Technical Memorandum
Additional Performance Evaluation of Dispersion Modeling Approaches
Miami SO_2 Nonattainment Area State Implementation Plan (SIP)
February 5, 2018

A performance evaluation was previously completed on five air quality dispersion model approaches for use in the Arizona Department of Environmental Quality's (ADEQ) Miami sulfur dioxide (SO_2) Nonattainment Area State Implementation Plan (SIP) submittal to the U.S. Environmental Protection Agency (EPA). This previous performance evaluation was summarized in a Technical Memorandum dated August 11, 2015, and included in the Miami SO_2 Nonattainment Area SIP Revision Attainment Demonstration Technical Support Document (TSD) as Appendix C. Additional analysis of model performance has since been conducted using the procedures documented in EPA's 1992 guidance, *Protocol For Determining The Best Performing Model*. This memo summarizes the additional analysis performed and its results.

Introduction

The additional analysis recommended in EPA's guidance involves two steps. The first step is a screening test to eliminate models that fail to perform at a minimum operational level. The second step applies only to those models that pass the screening test and is based on composite measures of performance that integrate both the operational and scientific components of model performance.

The additional analysis was performed for each of five modeling approaches based on implementation of EPA's preferred "BLP"[1] and "AERMOD"[2] dispersion models, both of which have features relevant to modeling the smelter:

- Additive BLP/AERMOD, Multi-Vent BLP Plume Rise
- Additive BLP/AERMOD, Single-Vent BLP Plume Rise
- Hybrid BLP/AERMOD
- AERMOD, Roofline Vents as point sources with Downwash
- AERMOD, Roofline Vents as point sources without Downwash

These approaches are described in further detail in the August 2015 performance evaluation.

The additional analysis was applied to observations from the three ambient SO_2 monitoring locations that operate around the FMMI facility: Jones Ranch, Ridgeline, and Miami Townsite. These monitoring locations are described in further detail in the August 2015 performance evaluation. The additional analysis, like the August 2015 performance evaluation, focused solely on 1-hour average SO_2

[1] BLP is a Gaussian plume dispersion model designed to handle unique modeling problems associated with industrial sources where buoyant plume rise and downwash effects from stationary line sources are important. With EPA's proposed changes to AERMOD, EPA is also proposing to delist BLP as a preferred model.

[2] AERMOD is a steady-state plume model that incorporates air dispersion based on planetary boundary layer turbulence structure and scaling concepts, including treatment of both surface and elevated sources, and both simple and complex terrain.

concentrations. Due to the nature of the standard, the highest 3 observed and predicted values were not considered in this analysis.

Screening Test

Fractional bias is used as the performance measure in the screening test and is calculated for each competing model as follows:

$$FB = 2\left[\frac{OB-PR}{OB+PR}\right] \quad (1)$$

where *FB* = fractional bias,
OB = mean or standard deviation of the 25 highest observed values, and
PR = mean or standard deviation of the 25 highest predicted values.

The fractional bias of both the mean and standard deviation is calculated, and the analysis is limited to the upper tails of both observed and predicted values (i.e., the highest 25 values). Due to the form of the equation, models that over-predict will have negative values of fractional bias while models that under-predict have positive values of fractional bias. Values of fractional bias between -0.67 and +0.67 are desired, as such values indicate model predictions are within a factor of two of observations.

Figure 1 provides a graphical representation of the screening test results. The modeling approaches with results that fall within the central box meet the screening test criteria of having calculated fractional bias values between -0.67 and +0.67. The results indicate that model performance for the Jones Ranch and Ridgeline monitor locations favors the Hybrid Approach, although the AERMOD Approach with no vent downwash is also a candidate approach for inclusion in the second step of the analysis. In contrast, the model performance for the Miami Townsite monitor location favors the AERMOD Approach with vent downwash as well as the Additive Approach where single-vent BLP plume rise is considered. Only one approach can be eliminated by the screening test: the Additive Approach where multiple-vent BLP plume rise is considered.

Composite Performance Measure

As noted previously, the second step is based on a composite measure of performance that integrates both operational and scientific evaluations of model performance. The purpose of the operational evaluation is to measure a model's ability to accurately predict concentrations most directly used for regulatory purposes. In the case of 1-hour SO_2 concentrations, the statistic of interest involves the network-wide highest concentrations. For the operational evaluation, the precise time, location, and meteorological condition is of minor concern compared to the magnitude of the observed and predicted highest concentrations. The purpose of the scientific evaluation is to measure a model's ability to perform accurately throughout (a) the range of meteorological conditions that might be expected to occur and (b) the geographic area immediately surrounding the monitor locations for which model estimates are needed. The performance measures obtained from the operational and scientific evaluations are combined to create a composite performance measure.

The operational and scientific evaluations are performed using a robust estimate of highest concentration (RHC). The RHC is based on a tail exponential fit to the upper end of the observed and predicted distributions and is calculated as follows:

$$RHC = X(n) + [\bar{X} - X(n)] \ln\left[\frac{3n-1}{2}\right] \tag{2}$$

where *RHC* = robust highest concentration,
X(n) = n^{th} largest value,
$\bar{X}$ = average of the *n*-1 largest values, and
n = number of values exceeding the threshold value ($n \leq 26$).

The value of *n* is arbitrarily set at a value of 26 or less such that the number of values averaged is no greater than 25. The threshold value is defined as a concentration near background which has no impact on the determination of the *RHC*. In cases where the value of *n* is less than 3, the *RHC* is set equal to the threshold value.

Operational Evaluation

The operational evaluation compares the performance of the models in terms of the largest network-wide RHC, which is calculated separately for each ambient monitoring location within the network. The largest observation-based RHC in the monitoring network and the largest prediction-based RHC value for each model are used to calculate absolute fractional biases (AFB) for each model.

Table 1 presents the RHC and AFB values calculated for the operational evaluation. The results show that the Hybrid Approach performs the best under the operational evaluation.

Table 1. Calculated RHC and AFB, Operational Evaluation

Description	**RHC (µg/m³)**	**AFB_O**
Observed, Actual Measurements	537	--
Predicted, Multi-Vent Additive BLP/AERMOD	1,642	0.847
Predicted, Single-Vent Additive BLP/AERMOD	2,356	1.119
Predicted, Hybrid BLP/AERMOD	547	0.195
Predicted, AERMOD, Roofline Vents with Downwash	1,609	0.830
Predicted, AERMOD, Roofline Vents without Downwash	342	0.641

Scientific Evaluation

The scientific evaluation compares the performance of the models in terms of the RHC, which is calculated separately for each ambient monitoring location and each of six meteorological conditions as follows:

- Wind speed less than 4.0 meters per second, stable atmospheric conditions (Pasquill-Gifford classifications E and F),
- Wind speed less than 4.0 meters per second, neutral atmospheric conditions (Pasquill-Gifford classification D),
- Wind speed less than 4.0 meters per second, unstable atmospheric conditions (Pasquill-Gifford classifications A, B, and C),
- Wind speed greater than or equal to 4.0 meters per second, stable atmospheric conditions (Pasquill-Gifford classifications E and F),

- Wind speed greater than or equal to 4.0 meters per second, neutral atmospheric conditions (Pasquill-Gifford classification D),
- Wind speed greater than or equal to 4.0 meters per second, unstable atmospheric conditions (Pasquill-Gifford classifications A, B, and C),

The largest observation-based RHC in the monitoring network for each meteorological condition and the largest prediction-based RHC value for each model and meteorological condition are used to calculate absolute fractional biases (AFB) for each model and meteorological condition.

Table 2 presents the RHC and AFB values calculated for the scientific evaluation. The results show that the different models perform best under different meteorological conditions.

Integration of Operational and Scientific Performance Evaluations

A composite performance measure is computed for each model as a weighted linear combination of the individual fractional bias components. Because the operational evaluation is deemed by EPA to be more important than the scientific evaluation, the operational evaluation is given a weight that is twice the weight of the scientific evaluation. Within the scientific evaluation, each of the combinations of the meteorological conditions is given equal weight. The composite performance measure (CPM) is therefore calculated as follows:

$$CPM = \frac{2}{3} AFB_O + \frac{1}{3} AFB_S \tag{3}$$

where CPM = composite performance measure,
AFB_O = absolute fractional bias calculated in the operational evaluation, and
AFB_S = average of the absolute fractional biases calculated for each meteorological condition in the scientific evaluation.

Table 3 presents the calculated CPM for each model. The Hybrid Approach has the lowest CPM and is the only approach where the AFB calculated in each evaluation is less than 0.67.

Conclusion

The additional analysis confirms that the Hybrid Approach is the selected approach for identifying the smelter critical emissions value (CEV) because the model performs best in consideration of composite performance measures for the three ambient monitoring locations. The alternative models are unacceptable because the calculated statistics do not meet the criteria established by EPA.

Table 2. Calculated RHC and AFB, Scientific Evaluation

Description	RHC (µg/m³)	AFB_S
Observed, Actual Measurements		
Wind Speed < 4.0 m/s, Stable Conditions	486	--
Wind Speed < 4.0 m/s, Neutral Conditions	192	--
Wind Speed < 4.0 m/s, Unstable Conditions	405	--
Wind Speed ≥ 4.0 m/s, Stable Conditions	146	--
Wind Speed ≥ 4.0 m/s, Neutral Conditions	220	--
Wind Speed ≥ 4.0 m/s, Unstable Conditions	214	--
Predicted, Multi-Vent Additive BLP/AERMOD		**0.724 (avg.)**
Wind Speed < 4.0 m/s, Stable Conditions	1,559	0.867
Wind Speed < 4.0 m/s, Neutral Conditions	179	0.606
Wind Speed < 4.0 m/s, Unstable Conditions	121	1.253
Wind Speed ≥ 4.0 m/s, Stable Conditions	602	0.944
Wind Speed ≥ 4.0 m/s, Neutral Conditions	402	0.319
Wind Speed ≥ 4.0 m/s, Unstable Conditions	200	0.356
Predicted, Single-Vent Additive BLP/AERMOD		**0.691 (avg.)**
Wind Speed < 4.0 m/s, Stable Conditions	2,368	1.174
Wind Speed < 4.0 m/s, Neutral Conditions	321	0.041
Wind Speed < 4.0 m/s, Unstable Conditions	228	0.792
Wind Speed ≥ 4.0 m/s, Stable Conditions	620	0.968
Wind Speed ≥ 4.0 m/s, Neutral Conditions	568	0.645
Wind Speed ≥ 4.0 m/s, Unstable Conditions	167	0.530
Predicted, Hybrid BLP/AERMOD		**0.510 (avg.)**
Wind Speed < 4.0 m/s, Stable Conditions	547	0.119
Wind Speed < 4.0 m/s, Neutral Conditions	164	0.686
Wind Speed < 4.0 m/s, Unstable Conditions	95	1.392
Wind Speed ≥ 4.0 m/s, Stable Conditions	199	0.083
Wind Speed ≥ 4.0 m/s, Neutral Conditions	316	0.083
Wind Speed ≥ 4.0 m/s, Unstable Conditions	139	0.698
Predicted, AERMOD, Roofline Vents with Downwash		**0.581 (avg.)**
Wind Speed < 4.0 m/s, Stable Conditions	344	0.568
Wind Speed < 4.0 m/s, Neutral Conditions	263	0.237
Wind Speed < 4.0 m/s, Unstable Conditions	189	0.945
Wind Speed ≥ 4.0 m/s, Stable Conditions	369	0.525
Wind Speed ≥ 4.0 m/s, Neutral Conditions	458	0.445
Wind Speed ≥ 4.0 m/s, Unstable Conditions	129	0.762
Predicted, AERMOD, Roofline Vents without Downwash		**0.949 (avg.)**
Wind Speed < 4.0 m/s, Stable Conditions	1,573	0.874
Wind Speed < 4.0 m/s, Neutral Conditions	658	0.652
Wind Speed < 4.0 m/s, Unstable Conditions	202	0.892
Wind Speed ≥ 4.0 m/s, Stable Conditions	574	0.907
Wind Speed ≥ 4.0 m/s, Neutral Conditions	565	0.641
Wind Speed ≥ 4.0 m/s, Unstable Conditions	21	1.725

Table 3. Calculated CPM, Integrated Evaluation

Model	**CPM**	**AFB_O**	**AFB_S**
Multi-Vent Additive BLP/AERMOD	0.806	0.847	0.724
Single-Vent Additive BLP/AERMOD	0.976	1.119	0.691
Hybrid BLP/AERMOD	0.300	0.195	0.510
AERMOD, Roofline Vents with Downwash	0.747	0.830	0.581
AERMOD, Roofline Vents without Downwash	0.743	0.641	0.949

Figure 1. Screening Test Results for Each Ambient Monitor Location

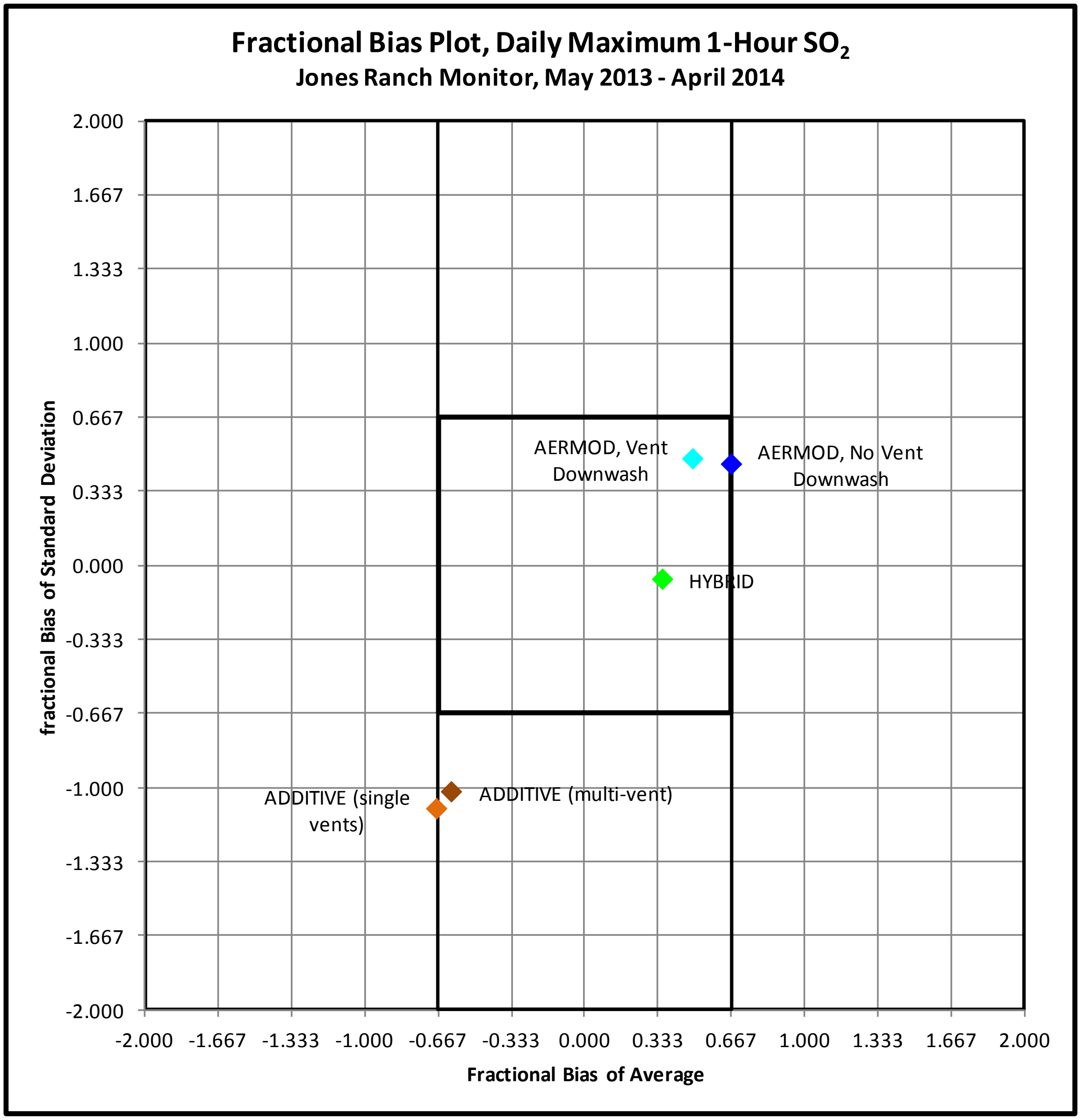

Figure 1. Screening Test Results for Each Ambient Monitor Location (Continued)

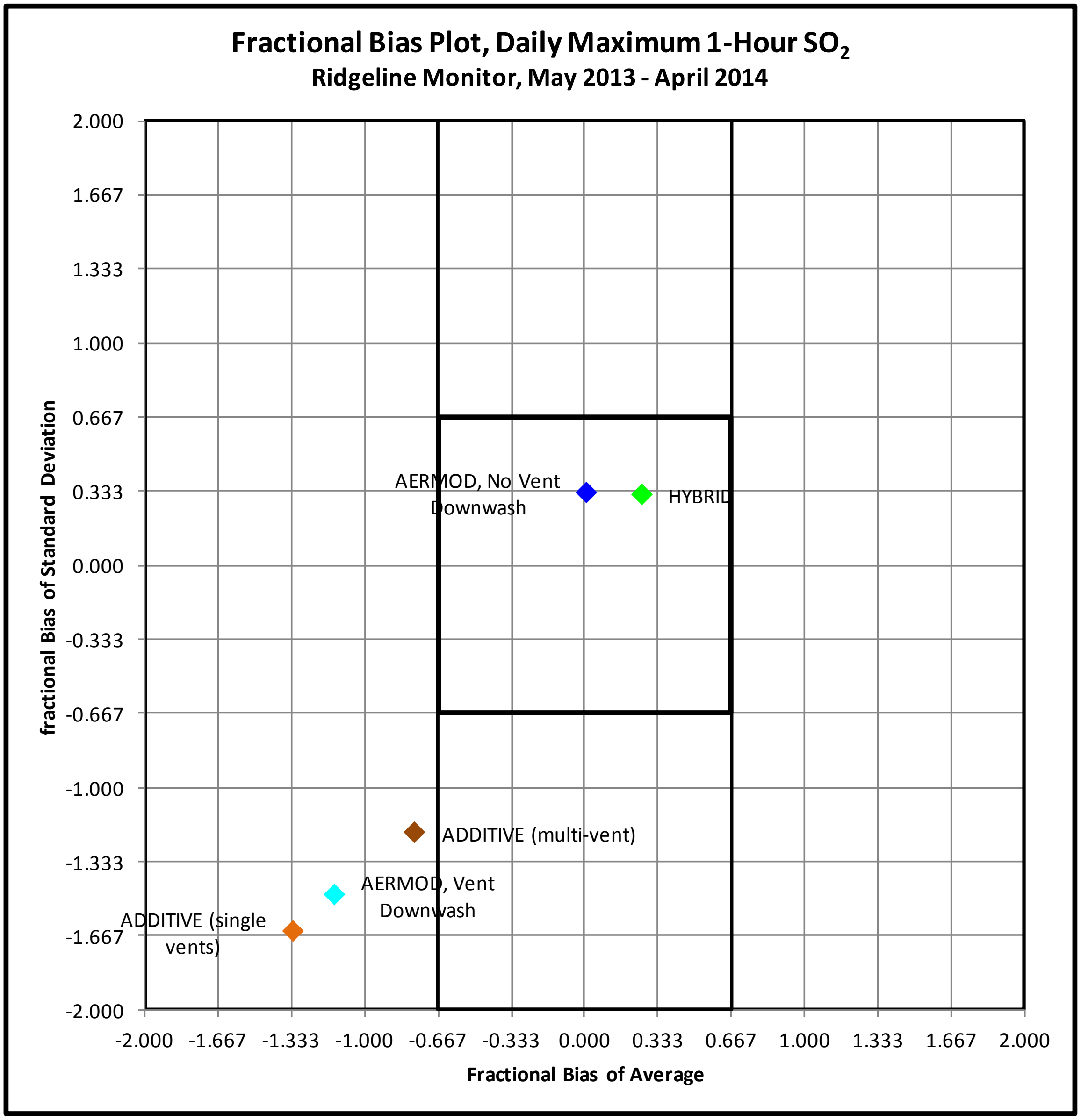

Figure 1. Screening Test Results for Each Ambient Monitor Location (Continued)