sub>2.5</sub> or PM<sub>2.5</sub> components at individual monitor sites for time periods with measurements. Can be used to identify specific hours (and locations within days) with good/poor performance.
3. Q-Q plots of hourly and 24-hour total PM<sub>2.5</sub> or PM<sub>2.5</sub> components at individual monitor sites for time periods with measurements. These plots can be used to assess the dynamic range of the model.
4. Bugle plots of hourly and 24-hour PM<sub>2.5</sub> at individual monitor sites for time periods with measurements: Used to assess the model performance as concentrations increase/decrease in magnitude.

Recommended statistical analyses for SO<sub>4</sub>, NO<sub>3</sub> and NH<sub>4</sub> deposition:

1. Weekly spatial plots of modeled and observed data for NADP wet and CASTNet dry deposition. These plots can be used to identify time periods where the model accurately simulates observed levels but does not produce the expected location of the high concentrations.
2. Scatter plots of weekly and seasonal averages at individual monitor sites for wet deposition. Can be used to assess the model performance for a range of concentration levels.

If issues are identified as part of the operational photochemical grid model evaluation (e.g., consistent spatial bias during periods with poor air quality, seasonal biases in a particular ozone precursor or PM<sub>2.5</sub> component, etc.), additional diagnostic assessments may be needed. Analyses for diagnostic evaluations could include re-running the air quality model with different boundary conditions, emissions inputs, or meteorological inputs based on issues identified during the other portions of the platform evaluation. Testing various configuration options should be conducted in an iterative approach to better understand which inputs or parameters are most relevant to model performance. Similar graphical displays and/or statistical analyses outlined for the operational evaluation should be completed for the diagnostic evaluation. Two specific analyses that could be completed as part of the diagnostic evaluation include:

- Plots of Indicator Ratios: Ratios of species can be useful for assessing if the model accurately represents the relative emissions rates of precursors. For example, the model may be biased in simulating absolute concentrations of VOC and NO<sub>x</sub> due to errors in dispersion but may accurately represent the ratio VOC/NO<sub>x</sub>. Indicator ratios are useful for assessing whether the model accurately represents different chemical regimes of ozone sensitivity to VOC and NO<sub>x</sub> (see EPA, 2014 for examples).
- Process Analysis: Provides diagnostic information on the chemical reactions and physical processes that determine species concentrations. Processes include: Emissions, Deposition, Advection, Diffusion, Vertical Mixing, and Net Change from Chemistry at various time steps. For instance, plots comparing modeled ozone concentrations, output rates of ozone, daily total odd oxygen production rates, and net ozone production in NO<sub>x</sub> limited and oxidant limited chemical regimes.

## SUMMARIZING AND INTERPRETING RESULTS OF MODEL PLATFORM PERFORMANCE EVALUATION

The goal of the preceding evaluation methods is to assess the degree of confidence in the use of the model platform as a tool to inform the planning process, and more specifically, to determine how reliably the model predicts the response of ozone, PM<sub>2.5</sub> and deposition to changes in emissions. Once all of the graphical displays and statistical analyses are generated for the MPE associated with each model component, it is important to compile all of the analyses into comprehensive assessments that describe the strengths and weaknesses and how the model platform should be used in light of the evaluation findings.

This section outlines an approach for summarizing and interpreting the results of the MPE associated with each model component to assist the IWDW-WAQS (and other entities that complete an MPE analysis) in documenting the model performance and determining the adequacy of the model platform for use in NEPA and other air quality planning projects. The model performance assessments should include:

1. A stand-alone report that summarizes the assessment of the boundary conditions used in the modeling, including:
  - a. Description of the boundary conditions used in the modeling.
  - b. Description of the expected influence of boundary conditions on simulated concentrations within the domain.
  - c. Description of any diagnostic analyses completed to investigate and improve the boundary condition inputs.
2. A stand-alone report that summarizes the assessment of the emissions inventory/modeling, including:
  - a. Summary of the datasets and model input and configuration.
  - b. Sample graphical displays that characterize individual pollutants by various source categories across the model domain.
  - c. Comparison to emissions inventories used in previous studies (where applicable).
  - d. Discussion of any issues or concerns revealed from the characterization and assessment of the emissions inventory that warranted additional investigation.
3. A stand-alone report that summarizes the meteorological modeling MPE, including:
  - a. Summary of the meteorological model, its inputs and configuration, and discussion of any changes applied to the methodology or configuration since the air quality modeling protocol.
  - b. Summary of the observational networks to be used in the meteorological modeling evaluations. This should include descriptions of monitor locations and density, the periods in which measurements were taken, and any general concerns about the sufficiency of the data for comparisons against model predictions.
  - c. Statistical and graphical comparisons focused and partitioned into the following subsets (additional details provided in subsequent sections):
    - i. observational site
    - ii. geographic sub-regions
    - iii. daily/monthly/seasonal/diurnal time periods
    - iv. vertical structure (aloft, surface)
    - v. synoptic regime (where applicable)
  - d. Comparison to evaluations completed in other similar studies.
  - e. Discussion of the strengths and weaknesses indicated by the evaluation for various parameters, time periods, and locations. Any identified model limitations and uncertainties that may require further model development, improvement, and/or diagnostic evaluation should also be discussed.
4. A stand-alone report that summarizes the photochemical grid modeling MPE, including:
  - a. Summary of the model input and configuration, and discussion of any changes applied to the methodology or configuration since the air quality modeling protocol.

- b. Summary of the observational networks to be used in the photochemical grid modeling evaluations. This should include descriptions of monitor locations and density, the periods in which measurements were taken, and any general concerns about the sufficiency of the data for comparisons against model predictions. In particular, a discussion related to the PM and deposition measurement artifacts and uncertainties should be included to assist in interpreting the PM and deposition performance evaluation. EPA's guidance notes some of the known measurement artifacts for PM for different monitoring networks (USEPA, 2014, Table 3.2). A discussion related to the method for pairing the observations and model results should also be included because the data may not always be directly comparable or commensurable to the model results. See EPA's guidance (USEPA, 2014) for additional details.
- c. Statistical analyses and graphical displays focused on or partitioned into the following subsets:
- i. Averaging Period: Performance statistics and graphical displays for the report should be based on hourly averages and averages that correspond with the form of the air quality standard or metric being evaluated for each pollutant. For example, ozone NAAQS should evaluate episodes during which the monitoring data or modeled 1-hour and 8-hour average ozone concentration approaches or exceeds the NAAQS.
  - ii. Episode Selection: Episode selection is a useful approach in screening the results and limiting the number of days for which detailed analyses are needed for the report. Episode selection also assists in providing a more intensive evaluation on days most relevant to the NAAQS. The analyses should focus on:
    - High ambient and modeled concentrations. For ozone, the analyses should consider model performance on days where ambient concentrations are greater than 60 ppb to provide insight into how well the model replicates ozone pollution episodes.
    - Low ambient and modeled concentration for PM<sub>2.5</sub>. Analyses screened for low concentrations of PM<sub>2.5</sub> assist in evaluating the model for visibility and deposition.
    - Days not associated with abnormal conditions. For instance, photochemical grid models often perform poorly for unusual events such as stratospheric intrusions (SI), wildfires or dust storms. Therefore, it is important to determine the time periods influenced by infrequent events and potentially exclude the days associated with these events from the analyses when they can be reliably identified.
    - Days not influenced by certain meteorological conditions. Meteorological conditions strongly influence photochemical reactions, formation, and deposition of various pollutants. For instance, the formation of ozone is reduced on days with cloud cover or precipitation. Further, sulfate can be formed in cloud aqueous chemistry and aerosol nitrate formation may be greater on cool, humid days. Therefore, it may be useful to associate model performance by days with and without cloud cover, unusually heavy precipitation episodes, or other meteorological conditions.
    - Days without certain surface characteristics. For instance, snow cover and surface albedo, are critical input for simulating winter oxidant formation, including ozone, NO<sub>2</sub>, and nitric acid. Therefore, it may be useful to screen the model performance into days and seasons with and without snow cover. Identifying the days associated with these episodes assists in determining the factors influencing the performance of the model.
  - iii. Evaluation of Model Performance at Monitoring Sites: Analyses of ozone and PM<sub>2.5</sub> should be evaluated at individual monitoring sites based on hourly averages and averages that correspond with the form of the air quality standard. Monitor sites selected for the evaluations should be relevant to the given model application. Analyses that average the model performance at multiple sites or multiple days are not very useful because biases can cancel out over multiple sites and long-term averages will reflect the model performance on days that may not be relevant to the NAAQS. Evaluations of performance over long time periods and large areas may also fail to distinguish specific episodes for which the model either performs especially well or especially poorly.

- d. Comparison to evaluations completed in other similar studies (where appropriate).
  - e. Discussion of the strengths and weaknesses indicated by the evaluation for various pollutants, time periods, and locations. In particular, the evaluation should outline whether the model platform can or cannot capture key processes to events or episodes important for the air quality application. For instance, the evaluation should summarize the model performance for high ozone and PM<sub>2.5</sub> events for supporting NAAQS demonstrations and visibility analyses. Any identified model limitations and uncertainties that may require further model development, improvement, and/or diagnostic evaluation should also be discussed.
5. A stand-alone memo that compiles and summarizes the results of the individual model component assessments (i.e., BC, EI, Meteorological, and PGM assessments). In particular, the memo should outline the overall strengths and weaknesses of the platform and any limitations that may require further model development, improvement, and/or diagnostic evaluation. The memo should also recommend the air quality applications that the model can or cannot be used for and how the evaluations could be used as a weight of evidence for certain applications. This recommendation should be based on:
- a. Comparisons to similar modeling exercises to ensure that the model performance approximates the quality of other applications. A recent literature review (Simon et al, 2012) summarized photochemical model performance for applications published in the peer-reviewed literature between 2006 and 2012. This review may serve as a resource for identifying typical model performance for state of the science modeling applications.
  - b. Performance statistics, which can provide information about the general model performance and identify model limitations and uncertainties that may require further model development, improvement, and/or diagnostic evaluation.
  - c. Specific benchmarks should not be applied to the statistical analyses because of concerns about potentially misleading comparisons of model performance findings across different analyses with differing model configurations and masking performance issues that may only be made apparent through graphical displays. Historically, ozone and PM<sub>2.5</sub> benchmarks have been commonly used in NEPA air quality analyses to determine whether the modeling platform “passes” or is acceptable for predicting future air quality impacts.

This memo should be reviewed and approved by the IWDW-WAQS Technical and Oversight Committees prior to releasing the modeling products for NEPA and air planning projects.

#### REFERENCES:

- DRAFT Modeling Guidance for Demonstrating Attainment of Air Quality Goals for Ozone, PM<sub>2.5</sub>, and Regional Haze [December 2014]  
[http://www.epa.gov/ttn/scram/guidance/guide/Draft\\_O3-PM-RH\\_Modeling\\_Guidance-2014.pdf](http://www.epa.gov/ttn/scram/guidance/guide/Draft_O3-PM-RH_Modeling_Guidance-2014.pdf)
- Guidance on the Use of Models and Other Analyses for Demonstrating Attainment of Air Quality Goals for Ozone, PM<sub>2.5</sub>, and Regional Haze [April 2007]  
<http://www.epa.gov/ttn/scram/guidance/guide/final-03-pm-rh-guidance.pdf>
- Compilations and Interpretation of Photochemical Model Performance Statistics Published between 2006 and 2012 [Simon, H., K. Baker and S. Phillips (2012), *Atmos. Env.*, 61, 124-139]  
<http://www.sciencedirect.com/science/article/pii/S135223101200684X>
- Comparison of Western U.S. Ozone using MOZART and GEOS-Chem for Boundary Conditions, ENVIRON, Western Regional Air Partnership, July 10, 2013  
[http://www.wrapair2.org/pdf/Morris\\_MOZART-GEOS\\_WestJumpAQMS\\_Jul10\\_2013\\_Draft1.pdf](http://www.wrapair2.org/pdf/Morris_MOZART-GEOS_WestJumpAQMS_Jul10_2013_Draft1.pdf)