NCHRP Project 25-55

NCHRP Research Report 1058
Assessing Air Pollution Dispersion Models for Emissions Regulation

Appendix D
Enhancing the Model Evaluation Process for the Transportation Sector

The National Cooperative Highway Research Program (NCHRP) is sponsored by the individual state departments of transportation of the American Association of State Highway and Transportation Officials. NCHRP is administered by the Transportation Research Board (TRB), part of the National Academies of Sciences, Engineering, and Medicine, under a cooperative agreement with the Federal Highway Administration (FHWA). Any opinions and conclusions expressed or implied in resulting research products are those of the individuals and organizations who performed the research and are not necessarily those of TRB; the National Academies of Sciences, Engineering, and Medicine; the FHWA; or NCHRP sponsors.

EXECUTIVE SUMMARY

This report provides recommendations for the establishment of an enhanced model evaluation process (MEP) for transportation projects.

The motivation for these recommendations follows the introduction of quantitative project-level air quality modeling requirements with the publication of EPA's "Transportation Conformity Guidance for Quantitative Hot-spot Analyses in PM2.5 and PM10 Nonattainment and Maintenance Areas" December 2010. Similar guidance was introduced for carbon monoxide in 2010 but building upon previous near roadway modeling guidance published in 1992. The new requirements were based on the use a preferred dispersion model (AERMOD) and the need for a complete modeling chain for use in regulatory analyses. The modeling chain includes traffic, emissions, and dispersion modeling and inclusion of background concentrations. This requirement was made without testing and evaluation using field data for typical transportation projects. Typical projects include high-volume freeways with and without noise walls, congested intersections with high diesel truck traffic, and truck and bus terminals. Field data would include tracer data used to assess the dispersion model and simultaneous, near-source air quality monitoring data for carbon monoxide (CO), PM, and potentially other pollutants that would be used to assess the entire regulatory modeling chain. The latter is needed to test if the required models and associated guidance adequately meet the intended regulatory purpose of project level modeling. That is, showing compliance with regulatory tests such as build/no-build (B/NB) tests and against applicable national ambient air quality standards (NAAQS) with statistical confidence.

The core purpose of this document is to provide general principles for evaluation of transportation air quality models, and a model evaluation plan consistent with these principles. Examples supporting the development and testing of the recommended model evaluation process (MEP) are provided as appendices. Attachment A provides an outline for a generic model evaluation. Attachment B applies that outline to evaluate AERMOD and other dispersion models against tracer data as well as the regulatory modeling chain against near-road monitoring data. A comparison of the results for the dispersion model (which showed good agreement with the tracer data) and the regulatory modeling chain (which showed less agreement with the data) indicates that the dispersion model is not the primary source of error in the modeling chain. The data used for this evaluation were obtained by the Berkeley Freeway Experiment – a field campaign undertaken as part of this NCHRP 25-55 research project for one of the highest priority transportation applications previously identified: a high-volume freeway without noise walls. This Experiment is described in attachment B.

The following summarizes the conclusions and recommendations obtained from the field studies, modeling, and analyses conducted in this NCHRP project:

Scope and purpose of the MEP

- An enhanced MEP is needed for transportation projects to be consistent with refined regulatory modeling requirements introduced in 2010. This is driven by the introduction of these requirements because of the limited validation of the dispersion models and even less for evaluation of the entire regulatory modeling chain against near-road monitoring data.

- Model evaluation for the regulatory modeling chain includes traffic, emissions and dispersion modeling and determination of representative background concentrations. Thus, such an evaluation requires both tracer studies and co-located observations of traffic-related air pollutants and their background concentration. Field studies supporting this should include the collection of both tracer and near-road air quality monitoring data.
- The use of co-located tracer and near-road monitoring sites as done in project 25-55 is critical to allow comparison of the modeling results for the dispersion component and the regulatory modeling chain. Comparison of the separate performance evaluation of these items can help to identify weaknesses in the overall chain. That is, if for example the dispersion model performs well, and traffic data and background concentrations are well characterized (i.e., are site-specific), then the focus for model improvements to improve accuracy and reduce uncertainty may be directed to the emission modeling step in the regulatory modeling chain.
- Identification and prioritization of the transportation applications (facilities, configurations, operating conditions, and setting) is needed. For cost-effectiveness and efficiency, consideration should be given to how to combine field studies, which are resource-intensive, for priority applications, e.g., combing field studies for a high-volume freeway arterial street interchange with adjacent congested intersections with diesel truck traffic instead of conducting two separate studies, although doing so will need to be carefully designed to avoid interference.

Findings from Application of the MEP

- In the tracer comparisons, results from different models and source configurations varied only slightly across tracers and the relative performance of the models was consistent across tracers. This implies that the tracers used here can be successfully used in mobile source experiments and that a single tracer could be used in future studies.
- For the tracer comparison all four models showed a tendency for underpredictions, as demonstrated by positive fractional bias values. However, this underprediction tends to occur mostly at lower concentrations that occur at further distances downwind from the roadway. This implies model performance is best near the source, where the predicted concentration is most critical to project evaluation. This is important, as the peak concentration is typically the deciding factor in regulatory applications, such as B/NB tests and NAAQS evaluation.
- All of the dispersion models' performance metrics were better for tracer data than for the pollutant-based modeling chain compared to near-road ambient air quality monitoring for $PM_{2.5}$ and CO. This implies other components of the modeling chain are limiting elements of overall performance and should be addressed in future evaluations.
- All of the dispersion models and their source configurations tended to underestimate individual hourly concentrations of CO for the pollutant-based approach. RLINE and ADMS-ROADS also underestimated $PM_{2.5}$, although AERMOD (in both volume and area configurations) tended to slightly overpredict for $PM_{2.5}$. Most $PM_{2.5}$ predictions were within a factor of 2 of observed, while most CO predictions were not. All models show low r-squared values in comparison to the tracer experiments suggesting much of the

uncertainty is associated with the emission inputs s used in the modeling, although background may play a role**.**

- For the regulatory comparison all four models/configurations showed design values that were close to observed values. However more than 80 percent of modeled design values were from the background concentrations, so only about 20 percent of modeled design values were due to near-receptor, traffic-related emissions. The findings are also dependent on both the emission factor model and background concentration making definitive conclusion on the contribution to discrepancy difficult to conclude. However, this test is most like regulatory use of these models.
- NCHRP 25-55 did not include a formal comparative analysis of different model formulations or technical application guidance. However, the following findings and insights were made based on the results of the statistical analyses:
- Model comparisons conducted for NCHRP 25-55 showed better performance for AERMOD when applied using AREAPOLY source type than RLINE. This is likely a result of the current EPA guidance on estimating release height and initial vertical dimension being developed prior to the incorporation of RLINE into the AERMOD model.
- Additionally, as noted by Gilles et. al (2005), in addition to atmospheric stability, the shape of the vehicle, and the angle of the ambient wind with respect to the direction of vehicle travel play a role in determining the vehicle scaling height factor. Further study is needed to better estimate this scaling factor.

Recommendations

- This enhanced MEP should be implemented for the transportation sector for future field experiments and model evaluations, including the use of co-located tracer and near-road ambient air quality monitoring. This MEP serves the objective for determining a scientifically rigorous and systematic approach for testing the adequacy of the regulatory modeling chain for its intended regulatory purpose for showing compliance for applicable regulatory tests (NAAQS compliance and Build/No-build analysis).
- Future model evaluations should identify and prioritize other transportation applications (facilities, configurations, operating conditions, and setting) not addressed by existing field studies. Potential priority settings to collect co-located tracer and near-road monitoring data include:
    - High-volume congested intersections with high percentages of diesel truck and/or bus traffic
    - High-volume freeway- interchanges high percentages of diesel truck and/or bus traffic
    - Truck and/or bus terminals with high volumes of diesel vehicles
    - Freeway segments with higher road grades
    - Arterial segments at-grade and with higher road grades

  - High volume freeways at-grade and with higher road grades, without and with noise walls, in settings distinct from those of the Berkeley Freeway Experiment.

- Future model evaluations should continue to focus on the full modeling chain in evaluation and evaluate performance in terms of regulatory use (B/NB test and NAAQS evaluation).
- The monitoring study was performed in California and used California's emissions factor model, EMFAC. Future studies should be conducted outside of California using MOVES and modeling approaches for pollutants-based and regulatory evaluations.
- Given the complexities of the scope of work and extensive stakeholder consultation needed in such field studies and model evaluation work, future work would most effectively be performed under a transportation pooled fund framework.
- Prioritize model validation for PM (especially $PM_{2.5}$) over other pollutants given typically high background concentrations relative to the applicable NAAQS, which places a premium on model accuracy for regulatory applications (i.e., modeling design values with statistical confidence sufficient to determine compliance with the NAAQS and B/NB tests).
- A detailed strength and weakness model comparison analysis should be included in future model evaluations. This was not included here due to limited resources.
- Other areas of focus for future research include:
  - Determining the appropriate release height and hence initial dispersion parameters for both cars and trucks by measuring the vertical concentration profile under a variety of meteorological conditions and vehicle types.
  - Researching emission models at project level, including determining uncertainty and the underlying basis for the uncertainty in emissions model predictions and developing recommendations for improving the project level emissions. This will require separating the various components of the particulate matter that is generated in the near roadway environment –resuspended road dust, brake and tire wear, and exhaust as well as the fleet characterization.
  - Research and guidance development to improving the determination of background concentrations for project level analyses.

Contents

1. Introduction

Transportation projects have unique circumstances and requirements for approval in an air quality regulatory setting. Depending on the scale and scope of the project, they can be valued in hundreds of millions, if not billions, of dollars. In addition, air quality considerations occur relatively late in the design process and are subject to various legal and regulatory requirements. The air quality analyses performed for these projects can result in significant delays, changes in the scope and design of the project, or even (in rare cases) scrapping of the project if the air quality analysis indicates the potential for significant impacts.

For these reasons, any air quality dispersion model used to estimate the potential for such air quality impacts of transportation projects must undergo a rigorous, comprehensive, systematic, credible, and transparent process to show that it is sufficiently accurate for the intended regulatory application before it is approved or required for use in a transportation project air quality analysis. This is particularly needed for pollutants for which the margins between the applicable national ambient air quality standards (NAAQS) and background concentrations are relatively small, such is often the case for fine particulate matter ($PM_{2.5}$), thus demanding the need for accuracy in modeling to be able to show compliance with the NAAQS and pass Build/No-Build tests with statistical confidence.

Dispersion modeling requirements for both $PM_{2.5}$ and coarse particulate matter (PM_{10}) were concurrently introduced in 2010 with the release of the MOVES emission model. One of the specified air quality dispersion models for use in PM analyses was AERMOD, with which most state DOTs limited experience as it had not previously been required for transportation projects. It was therefore only later realized by state DOTs that AERMOD, though now required for transportation applications, had not been validated for the full range of typical facility types, configurations, operating conditions, and settings (urban/rural) for which the models were now required by regulation to be applied but only against tracer data for an empty field (i.e., no roadway) and a relatively low volume roadway. In other words, it had not been validated for its full range of intended regulatory applications or purposes for transportation, including high-volume freeways with and without noise walls, congested intersections with high percentages of diesel trucks, truck and bus terminals, etc. Additionally, the project-level modeling chain (traffic, emissions and dispersion including the determination of representative background concentrations) had not been validated at all for their intended regulatory applications, and there was no mechanism in place or planned to do so.

This raised questions about the accuracy of the models and the project-level modeling chain for their intended regulatory applications in transportation and particularly so given the typical small margins between background concentrations for PM and the applicable NAAQS. Resolving these issues requires the collection for appropriate field data and model validation of both dispersion models and the modeling chain for the full range of typical transportation applications to determine those for which the models may be applied with statistical confidence and those for which limitations to their application may be warranted.

The question of how to evaluate air quality dispersion models has been debated for decades. A number of recent studies or reports (outlined below) together underline the need for the development of an *enhanced* model evaluation program (MEP) for transportation, one that goes further than current practice and enables DOTs to have greater confidence in the modeling results for both dispersion models and the entire regulatory modeling chain for the full range of typical transportation applications (project types, configurations, operating conditions, and settings) for each pollutant for which near-road modeling is required.

In 2007, the National Research Council published "*Models in Environmental Regulatory Decision Making*" (NRC 2007). This study made general recommendations for model evaluation; principles for model development, selection, and application; and model management. Key recommendations are described below.

- *Model evaluation* is "… the process of deciding whether and when a model is suitable for its intended purpose", involves "peer review, corroboration of results with data and other information, quality assurance and quality control checks, uncertainty and sensitivity analyses, and other activities", and "…should continue throughout the life of a model." Note the evaluation of models for their intended regulatory purpose requires field data for all typical applications (project type etc.) in order to test the models in regulatory applications.
- *Peer review* "…requires an effort commensurate with the complexity and significance of the model application", which translates to peer review needs for transportation applications.
- *Quantifying and communicating uncertainty* require addressing both the "…kinds of analyses [that] should be done to quantify uncertainty, and how these uncertainties should be communicated to policy makers."
- *General principles for model development, selection, and application* of interest here include model parsimony ("Models used in the regulatory process should be no more complicated than is necessary to inform regulatory decisions"), caution on extrapolation (or going "beyond conditions for which the model was constructed or calibrated or conditions for which the model outputs cannot be verified"), and a preference for non-proprietary models for environmental modeling.
- *Model management* involves the institution of "…best practice standards for the evaluation of regulatory models" (the development of which for transportation applications is the underlying intent of this study) and ensuring model accessibility.

In 2016, NASA issued "*Standard for Models and Simulations*" (NASA Technical Standard NASA-Std-7009a W/ Change 1: Administrative/ Editorial Changes 2016-12-07) (NASA 2016). While this standard was not written for environmental or air quality analyses specifically, its comprehensive approach is generally consistent with recommendations of the 2007 NRC report as it includes planning for the lifecycle of the models, model validation, characterization of uncertainty, sensitivity analyses, and reporting to decision-makers. It effectively demonstrates the successful implementation of those recommendations.

The US EPA Inspector General issued a report "*EPA Can Strengthen its Process for Revising Air Quality Dispersion Models that Predict Impact of Pollutant Emissions*" (USEPA 2018). It concluded in part that "the quality assurance and control activities undertaken for these revisions [to AERMOD from 2006 through 2016 were not as extensive as what EPA guidance recommends for new model development and evaluation." This lends weight to the need for enhancements to the current model evaluation process for AERMOD.

A summary paper "*Near-Road Air Quality Insights from a US DOT Five-Year Transportation Pooled Fund Study*" (Eisinger, et al. 2021) presented results from two case studies conducted for the Transportation Pooled Fund for Near-Road Air Quality. The results raised questions with the accuracy of the regulatory modeling chain (background, traffic, emissions, and dispersion) for project-level air quality analyses, again lending weight to the need for an enhanced model evaluation process for AERMOD and the entire regulatory modeling chain. The latter is needed to assess the adequacy of the project-level modeling chain and associated guidance in being able to meet the intended regulatory purposes of showing compliance with the NAAQS and passing the Build/No-Build (B/NB). Additional detail is provided in Craig et al. (2020). (Reference 6).

In these case studies, model-to-monitor comparisons were conducted for high volume freeway segments in Indianapolis and Providence using near-road monitoring data for $PM_{2.5}$. The AERMOD dispersion modeling results showed substantial overestimation for the average concentration for near-road $PM_{2.5}$ concentrations. Road dust emissions accounted for about 50% of the emissions, followed by exhaust, brake and tire wear.   However, since the case studies did not involve the use of data from co-located tracer and near-road monitoring data, they could not be used to clearly identify the source of the overestimation as either the background, emissions, or the dispersion modeling step. Note, as the traffic inputs to the emission model were based on detailed local traffic data collected for this purpose, their contribution if any to the overestimates for near-road concentrations would be minor. For Indianapolis, AERMOD was run for 152 analysis days in 2016. The average modeled PM2.5 near-road increments for these days were compared to the monitored near-road PM2.5 increments. The average modeled increment (3.7 $\mu g/m^3$) was three to four times larger than the average measured increments obtained from Federal Reference Method (FRM) or Federal Equivalent Method (FEM) monitoring instruments (1.2 $\mu g/m^3$ for FRM and 0.9 $\mu g/m^3$ for FEM). However, these biases appeared less of a concern for regulatory NAAQS comparison which uses the 98th percentile of the modeled concentration which was modeled as 6.1 $\mu g/m^3$ and the monitors reported values of 5.7 $\mu g/m^3$ for FRM and 7.9 $\mu g/m^3$ for FEM.

A similar modeling exercise was performed using AERMOD for Providence for 2015–2016 for 382 analysis days. The AERMOD-based analysis for Providence also showed an overpredicted bias for the average measured near-road PM2.5 increment. The average modeled PM2.5 increment (8.8 µg/m3) was more than six times, or 530 percent, greater than the average measured increment (1.4 $\mu g/m^3$) however this included a number of days with nearby construction. The bias is less when looking at only non-construction days and using the regulatory 98th percentile for NAAQS compliance here the 98th percentile modeled concentration was 10.8 $\mu g/m^3$ and the monitored value was 4.5 $\mu g/m^3$.

In the report "*AERMOD Source Types RLINE and RLINEXT Testing*" (Wayson and Voigt, 2022) results were presented from testing of the new alpha feature RLINEXT added to AERMOD by EPA to enable testing the model for noise walls near highways. The studied showed a number of deficiencies with the new feature, including applications involving a high-volume freeway with a noise wall, and noted as high-priority issue that AERMOD had not yet been validated against field data for high volume freeways or ones with noise walls. More broadly, the regulatory modeling chain of traffic, emissions, and dispersion (including the determination of background concentrations) has not been validated for its intended regulatory purpose of showing compliance with the NAAQS and Build/NoBuild (B/NB) tests for the full range of typical transportation applications for each pollutant for which regulatory project-level air quality analyses are required.

The report therefore applied the concepts from the NRC 2007 study to make recommendations for an *enhanced* model validation program for transportation projects that includes both assessments of dispersion models against tracer data (which is already standard practice) and, in order to also test the models for their intended regulatory purposes (showing compliance with statistical confidence in the NAAQS and B/NB tests), assessments of the full traffic, emission and dispersion modeling chain (including the determination of background concentrations) against near-road monitoring data. Typical transportation applications for which the recommended enhanced model validations would be needed include the full array of typical transportation project types (highway, street intersection, urban canyon, truck and bus terminals etc.), configurations (with or without noise walls, high road grades etc.), operating conditions (congested or uncongested), and setting (urban or rural).

The recommendations for the MEP for transportation projects presented in this document first discuss the general principals of enhanced model evaluation. The following section includes a discussion of the MEP, analysis of results, conclusions, and recommendations. This is followed by a discussion of the selection of the regulatory applications considered for model evaluations. We then present summary conclusions and recommendations. Finally, we present an outline for a sample model evaluation in Attachment A for reference. Attachment B presents a sample assessment based on the work done in NCHRP 25-55. For the latter, we note that this is a partial evaluation only. Given resource constraints of this project, completion of the field experiment and model performance evaluation were prioritized.

2. General Principles and Assumptions for Enhanced Model Evaluation for Transportation Projects

The following general principles are for an *enhanced* model evaluation process (MEP) for project-level air quality analyses for transportation that is comprehensive, rigorous, and systematic and, based on recommendations discussed in the introduction. This includes validating not only the dispersion model, but the entire modeling chain (traffic, emissions and dispersion including the determination of background concentrations). Additional field studies are needed to cover the full range of typical transportation applications and implement recommended enhanced model evaluation program. Ultimately, state DOTs and other transportation agencies will benefit from improved model performance and confidence in regulatory applications for transportation to the extent that these principles are incorporated into future model evaluation processes.

2.1 OBJECTIVE: DETERMINATIONS OF ADEQUACY OF THE DISPERSION MODEL AND REGULATORY MODELING CHAIN FOR THE INTENDED REGULATORY APPLICATIONS

The key objective for the recommended enhanced model evaluation process for project-level air quality analyses is to validate (or determine the adequacy of) both the dispersion model and the regulatory modeling chain for their intended regulatory purposes. For dispersion models, the validation is against tracer data from a moving release platform. For the regulatory modeling chain, in order to show compliance with statistical accuracy in the applicable regulatory tests for transportation (NAAQS and Build/N-Build) for the specified transportation project, the validation is against near-road ambient air quality data for pollutants (namely, $PM_{2.5}$, PM_{10}, and CO at present) for which project-level air quality analyses are required for the full range of typical transportation applications (facility type, configuration, operating condition, setting etc.)

Another objective is to compare the results of the dispersion model and the regulatory modeling chain to draw reasonable inferences about sources of error and uncertainty. For example, if the uncertainty in the results for the regulatory modeling chain is significantly greater than that observed for the dispersion model, then the conclusion may be drawn that the additional uncertainty stems from parts of the regulatory modeling chain other than the dispersion model, i.e., from the estimates for traffic, emissions and/or background concentrations. Further, if there is reasonable confidence (e.g., due to site-specific data collection) in the traffic data and background concentrations, then it may be concluded that the added uncertainty is primarily from the emission model. Recommendations from the model evaluation process may then conclude that improvements to the emission model are needed for the project level analysis. To the extent cost feasible, field data collection and analysis should be designed to allow identifying the step(s) in the modeling chain in which any error is originating, e.g., using high quality site-specific data for traffic, meteorology, background concentrations and land use.

The validations or determinations of adequacy may also be conditional, i.e., adequate but with limitations to specific applications based on performance and/or, perhaps more commonly, the absence of validation of other applications against representative field data. For example, the dispersion model and associated regulatory modeling chain may be found to be adequate (able to show compliance with the regulatory tests with statistical confidence) for a specific facility type and configuration such as a high-volume freeway without noise walls, if that is the only application for which the dispersion model and regulatory modeling chain have been validated against representative field (co-located tracer and near-road monitoring) data. The model evaluation process in these cases may conclude that the models may still be applied in other regulatory applications with the recognition (or warning) that they have not been validated for those applications as yet or, in some cases, may reasonably be extended to other applications given their similarity. For example, a validation for $PM_{2.5}$ for a specific application (e.g., facility type and configuration) may be extended to CO in the absence of pollutant removal or formation processes (deposition or chemical transformation, respectively) that could significantly affecting near-road concentrations.

Limitations can be specified using any number of reasonable measures, including but not limited to:

- Facility type, e.g., limitations to: highways where model validation has been completed, with warnings that the model has not been validated against field data for other facility types.
- Regulatory test type, e.g., for the NAAQS test but not the Build/No-Build test, if the latter cannot be met with statistical confidence.
- Receptor exclusion zones, e.g., near noise walls, if the model has not been validated for receptors close to the wall such as where a bike path may be located.

If the results of the model evaluation process shows that the regulatory model chain cannot be applied to meet its intended regulatory purpose for PM2.5, i.e., cannot be used to show compliance with statistical confidence for the NAAQS test and/or Build/No-Build tests, it may recommend that qualitative analyses be undertaken consistent with the conformity rule at 40 CFR 93.123(b)(2).[1]

Adequacy determinations are needed for both the dispersion models and the full regulatory modeling chain.

- The determination for the dispersion model alone is whether it meets specified technical criteria, e.g., accuracy, parsimony etc.
- The determination for the regulatory modeling chain in contrast is for the applicable regulatory tests, i.e., the NAAQS and Build/No-Build tests, it may recommend that qualitative analyses be undertaken consistent with the conformity tests. It must be done against near-road air quality monitoring data, as well as on-site measurements of traffic, meteorological, and background concentration.

Documentation for the model evaluations for each model should list the applications for which it has been validated against representative field data, with any limitations as appropriate.

[1] https://www.govinfo.gov/content/pkg/CFR-2014-title40-vol20/xml/CFR-2014-title40-vol20-sec93-123.xml

2.2 LIFE-CYCLE APPROACH

Model evaluations should be conducted on a life-cycle basis, i.e., whenever it may reasonably be considered to be warranted throughout the time that the model is required for use in regulatory applications. Assessments are generally warranted whenever the model is fundamentally updated and new capabilities are added. The overall process should result in continuous improvements over time in the models and associated guidance.

Once a model has completed an assessment, subsequent assessments may focus on changes to the model and/or other triggers. Triggers for initiating an assessment generally include but are not limited to:

- Periodic updates to the preferred model, for the improvements associated with that update as appropriate. For example, adding a significant new capability such as modeling the effects of sound walls would be cause for validating that new capability against appropriate field data.
- Testing of potential alternative models. For example, a new air quality dispersion model (e.g., a Gaussian puff model) is developed for transportation projects.
- Compelling evidence, e.g., as presented in multiple peer-reviewed paper, that existing model(s) get the "wrong" answer in specific regulatory applications, e.g., the model generates concentration estimates that are substantially greater or less (by a factor of 2 or more) than observed concentrations for a particular facility type.

  Another example would be if multiple peer-reviewed paper presented evidence that, while the model may work well for specific source types, it may not work well with one or more different source types at the same time, e.g., mixed source types such as roads with an intermodal facility in the same project area.

  Another example might be the identification of unique characteristics of some kinds of transportation facilities that would be better served by some other form of the model or a different model (e.g., computational fluid dynamic model).
- Development of a new NAAQS whose concentration levels are expected to vary with microscale components (short-time and distance intervals – e.g., 5-minute standard), if there is reasonable expectation that the change(s) could affect the determination of model adequacy in regulatory applications.

2.3 COMPREHENSIVE COVERAGE AND PRIORITIZATION OF REGULATORY APPLICATIONS FOR TRANSPORTATION

For purposes of establishing priorities and budgeting for future model evaluations, a prioritized list of regulatory applications for transportation that necessitates a model evaluation is needed. These evaluations can then be performed on a priority basis for each one in order as resources permit. Regulatory applications that cannot be assessed within current resource constraints can be identified and prioritized for future assessments as resources become available. Note multiple data sets are needed to validate the models against one set of data and then test the model performance against another set. The identification and prioritization of regulatory applications for model evaluations should involve stakeholder consultation and may be done as a separate exercise or study apart from any specific model evaluation.

Appropriate criteria for deciding which facility types and settings would warrant a separate assessment are needed. In general, assessments are needed for regulatory applications (both screening and refined) for which the model is required for use by regulation for transportation projects. Consideration should be given for separate model evaluations for which near-road ambient concentrations, and specifically design values, may reasonably be expected to differ to the extent that they may affect the determination with statistical confidence of compliance with the applicable regulatory tests.

Potential applications include typical transportation facility types, configurations, operating conditions, settings (urban/rural), pollutants etc. While the list of potential applications may be lengthy it may be reduced in the prioritization process by combining related applications where significant differences in model concentrations and design values are not expected or by conducting field work concurrently wherever feasible. For example, separate field studies and model evaluations for a high-volume freeway and other high-volume arterial highways may be determined to not be warranted or of low priority given the expected similarities in the facility types. Another example is to combining field work for different facility types that warrant separate model evaluations, e.g., a freeway segment with an adjacent congested intersection with high diesel truck and/or bus traffic and/or with an adjacent (diesel) truck and/or bus terminal.

Potential regulatory applications to be considered for separate model evaluations include but are not limited to the following:

- Regulatory Tests – NAAQS, Build/No Build (see description below) and/or state environmental requirements. Obtaining field data for Build/No Build tests may be particularly challenging; determinations of model adequacy therefore may rely upon measures of uncertainty for the NAAQS test compared to estimates of uncertainty in estimates for Build/No Build traffic and emissions.
- Screening versus refined modeling.
- Pollutants: CO, $PM_{2.5}$, PM_{10}, other microscale pollutants that may be deemed necessary for an air quality analysis of a transportation project.
- Project or facility type, e.g.:
    - High volume highway segments (where the added turbulence from traffic reduces peak concentrations),
    - Interchanges (freeway-to-freeway, and arterial interchanges), with and without adjacent congested signalized intersections with high diesel truck and/or bus traffic,
    - Signalized intersections with high diesel truck and/or bus volumes, and
    - Off-network sites including:
        - Truck and bus terminals operating diesel fueled vehicles,
        - Park & ride lots, and

        - Other off-network sites, including consideration of adjacent industrial sources (industrial complex, railyards with diesel locomotives) within x meters of the roadway.

- Typical range of configurations
    - With and without noise walls. A range of typical noise wall heights and distances from the roadway edge should be tested. One option for field testing may be to set up temporary barriers for this purpose.
    - Urban street canyon
    - Depressed and elevated sections
    - Skewed or unskewed intersections and interchanges,
    - Roadways with significant changes in road grades, to test both the emission and dispersion models. This is important since MOVES generates anomalous emission factors at high road grades, the effect of which varies by pollutant, fuel type, and facility type, as reported to the MOVES Review Workgroup[2] of the EPA Mobile Source Technical Review Subcommittee (MSTRS) in 2019. Onboard portable emission monitoring systems may be useful for these tests. One option may be to test roads with truck climbing lanes, with and without noise walls
- Operating conditions (e.g., high or low volume, high or low diesel truck and bus volumes, level of service, etc.) …"
- Setting – urban or rural

2.4 FIELD DATA (CO-LOCATED TRACER AND NEAR-ROAD MONITORING DATA) COLLECTION FOR PRIORITIZED REGULATORY APPLICATIONS

For each regulatory application (facility type etc.) that has been prioritized for model validation, field data (co-located tracer from a moving vehicle platform and near-road ambient air quality monitoring data is needed. In addition, site-specific traffic volume, fleet age distribution, fleet mix, meteorology, and background concentration data are needed as well as land use for the area surrounding the monitoring site. If these data are not otherwise available, field data collection will be needed for this information. Given the complexity of the field work, a detailed plan is needed to collect and analyze, QA and QC the collected tracer and near-road ambient air monitoring data. The plan should be comprehensive and cover related elements including traffic data collection, on-siter meteorology, and the determination of representative background concentrations (preferably on-site) in addition to the tracer and near-road ambient air quality monitoring data.

[2] See: https://www.epa.gov/sites/default/files/2019-12/documents/03-moves-project-level-analyses-2019-10-09.pdf

2.5 USE OF MULTIPLE TECHNICAL CRITERIA

Consistent with the recommendations of the 2007 NRC study, multiple criteria should be specified for regulatory model evaluation, which differ in priority for refined and screening models. Recommended criteria for project-level air quality analyses for transportation include:

- **Accuracy**, for which there are typically several statistical measures. For regulatory modeling, the focus is typically on accuracy for maximum concentrations that are used to determine design values for comparison with the applicable NAAQS. Here the statistical measure as the robust highest concentration is widely used in air quality modeling to assess the model's performance for a set of highest concentrations. To assess the accuracy of the model for the entire near-road concentration field statistical measures are used to assess the models performance statistical tests such as the fraction and absolute bias are routinely used.

  One area of concern is whether the modeling chain is sufficiently accurate to be used in Build/No Build tests if the uncertainty in the model estimates for design values is comparable to or higher than the percent change in travel activity and emissions. In the absence of field data for the Build/No Build test, which can be difficult to obtain, it may still be possible to draw reasonable inferences on accuracy and whether it is sufficient to demonstrate model adequacy for Build/No Build applications. One approach may be to compare the uncertainty in the field data results for design values to typical values for the percent change that may be expected in travel activity for a build scenario. For example, if the field data indicate the uncertainty in modeled design values to be substantial, e.g., a factor of two, and the percent change in vehicle-miles-traveled for the build versus the no-build scenario is substantially lower than that, e.g., 20 percent, then it may reasonably be concluded that the modeling chain (and the dispersion model as applicable) is not adequate or not expected to be adequate for the Build/No Build test for that regulatory application.

- **Parsimony or Proportionality**. The goal for regulatory models is to not be more complex or resource-intensive to apply than needed for the specific regulatory need. This is important for all regulatory models and particularly so for screening models.
  - Aggregate measures for this include the time and cost for conducting typical analyses using the model following applicable EPA and other (e.g., state) guidance.
  - More specific or targeted measures may also be used, e.g., considering both the sensitivity of a model to specific inputs and time or cost estimates for generating those inputs. If the added time or cost is significant but the model is relatively less sensitive to that input, then appropriate measures to reduce the time and cost for analyses may be considered, e.g., using default representative values.

  In general, there may be tradeoffs between model parsimony and accuracy. This may require weighing model complexity (e.g., as reflected in the number and type of inputs) against how accurate and precise the model results must be to meet the intended regulatory requirements.[3]

[3] As air quality standards have lowered over the past decades air quality model's parsimony has increased to achieve the desired level of accuracy.

Note this criterion may not be needed for research applications of the model, where the focus tends to be much more on accuracy than efficiency.

- **Ease-of-Use**. This applies for both screening models and refined models for regulatory application. Ease-of-use measures may include whether a graphical user interface is available, a subjective measure of its quality, the provision of default model inputs consistent applicable federal guidance, etc. This criterion is typically of higher weight for screening applications than refined.
- **Quality Assurance and Control (QA/QC)**: This includes the degree to which the model helps users in specifying and checking input data. It may also include the degree to which the model generates standard output tables and charts that can be directly used in environmental clearance documentation without post-processing or other user modification, i.e., minimizing the potential for user error in output processing and documentation.

2.6 COMPARATIVE ANALYSES

Model evaluations should include rigorous comparative analyses with other models (regulatory and non-regulatory). The general objectives are to identify the best performing models for specific applications as well as the features or elements (algorithms, parameterizations) of the science in the comparison models that contribute to that better performance and so may be worth considering for incorporation (modified as needed) into the regulatory model in a future update or as a beta option in the regulatory model for further testing and evaluation. The comparative analysis should therefore document any difference(s) in performance between the models and provide expert insight and analysis to explain why the observed difference(s) occur.

Comparative analyses also provide a useful means of assessing and validating potential alternative models, which may have better performance on one or more criteria than the preferred model. This in turn may be helpful in informing future updates to both the preferred and alternative models.

Multiple stakeholders may be involved in comparative analyses as a means to apply the best resources for this task. A workshop may be needed in these cases to generate consensus findings. Issues for which consensus findings are not obtained may be noted with recommendations for future research to resolve any questions.

2.7 STAKEHOLDER CONSULTATION AND PEER REVIEW

The model evaluation process should be open and transparent and involve a robust and ongoing peer review and stakeholder involvement element. It should generally follow the cooperative, continuous, and comprehensive approach that has long been used by transportation agencies in other modeling work.[4] For model evaluations specifically:

- Key stakeholders include state DOTs, US DOT, and other transportation organizations that typically sponsor transportation projects. Other potential stakeholders include transportation-related organizations such as AASHTO as well as technical experts in

[4] See: https://www.federalregister.gov/documents/2016/05/27/2016-11964/statewide-and-nonmetropolitan-transportation-planning-metropolitan-transportation-planning

related subject matter from academia, the consulting community, and research institutions. A pooled fund approach may be one way to not only generate funds for model validation efforts including field data collection but also include the engagement of state DOTs in the process.

- All stakeholders should be provided complete, timely and ongoing access to all information, without delay or redaction. For this purpose, access is provided (e.g., via a web site, email etc.) to all stakeholders on an ongoing basis to all aspects of the assessment, including but not limited to the planning, specification of methodologies including both field testing and modeling, meeting materials, modeling code and executables, raw and post-processed field data, modeling results, utility files/spreadsheets used for pre- and post-processing, data and sources for all model inputs, etc.
- Stakeholder engagement and involvement should occur throughout the process. Additionally, specific opportunities for comment should be provided at the beginning of the model evaluation process, to get feedback and recommendations regarding existing or known issues or challenges with the model for specific transportation applications before the model assessment is initiated; at the end, for review and comment on the draft model assessment; and at key milestones as appropriate in the process. While this will slow the process it is key that stakeholders be engaged in the process.

All comments from stakeholders must be provide as written comments with all stakeholders receiving other comments so that differences can be resolved by all parties. Stakeholder and peer review comments on desired model features or capabilities must be proactively sought in addition to comments on model performance or other issues, to help inform and prioritize future updates to the models and, as appropriate, guidance.

2.8 SUMMARY

Given the volume of complex and highly detailed information considered in the development of a model evaluation, an executive summary that concisely captures its key elements and presents recommendations for future priorities and funding is needed for decision-makers and stakeholders. As the information presented in the summary may be derived from multiple sources, and the summary effectively represents the thoughts and opinions of the project manager (or chair of the steering committee if one is established) on all matters relevant to the model evaluation, it may be more accurately referred to as the Project Manager Discussion and Analysis (PMDA). It serves to make transparent to all stakeholders the key issues and how and why specific conclusions and recommendations were made related to determinations of model adequacy, model improvements, future research, future model evaluation needs, and other topics of interest. It will be the focus of decision-makers in their determination of which recommendations to accept and fund for future model improvements.

The PMDA provides a review of the most relevant factors from the model evaluation including the results of the assessment against technical criteria (accuracy, parsimony etc.), model comparative analysis, comments from state DOT, FHWA/FTA and other stakeholders and peer reviewers (including on the draft PMDA as part of the draft model evaluation), insights from a focused literature review, a prioritized sets of issues to be addressed in the future (including deficiencies in the model and associated guidance) along with recommendations for improvements including future research needs, and how each factor was weighed in the determination of the final conclusions and recommendations. PM, CO, and any other pollutants subject to regulatory applications may be considered separately or together as appropriate.

Dissenting opinions from the peer review and stakeholder involvement process may be included to provide any needed context to decision-makers and other stakeholders and peer reviewers. Interested readers may be referred to the working papers or task reports developed in the course of the model evaluation (e.g., field study, modeling, and the comparative analyses) for more detail, each of which may have detailed executive summaries of a more technical nature.

The PMDA may be several pages in length, and vary in format, length, and content from one assessment to the next depending on the specific issues encountered in the assessment. It focuses on the most critical issues of each assessment – particularly those affecting decisions on next steps - rather than simply summarizing all its various aspects.

3. Model Evaluation Plan

For each model validation, a model evaluation plan should be developed implementing the general principles. The plan should document the proposed methodology for stakeholder review and may be updated as needed during the course of the assessment. An outline of potential key elements of a model evaluation plan is presented below.

The plan may be accomplished using consultants and/or staff with relevant expertise as appropriate to each task. Deliverables for each task may be in the form of memoranda, working papers or task reports that can be combined (modified as needed) into one model evaluation report at the completion of the model evaluation process. Note the ordering of the sections in the final report will generally differ from those in the plan as outlined below, which are presented in the expected order of preparation and not in the typical order for the final report as presented in Attachment A.

3.1 SELECTION OF MODELS FOR EVALUATION

Specify the:

- Dispersion models (including version numbers) to be evaluated which typically include the preferred model as well as one or more other models that may be proposed as alternative models.
- Emission model (i.e., MOVES or EMFAC) to be applied for the test of the regulatory modeling chain.
- Guidance documents that will be applied for testing the regulatory modeling chain.

3.2 CONSULTATION

A consultation plan with a schedule should be developed for the model evaluation. The plan should include:

- Identification of key stakeholders to be consulted on the model evaluation, including state DOTs, the US DOT, other transportation project sponsors, and others as appropriate including technical experts from other agencies, the consulting community and research institutions.
- A list of major consultation activities and provide milestone date(s) for delivery to them of key interim products of the model evaluation, e.g., draft task reports, for review and comment.

  At the beginning of the process, a request should specifically be made for stakeholder feedback on the proposed scope of the model evaluation and the draft model evaluation plan as well as any other topic of interest to the stakeholder that they would like to see addressed in the study, e.g., any specific issues that they have experienced with the preferred model, or desired features for consideration in future updates.

- Specification of a website where all information pertaining to the model evaluation will be posted and may be obtained by the stakeholders.

- A request for feedback on the consultation plan and on any specific topics of interest for each stakeholder.
- A schedule for the review of the draft model evaluation plan.

The model evaluation report should include documentation of all consultation, including a summary of stakeholder input. This includes written comments on the draft deliverables, and specific issues that they may have experienced with the model, or desired features for consideration in future updates. A summary of key issues (if any) identified in the review for the draft model evaluation should also be provided.

3.3 LITERATURE REVIEW

A targeted literature review may be needed to help identify potential issues that may be of interest to the model evaluation and in particular to identify:

- Documented model deficiencies for regulatory applications for pollutants subject to the NAAQS and Build/No-Build (B/NB) tests.
- Documented recommendations for improvements for the preferred model and/or potential alternative models that may warrant testing as part of the model assessment.

The literature review may also help prioritize and support the determination of the specific regulatory application (e.g., facility type) to be selected for the model evaluation. Previous model evaluations may be included as part of the literature review, particularly if they identified unresolved deficiencies or issues with one or more models that could be addressed in the present model evaluation process.

3.4 SELECTION OF THE REGULATORY APPLICATION(S) FOR MODEL EVALUATION

Specify the regulatory application to be subjected to model evaluation, i.e., the specific facility type, configuration, operating condition and setting. The selection may be made considering established priorities for regulatory applications (if available), the results of the literature review, stakeholder feedback, and any other relevant information as available.

Note, multiple field data sets may be needed to ensure that the results are reproducible and to check model validation and avoid bias, i.e., by validating the model against one data set and checking model performance against another data set.

3.5 FIELD DATA COLLECTION AND ANALYSIS

The model evaluation plan should include a detailed field data collection and analysis section that addresses collection and analysis of site-specific tracer and near-road monitoring data, traffic, meteorological, background concentration, and land use.

If field data from prior studies are planned for use in the model evaluation and additional data collection is therefore not needed, then this should be documented in the plan and provide a brief overview of the methodology and results for the field data collection from that prior study. Any planned data analysis as needed to use the available field data should be specified, e.g., if the plan is to use the raw data from the prior study and conduct the needed data processing as part of the new model evaluation.

In either case, co-located tracer and near-road ambient air quality monitoring data collection is needed for model evaluations to allow for evaluation of the dispersion model component against the tracer data and the regulatory modeling chain of traffic, emissions, and dispersion against the near-road air quality monitoring data to assess its ability for compliance with the NAAQS and build/no-build tests. Comparison of the results for the tracer and regulatory chain also allows assessments of whether any identified issues with accuracy for the regulatory modeling chain are attributable to the dispersion model or to other parts of the modeling chain, i.e., traffic, emission modeling and/or background concentrations.

Other specific needs include:

- A facility layout of the transportation setting for a broad audience including decision-makers, stakeholders, public, and other non-technical parties.
- Site selection for the field studies that allows background concentrations for each pollutant to be well characterized.
- Detailed traffic data collection is strongly recommended to support the regulatory modeling chain test, to minimize or eliminate the potential error in estimating traffic volumes, speeds, and fleet mix and age distribution.
- Meteorological data collection (wind speed, direction, sensor threshold for minimum wind speed, standard deviation in wind direction, temperature, cloud cover, solar radiation, rainfall, source of meteorological data; on-site, modeled derived, etc.)
- Tracer Details
    - Release rate, location, duration of release, use of multiple tracers etc.
    - Tracer monitors (location and duration).
- Near-Road Monitoring Details
    - Pollutants, location, monitor instrumentation type and accuracy etc.
- Other monitoring (background) location and analysis details.

Overall, the plan for field data collection must provide sufficient detail for reviewers to conclude that the proposed approach is reasonable and should result in data of suitable quality for use in the model evaluation.

3.6 MODELING

If travel demand modeling and/or simulation is planned, e.g., in addition to traffic data collection planned as part of the field study, the planned approach should also be summarized.

3.7 ANALYSIS OF RESULTS

The plan should detail how the modeling results will be analyzed and assessed against the field data for the dispersion models against tracer data and the regulatory modeling chain against near-road ambient air quality monitoring data.

3.7.1 Technical Criteria for Model Evaluation

Specify the proposed criteria and their relative weights for the model evaluations against field data, which typically include accuracy, parsimony, ease-of-use, and uncertainty features as well as additional criteria as deemed appropriate. Also specify the statistical tests to be applied in the model evaluation, which typically include robust highest concentration (RHC), fractional bias, Q-Q and scatter plots, etc. Discussion on uncertainty should also be addressed and may be determined separately for receptors nearest the road (e.g., less than 50 meters away) and those at greater distance.

Given the need to test the regulatory modeling chain for the intended regulatory purpose of design values for comparisons with the applicable NAAQS, the criteria should include accuracy for design values in addition to the overall near-road concentration profile.

Indicate if the model evaluation is for refined or screening applications of the models. This may affect the relative weight for each criterion, i.e., with more weight on accuracy for refined applications, and more on parsimony and ease-of-use for screening applications.

3.7.2 Comparative Analyses

Specify the planned approach for the comparative analysis for the dispersion models tested. Indicate who will be performing the analyses and if multiple groups a workshop or meeting may be needed to generate consensus findings. Issues for which consensus findings are not obtained should be documented with recommendations as appropriate for future research to resolve any questions.

3.7.3 Determinations of Model Adequacy

Based on the results of the various analyses, provide separate determinations of model adequacy for the dispersion models and the regulatory modeling chain, with any limitations to specific applications (facility type, configuration as etc.) appropriate.

- For the dispersion model, this means primarily whether its meets prescribed standards for accuracy.
- For the regulatory modeling chain, this means, this means its verified capability to show compliance with statistical confidence in the applicable regulatory tests for transportation (NAAQS and build/no-build) for the specified transportation application. The validation is against near-road air quality monitoring data for pollutants (namely, $PM_{2.5}$, PM_{10}, and CO at present) for which project-level air quality analyses are required for the typical transportation applications (facility type, configuration, operating condition, setting etc.)

If the results of the model evaluation process shows that the regulatory model chain cannot be applied to meet its intended regulatory purpose for PM, i.e., cannot be used to show compliance with statistical confidence for the NAAQS test and/or build/no-build tests, it may recommend that qualitative analyses be undertaken consistent with the conformity rule at 40 CFR 93.123(b)(2).[5]

Examples of determinations of model adequacy are provided below in two main categories:

1) Adequate, with or without identified limitations,

2) Inadequate and should not be used in the specified regulatory applications.

   1. ADEQUATE, WITH OR WITHOUT LIMITATIONS: Cases in which model performance has been determined to be adequate, with any limitations as stated, based on an assessment against standard criteria in an approved model evaluation.

      Examples:

      - Mode: Refined for PM2.5
      - Regulatory Test: NAAQS only in the PM Hot Spot Guidance. Due to uncertainty in model estimates as well as background concentrations, use of this model for regulatory purposes is limited to cases with a minimum margin of "x" micrograms/m3 for PM2.5 between background concentrations and the annual NAAQS
      - Setting: Urban only (field data not available for rural settings)
      - Facility Types
        - Freeway-Freeway (Free-Flow) Interchange
          - Configuration: "T"
          - Operating Condition: Free-flow and congested, low truck%
          - Near-Road Barrier: Absent (not within x meters from roadway)
          - Nearby Sources (rail, stacks, etc.): Absent (not within x meters from roadway)
        - Freeway Segment
          - List of conditions as above
        - Arterial Street Intersection
          - List of conditions as above
        - Other Facility Types

[5] https://www.govinfo.gov/content/pkg/CFR-2014-title40-vol20/xml/CFR-2014-title40-vol20-sec93-123.xml

2. <u>NOT SHOWN TO BE ADEQUATE</u>: This includes both cases in which:

- Model performance has been determined to NOT be adequate based on an assessment against standard criteria in an approved model evaluation, and
- The model has not (yet) been subjected to a model evaluation

Note: Qualitative analyses are not precluded for these regulatory applications and may be appropriate in the absence of a validated model.

Examples:

- Regulatory Test:
  - B/NB from the PM Hot Spot Guidance for all facility types. Model uncertainty, which is high, effectively sets a limit below which B/NB comparisons cannot be made with statistical confidence
  - NAAQS Tests from the PM Hot Spot Guidance that do not meet the margin limits or other limits specified above, i.e., a minimum margin of "x" micrograms/m^3 for PM2.5 between background concentrations and the annual NAAQS.
- Setting: Rural, due to lack of field data and/or poor model performance in the model evaluation.
- Near-Road Barriers: Limited to facilities without near-road barriers
- Facility Types
  - Truck Terminals: Based on poor performance in tests against field data, as documented in the XXX model assessment
  - Bus terminals: No model assessment yet, as no field data/tracer study
  - Highly skewed intersections (> x degrees): No model assessment yet.

3.8 CURRENT LIMITATIONS WITH DISPERSION MODEL APPLICATIONS

Deficiencies in the dispersion models, modeling chain and associated guidance may be identified in the literature review, stakeholder consultation, evaluation against technical criteria, comparative analyses, and previous model evaluations. Deficiencies identified early in the process (e.g., in the literature review or stakeholder consultation) may be addressed in the model evaluation process (e.g., in the field study to collect data) that can be used to address or further define the issue. Identified deficiencies should be tabulated with initial responses as appropriate (e.g., the deficiency may be addressed in future updates to the model data needed to resolve the question were collected as part of the field study and the issue may be resolved in the next major model update) or further research will be required to resolve.

An example is provided below with an additional example in Attachment B. The list in an actual model assessment may be considerably longer or shorter depending on the field data collected.

Deficiency: The modeling chain uncertainty is too high to apply the NAAQS assessment with statistical confidence following the PM Hot Spot Guidance.

Recommendation: Research is needed to improve emission estimate for all components for project level assessments which should primarily focus on improved estimates of PM emission rates for brake wear, tire wear, exhaust, and re-entrained road dust. In the interim, the regulatory requirements for these model applications should be limited until better estimates of emissions are determined for project level analysis and the model evaluation determines the model is adequate for regulatory application. Qualitative analyses may still be conducted for these cases, as provided in the conformity rule for PM analyses at 40 CFR 93.123(b)(2) "Where quantitative analysis methods are not available, the demonstration required by § 93.116 for projects described in paragraph (b)(1) of this section must be based on a qualitative consideration of local factors."

3.9 CONCLUSIONS AND RECOMMENDATIONS

Present a comprehensive set of conclusions and recommendations. Summarize the work done including specifically the determinations of model adequacy for the intended regulatory applications, with limitations as appropriate, and the basis for those determinations.

Recommendations for future model evaluations, field studies, research and any other work that would require new funding or prioritization by decision-makers for future efforts should be discussed, including preliminary budget estimates to support funding requests. This includes recommendations for:

- Model evaluations for specific transportation applications (facility type and configuration, etc.)
- Field studies or other research needed to support future model evaluations or improvements
- Model improvements (near, intermediate, and long-term) by pollutant as appropriate. This includes recommendations to address any model deficiencies identified in the model evaluation, with a priority for the preferred model.
- Future research

3.10 SUMMARY

The project manager should provide a comprehensive summary (discussion and analysis) of the model evaluation process and results, written for an audience of stakeholders and decision-makers that need to understand the key issues and how they affect the conclusions and recommendations. Dissenting opinions, e.g., from one or more stakeholders, may also be noted and addressed. Details should refer readers to the conclusions and recommendations section and/or to the underlying working papers or task reports developed in the course of the model evaluation.

The summary focuses on the determinations of model adequacy (including the results of the model evaluation against performance criteria and the conclusions and insights from the comparative analyses), identification of deficiencies and recommendations for model improvements, recommendations for future research.

4. Selection of the Regulatory Application(s) for Model Evaluation

In support of the recommended enhanced model evaluation process, transportation applications for which model evaluations and field data collection are needed must be identified and prioritized, given the associated high labor and equipment cost needed to collect all of the necessary data needed for the model evaluation. The model evaluation can then be performed on a priority basis for each application in priority order as resources permit. The identification and prioritization of regulatory applications for model evaluations may be done as a separate exercise (e.g., a DOT pooled-fund study) apart from any specific model evaluation, and then referenced as appropriate in subsequent model evaluations.

Appropriate criteria for deciding which transportation applications (facility types etc.) would warrant a separate assessment are needed. In general, assessments are needed for all regulatory applications (both screening and refined) for which the model is required for use by regulation for transportation improvement projects. This includes near-road environment for which the design values maybe expected to differ to the extent that they may affect the determination with of compliance with the applicable regulatory tests (NAAQS and B/NB) or otherwise present a unique challenge to modeling.

For reference, Table 1 below identifies examples of potential transportation applications. Only two applications to date listed in the table involved the collection of the co-located tracer and near-road monitoring data as recommended here. The list may be reduced if it determined, via stakeholder consultation, that some applications are reasonably similar, i.e., if one of those applications is validated, validation of the others may reasonably be expected due to their similarity.

Table 1. Potential Regulatory Applications for Transportation (Not Prioritized)

Facilities (Highway, Transit and Intermodal)	Field Data (not tracer)	Tracer Data	Model Evaluation Completed	Models Assessed
Freeway segments, Congested and Uncongested Operations, without Near-Road Barrier	NCHRP 25-55 (2022): CO & PM2.5, and traffic data	NCHRP 25-55 (2022)	NCHRP 25-55 (2022)	AERMOD, AERMOD-RLINE, ADMS-ROADS, screening models CAL3QHC (CO -only), AERSCREEN
Freeway segments, Congested and Uncongested Operations, with one and two near-road barriers	UCR-Caltrans (2023) CO, CO2	UCR-Caltrans (202?) SF6	In progress	AERMOD, AERMOD-RLINE
Freeway segments, Congested and Uncongested Operations, with Near-Road Barrier, High Truck%, High Road grades Etc.	n/a	n/a	No	n/a
Interchanges: Freeway-Freeway, Unskewed, Congested and Uncongested Operations	n/a	n/a	No	n/a

Facilities (Highway, Transit and Intermodal)	Field Data (not tracer)	Tracer Data	Model Evaluation Completed	Models Assessed
Interchanges: Freeway-Arterial	n/a	n/a	No	
Signalized Intersections: Congested and Uncongested Operations, Unskewed	n/a	n/a	No	n/a
Signalized Intersections: Congested and Uncongested Operations, Skewed	n/a	n/a	No	n/a
Signalized Intersections:	n/a	n/a	No	n/a
Truck terminals	n/a	n/a	No	n/a
Bus terminals	n/a	n/a	No	n/a
Arterial streets	n/a	n/a	No	n/a
Park and ride lots	n/a	n/a	No	n/a
Street canyons	n/a	n/a	No	n/a
Port (drayage) operations	n/a	n/a	No	n/a

This list of transportation applications is partial and may be expanded to include for example a range of settings (e.g., urban, rural, urban canyon), operating conditions (congested, uncongested, high diesel truck percent etc.), and configurations. Note: this list is not intended to be a complete list of potential project types and configurations.

Prioritized studies may include:

- High-volume congested intersections with high percentages of diesel truck and/or bus traffic
- High-volume freeway-arterial street interchanges high percentages of diesel truck and/or bus traffic
- Truck and/or bus terminals with high volumes of diesel vehicles
- Freeway segments with higher road grades
- Arterial segments at-grade and with higher road grades
- High volume freeways at-grade and with higher road grades, without and with noise walls covering the full range of noise wall heights and distances.

Field studies that provide an opportunity to collect data for the build/no-build test (in addition to the NAAQS) test are needed. Note, multiple field data sets for prioritized regulatory application are needed to ensure that the results are reproducible and to check model validation and avoid bias.

5. Conclusions and Recommendations

The following conclusions and recommendations are based on the work conducted in NCHRP 25-55, including the field studies, modeling, and analyses.

5.1 CONCLUSIONS

- An enhanced model evaluation process is needed for the transportation sector, as refined modeling requirements were introduced without extensive validation of the dispersion models and regulatory modeling chain against tracer and near-road ambient air quality monitoring data respectively for the full range of typical transportation applications.
- Model validation or evaluation for the regulatory modeling chain is critical for determining its adequacy for the intended regulatory purpose of showing compliance with the applicable regulatory tests.
- Model evaluation for the regulatory modeling chain cannot be done using just tracer studies but requires co-located field studies to collect near-road air quality monitoring data.
- The use of co-located tracer and near-road ambient air monitoring sites as done in 25-55 allows comparisons on the modeling results for both the dispersion model and the regulatory modeling chain. If the dispersion model performs well, and traffic data and background concentrations are well characterized (i.e., are site-specific), then the focus for model improvements to improve accuracy and reduce uncertainty is on the emission modeling step in the regulatory modeling chain.
- Based on lessons learned in carrying out the field experiment it is essential the program be carefully planned and designed and that the a pilot study is carried out prior to the full program to assure that equipment operates as intended, tracer gas release amounts are sufficient, identify if other sources of tracer gas may be present that will interfere with the experiment, and assure collection equipment is free from cross-contamination.
- Identification and prioritization of the transportation applications (facilities, configurations, operating conditions, and setting) is needed. For cost-effectiveness and efficiency, consideration should be given to how to combine field studies for priority applications (e.g., combing field studies for a high-volume freeway arterial street interchange with adjacent congested intersections with diesel truck traffic instead of conducting two separate studies).
- Extensive involvement of transportation stakeholders is essential to have a successful outcome. This includes the participation of USDOT, state DOTs and other transportation project sponsors, EPA, and other technical experts from research organizations, academia, and the consulting community.
- Dispersion Models
    - Of the four models (AERMOD-AREAPOLY, AERMOD-VOLUME, AERMOD-RLINE, ADMS-ROADS) examined for the tracer study, AERMOD with source configuration of AREAPOLY showed better performance when compared to the

other three models/source evaluated. AERMOD-AREAPOLY had a lower absolute average residual, fractional bias, and importantly the best pairing of the RHC observed to modeled.

  - All four models showed a small tendency for underpredictions, as demonstrated by positive fractional bias values.
  - At the location where the highest concentrations occur (1-m from the roadway) models the AERMOD-AREAPOLY has the closest mean and maximum concentrations. The suspected cause for this better performance is that given the 10 lanes of freeway and close proximity of the receptors to the freeway that the emissions are best characterized as an area source rather than a line or set of volume sources

- Regulatory Modeling Chain
  - The regulatory evaluation based on the design value for all four model/source configuration showed that AERMOD-AREAPOLY were closest to the observed values.
  - Determination of model adequacy for the intended regulatory purpose:
    - All four models/configurations showed design values that were close to observed values with a high fraction (more than 80 percent) of modeled design values driven by background concentrations, so only 20 percent of modeled design values were due to traffic-related emission. Thus, background determination is an important determinant in the model evaluation process.
- Comparison of Tracer and Near Road Modeling
  - The results show that the dispersion models all performed better against tracer data than for the full regulatory modeling chain compared to near-road monitoring data for PM2.5 and CO. For example, the r-squared for the tracer experiments were all greater than or equal to 0.70, while for the PM2.5 and CO the highest r-squared was 0.39.
    - The traffic data were obtained primarily via site-specific monitoring and so is considered well-characterized and therefore not likely to be a significant source of error in the regulatory modeling chain for this study.
    - Backgrounds concentrations were determined from nearby monitors and are not likely to be a significant source of error in the regulatory modeling chain.
    - The emission modeling step (EMFAC, AP-42 for road dust) in the regulatory modeling chain therefore seems the likely source of uncertainty observed for the regulatory modeling chain.

- Comparative Analyses - No formal comparative analysis was conducted under the NCHRP 25-55 study. However the following findings and insights were made based on the results of the technical analyses.
    - Model comparisons showed better performance for AERMOD when applied using AREAPOLY source type than RLINE. This is likely a result of the current EPA guidance on estimating release height and initial vertical dimension being developed prior to the incorporation of RLINE into the AERMOD model.
        - Key papers used by EPA to recommend plume height and initial vertical dimension from vehicles most notably under stable conditions were reviewed. A key reference was the paper by Gilles et. al (2005) which reported that the initial vehicle height is scaled by a factor of 1.7 to determine release height. However, this was based on a small fleet of nine vehicles with no vehicle higher than 3.3 meters. We suspect that for taller vehicles this 1.7 factor may be an overestimate.
        - Additionally, as noted by Gilles that in addition to atmospheric stability, the shape of the vehicle, and the angle of the ambient wind with respect to the direction of vehicle travel play a role in determining this scaling height factor. Further study is needed to better estimate this scaling factor as currently estimated it likely overestimates the release height and initial vertical dispersion so that when this parameterization is coupled with the better source characterization line source algorithm in AERMOD-RLINE (numerically integrating point sources) the model underpredicts.

5.2 RECOMMENDATIONS

- It is recommended that enhanced model evaluation process for the transportation sector following the general principles and approach presented in this report be followed. This includes the use of co-located tracer from a mobile platform and near-road air quality monitoring data, with the key objective of determining the adequacy of the regulatory modeling chain for its intended regulatory purpose of showing compliance with applicable regulatory tests (NAAQS and Build/No Build).
    - Identify and prioritize the transportation applications (facilities, configurations, operating conditions, and setting) for future model evaluations, and field studies to collect co-located tracer and near-road monitoring data as needed for this purpose.
        - Potential priorities include:
            - High-volume congested intersections with high percentages of diesel truck and/or bus traffic
            - High-volume freeway-arterial street interchanges high percentages of diesel truck and/or bus traffic
            - Truck and/or bus terminals with high volumes of diesel vehicles
            - Freeway segments with higher road grades

        - Arterial segments at-grade and with higher road grades
        - High volume freeways at-grade and with higher road grades, without and with noise walls covering the full range of noise wall heights and distances.
    - Field studies that provide an opportunity to collect data for the Build/No Build test (in addition to the NAAQS) test are needed.
- Given the scope of the work including extensive interactive stakeholder discussion, a transportation pooled fund may serve as the most effective forum for financing and overseeing the field studies and model evaluation efforts.
- The present study was performed in California and employed California's emissions factor model EMFAC. Future studies should be conducted in other states using MOVES for pollutants-based and regulatory evaluations. (Tracer results are independent of the state in which they were conducted.)
- Prioritize model validation for PM (especially $PM_{2.5}$) over other pollutants given (1) typically high background concentrations relative to the applicable NAAQS, and (2) complicated source contribution (brake wear, tire wear, exhaust and road dust).
- Comparative analyses are recommended for inclusion in future model evaluations. They were only not done in this study given resource constraints.
- Future research:
    - Comparative analyses for the models and results from the NCHRP 25-55 study
    - Research is needed to determine the appropriate release height for both cars and trucks by measuring the vertical concentration profile under a variety of meteorological conditions. This should lead to an improved characterization of the release height and initial vertical dimension.
    - Research to determine the underlying basis for the uncertainty in the emission modeling step (especially for PM2.5) in the regulatory modeling chain and develop recommendations for improving the emission model.
    - Research to improve the determination of background concentrations for project level analyses.

6. References

Craig, K., Lynn M. Baringer, Shih-Ying Chang, Michael C. McCarthy, Song Bai, Annie F. Seagram, Vikram Rav, Karin Landsberg, Douglas S. Eisinger, (2020) "Modeled and measured near-road $PM_{2.5}$ concentrations: Indianapolis and Providence cases," Atmospheric Environment, Volume 240, 1 November 2020, 117775. https://doi.org/10.1016/j.atmosenv.2020.117775https://www.sciencedirect.com/science/article/abs/pii/S1352231020305070?via=ihub

Eisinger, D., Kenneth Craig, Karin Landsberg, Anondo Mukherjee, Jennifer DeWinter, Michael McCarthy, And Steven Brown, (2021) "Near-Road Air Quality Insights from a U.S. DOT Five-Year Transportation Pooled Fund Study", TR News March-April 2021. https://onlinepubs.trb.org/onlinepubs/trnews/trnews332.pdf

Gillies,J.A., V.Etyemezian, H.Kuhns, D.Nikolic, D.A. Gillette, Effect of vehicle characteristics on unpaved road dust emissions, (2005) Atmospheric Environment 39 (2005) 2341–2347. https://doi.org/10.1016/j.atmosenv.2004.05.064

NASA (2016) "Standard for Models and Simulations" ,NASA Technical Standard NASA-Std-7009a W/ Change 1: Administrative/ Editorial Changes 2016-12-07), https://standards.nasa.gov/standard/nasa/nasa-std-7009

National Research Council (NRC) 2007. *Models in Environmental Regulatory Decision Making* Committee on Models in the Regulatory Decision Process Board on Environmental Studies Toxicology Division on Earth and Life Studies, National Academics Press, June 2007. ISBN 978-309-38716-3. http://nap.edu/11972

USEPA (2018) "EPA Can Strengthen Its Process for Revising Air Quality Dispersion Models that Predict Impact of Pollutant Emissions", USEPA Inspector General, Report #18-P-0241, September 2018. Available: https://www.epa.gov/office-inspector-general/report-epa-can-strengthen-its-process-revising-air-quality-dispersion

Wayson, R. and Voigt, C. (2022) "AERMOD Source Types RLINE and RLINEXT Testing", Report for the FHWA Office of Natural Environment and FHWA Virginia Division Office, March 2022, pp.A-84 to A-88. The report is available on the AECOM.com website (link) and will also be posted on the page for MOVES-Matrix modeling on the Georgia Institute of Technology (https://www.gatech.edu/) website as part of the set of three reports that were developed for FHWA addressing the AERMOD RLINE model.

Attachment A: Outline for a Sample Model Evaluation

The final report for the model evaluation is comprised of the task reports or memorandums developed following the model evaluation plan, with some reordering as appropriate, e.g., the summary is presented first in the final report. The recommended order is presented below, with explanatory text added where needed.

- EXECUTIVE SUMMARY (Project Manager Discussion and Analysis)
- Models Evaluated
- Regulatory Application(s) for Model Evaluation
- Consultation
- Literature Review
- Field Data Collection and Analysis
    - Tracer Data
    - Near-Road Monitoring Data
    - Traffic and Emissions Data
    - Meteorological Data
    - Land use to support determinations of surface roughness
- Modeling
    - Dispersion Model Testing
    - Regulatory Modeling Chain Testing
- Analysis of Results
    - Dispersion Model Testing Against Tracer Data
    - Regulatory Modeling Chain Testing against Near-Road Monitoring Data
    - Comparison of Dispersion and Regulatory Modeling Chain Testing *(to support determinations of sources of error in the regulatory modeling chain)*
    - Comparative Analyses
    - Determinations of Model Adequacy
    - Identified Deficiencies in the Model(s) and Possible Fixes
- Conclusions

- Present a set of detailed conclusions as appropriate from the model evaluation. These should be written for a technical audience.
- The conclusions reached should be based on the model evaluation study and can extend across the modeling chain. Conclusion can be both definitive and speculative as long as they clearly communicated.

- Recommendations
  - Recommendations on model improvements needed (near, intermediate, and long-term) by pollutant:
    - The recommendations can be directed across the modeling chain but should tie back to findings from the model evaluation
  - Field studies or other research needed to support future model evaluations or improvements with preliminary budget estimates for reference in future funding requests or proposals, for example:
    - field studies directed at particular elements across the modeling chain (e.g., improved methodology for determining re-entrained road dust emission rates at varying traffic volumes)
    - field studies for particulate matter that can distinguish the source contribution from in field measurements in comparison with those from regulatory determined emission rates
    - field studies for specific transportation facility types that need to have model validations studies performed to support the needs of the transportation community

Attachment B: Example of a Partial Model Evaluation Based on the Berkeley Freeway Field Study Conducted Under NCHRP 25-55

CONTENTS

EXECUTIVE SUMMARY

Italicized material is guidance information on material that should be discussed. A model evaluation plan should be developed implementing the general principles for the assessment for stakeholder review. The key elements for each component of the model evaluation plan for 25-55 project are summarized in the executive summary.

The model evaluation plan identified the most relevant models for evaluation were the ADMS-Roads model and AERMOD dispersion models with three modeling options AERMOD-AREAPOLY, AERMOD-VOLUME and AERMOD-RLINE with AREAPOLY or VOLUME the current regulatory version. The evaluation focused first on the dispersion model using a tracer dataset collected as part of this study. The tracer study provides a well-known release rate so that the evaluation focuses on how well the dispersion model simulates the downwind concentration after its release to the atmosphere. The evaluation also included an assessment of how the modeling system performed when the entire modeling chain (emission factors, traffic volume and dispersion) is included in the assessment. The modeling chain requires inputs from an emission factor model. In this study we used California's EMFAC. Current regulatory guidance documents were followed to extent applicable. The regulatory evaluation used co-located measurements of air pollutants operated at a near-road air quality monitoring trailer operated by the local air district over a 1-year period. The regulatory evaluation of the model included hourly and daily measurements of CO and PM2.5 with 1 year of meteorological data for the same year measured at nearby meteorological station. Design values (DV) were determined consistent with the form of the NAAQS standards for 1-hour and 8-hour CO and 24-hour and annual PM2.5. Traffic data was based on local site-specific traffic data collection. A nearby air monitoring station served to provide background concentrations. The on-site traffic data exceeds that of typically available for regulatory assessments. This high-quality dataset was designed to mirror how a regulatory evaluation would be performed, but with higher quality information so that more robust model evaluation statistics are determined.

While not performed as part of the 25-55 project because of resource limitations key stakeholders should be consulted on the model evaluation plan, this would typically include state DOTs and US DOT and other technical experts from other agencies, academic research institutions and consultants.

A targeted literature review was needed in this study to help identify a potential issue found with the plume rise height and initial vertical dimension as currently recommended for use in modeling guidance. We identified the key reference papers on this issue and recommended that further study and evaluation is needed to better estimate these parameters as currently estimated as currently recommended likely overestimates the release height and initial vertical dispersion.

The field experiment was undertaken to collect air quality data along with meteorology and traffic information along a section of the Interstate 80 freeway running through Berkeley, CA. The site offered an excellent platform under which to evaluate the air quality models. A total of ten tracer experiments were conducted on different days over a variety of traffic and meteorological conditions. This high-volume freeway (280,400 AADT) with 4.8% truck fraction (13,460 AADT) segment runs parallel to Aquatic Park. In addition to the tracer data collection, co-located measurements of air pollutants as measured by the local air district (Bay Area Air Quality Monitoring District) near road monitor were also collected for this study. Hourly concentrations of CO and PM2.5 were paired with background concentrations as measured nearby middle scale air monitoring station.

Traffic data was collected to support the regulatory modeling analysis as the emission rates are based on vehicle traffic volume and speeds. Emission factors for CO and PM2.5 were based on California's emission factor model (EMFAC), with the exception of road dust based on EPA's AP-42 method. Traffic volume and speed information was readily available through the Caltrans Performance Measurement System (PeMS).

For the tracer modeling, we used the on-site wind measurements collected using a 2-D sonic anemometer. The tracer experiment dataset included hourly averaged data from May 3, 2021, June 6, 2021, and June 24, 2021, experiment days. Additional meteorological and surface characteristic parameters for surface air temperature, cloud cover, and upper-air temperature from the Oakland Airport Automated Surface Observing System with the Bowen ratio, albedo and surface roughness derived from land use and land cover information.

For the regulatory modeling we followed EPA and local BAAQMD guidance on how to prepare meteorological data for use in air dispersion models. Here we used the most representative one-year set of high-quality hourly meteorological data available which was from the 2019 Oakland Sewage Treatment Plant (STP) supplemented for missing wind measurements by the automated weather station operated at Pier 34 by the National Buoy Center. This dataset included an on-site turbulence parameter of standard deviation in wind direction that is frequently used in a regulatory application.

With EPA's release of AERMOD (version 19191) a new model feature, RLINE, became available to the air quality transportation specialist. RLINE is a line source algorithm that is based on the numerical integration of a series of point source emissions. This provides a more realistic approximation of a moving line source release. **This model evaluation assessment primary purpose was undertaken to evaluate AERMOD-RLINE in comparison to the area source algorithm (AREAPOLY) currently in use in modeling studies**. This evaluation used both the tracer dataset for evaluating the dispersion algorithm and in the regulatory application for the modeling chain evaluation. The key statistical tests for model acceptability are the robust highest concentration (RHC) and the fractional bias. These two statistics are the most succinct measures of model performance due to the importance of reproducing the upper end of the distribution of highest values.

This section summarizes the key findings and recommendations needed for decision-makers to potentially fund future model evaluation studies.

Key findings from the tracer evaluation study identified that AERMOD-AREAPOLY had the highest accuracy amongst the models evaluated which include AERMOD-RLINE, however the advantage appeared link to an overestimate in the current practice for estimating the release height particularly during stable conditions and for trucks. A field study program should be undertaken to provide a more robust estimate of the initial plume rise height under a variety of atmospheric conditions, wind speeds and roadway widths that includes a minimum of three FHWA vehicle size classifications (Class 1-2, class 3-5, class 6-12)

For the modeling chain evaluation both the AERMOD-AREAPOLY and AERMOD-RLINE dispersion models performed poorly in comparison to the tracer evaluation. The most likely cause is the emission modeling step using EMFAC and EPA's AP-42 emission factor procedure for road dust emissions. Further testing of the regulatory modeling chain is needed to clearly identify the sources of uncertainty and the relevant importance of emission source contribution. This would include a longer period of air quality monitoring both on-site for ambient background and downwind concentration in the near road environment for both PM2.5 and CO along with on-site traffic data. Additionally, other types of field studies could also be considering such as tunnel studies and real-time instrumentation for more direct measure of particulate matter.

B.1 MODELS EVALUATED

For this example of a partial model evaluation assessment, we evaluate and compare the refined air quality models

- AERMOD-RLINE
- AERMOD

AERMOD is EPA's preferred model for near-field air quality dispersion. It offers flexibility in terms of specifying terrain, meteorology, source setup, and outputs, which can include hourly air concentrations if needed, though it has not been validated against field data for transportation features like barriers and depressed roadways. EPA maintains and regularly updates the model. AERSURFACE (current version released in 2020), while not a regulatory program in the AERMOD system, can help supply modeling values for surface roughness, albedo, and Bowen ratio.

An update in 2019 incorporated the RLINE model's treatment of dispersion from emissions from near-surface transportation sources. The model formulation is considered a true line source treated as a series of point sources whose dispersion results are integrated. However, use of the RLINE feature in this model update is limited to flat terrain settings (as complex terrain cannot be used). The line source formulation as used in RLINE is a beta option in the current version of AERMOD (v21112) and in EPA's current PM Hot-spot Guidance. The beta option is considered to have had sufficient review to meet requirements for consideration as an alternative model when there is no preferred model.[6] This assessment focuses on evaluating how the RLINE source characterization algorithm compares with the current line source algorithm for a variety of metrological conditions both in the near and far-field distances from the freeway.

B.2 REGULATORY APPLICATION(S) FOR MODEL EVALUATION

The regulatory evaluation uses co-located measurements of air pollutants operated at the near-road SLAMS air quality monitoring trailer collected over a 1-year period. The regulatory

---

6 Per the updated PM Hotspot Guidance (EPA-420-B-21-037, October 2021):
The latest version of AERMOD now also includes two additional source types to represent line sources: "RLINE" and "RLINEXT" (for RLINE-extended). RLINE is a Beta feature, meaning its use requires alternative model approval (see Section 7.3.3), and RLINEXT is an Alpha feature, meaning it is for research purposes only. In limited cases, an alternate model for use in a PM hot-spot analysis may be considered. … However, should a project sponsor seek to employ a new or alternate model for a particular transit or highway project, that model must address the criteria set forth in Section 3.2 of Appendix W. Determining model acceptability in a particular application is an EPA Regional Office responsibility involving consultation with EPA Headquarters, when appropriate. https://www.epa.gov/system/files/documents/2021-10/420b21037.pdf

evaluation includes measurements of CO and PM2.5 with 1 year of meteorological data for the same year measured at nearby meteorological station. All data cover the year 2019.

These two pollutants are released as a result of vehicles traveling on the roadway. Consistent with regulatory evaluations, we determine the design values (DV) consistent with the form of the NAAQS standards for 1-hour and 8-hour CO and 24-hour and annual PM2.5. Traffic data is based on local site-specific traffic data collection for year 2019 as available through Caltrans PeMS records. Along with the observed air quality values measured at the SLAMS site, we pair background concentrations from the nearby middle scale air monitoring station (AIRS ID: 06-013-1004) located to the west of the freeway so has frequent exposure to the relatively clean air. Both sites employed the same instrument types to measure CO and PM2.5.

The on-site traffic data exceeds that of typically available for regulatory assessments. That is, any error attributable to the traffic data in this analysis is expected to be less than that for future year transportation modelling used for regulatory analyses for NEPA and conformity. This high-quality dataset is designed to mirror how a regulatory evaluation would be performed, but with higher quality information so that more robust model evaluation statistics can be determined. Thus, it includes both a design value and scatter plots involving all modeled hours. That is, the graphical and quantitative evaluations include a full year of data points for 24-hour average PM2.5, 1-hour CO and 8-hour CO concentrations, along with a comparison of the design values for each pollutant and averaging period.

B.3 CONSULTATION

<Example> *Consultation on the draft model evaluation plan was conducted with stakeholders including sponsoring state DOTs, US DOT, AMPO, MPOs, and other transportation organizations that sponsored the research. All data and information considered in the assessment were made available for the model evaluation*.

- *Key results from peer review included: <list in priority order, if possible>*
- *Key results from stakeholder consultation included: <list in priority order, if possible>*
- *Key results from public consultation included: <list in priority order, if possible>*

Table B-3 shows the statistical measures for AERMOD with AREAPOLY source configuration when compared against observed data for the tracer evaluation. The RHC for both PDCB and PMCP are only slightly above the observed RHC with only PMCH tracer below. The fractional bias is best for PMCH but with all values well within a factor of two showing good overall model performance when emissions rates are well defined.

B.4 LITERATURE REVIEW

We reviewed key papers used by EPA to recommend plume height and initial vertical dimension from vehicles most notably under stable conditions. The key reference was the paper by Gilles et. al (2005) which reported that the initial vehicle height is scaled by a factor of 1.7 to determine release height. However, this was based on a small fleet of nine vehicles with no vehicle higher than 3.3 meters. We suspect that for taller vehicles this 1.7 factor may be an overestimate. Additionally, as noted by Gilles that in addition to atmospheric stability, the shape of the vehicle, and the angle of the wind direction with respect to the direction of vehicle travel play a role in determining this scaling height factor. Further study is needed to better estimate this scaling factor as currently estimated as it likely overestimates the release height and initial vertical dispersion so that when this parameterization is coupled with the better source characterization line source algorithm in AERMOD-RLINE (numerically integrating point sources) the model underpredicts at near and far-field locations.

B.5 FIELD DATA COLLECTION AND ANALYSIS

The field experiment was undertaken in 2021 along a section of the Interstate 80 freeway running through Berkeley, CA in Alameda County. The site offered an excellent platform under which to evaluate the AERMOD-RLINE source type in comparison to the current line source algorithm used in AERMOD. A total of ten tracer experiments were conducted on different days over a variety of traffic and meteorological conditions. This high-volume freeway (280,400 AADT) with 4.8% truck fraction (13,460 AADT) segment runs parallel to Aquatic Park in Berkely, CA oriented in nearly a north-south direction (Figure B-1). There are jersey barriers but no noise barriers on either side of the freeway. Land-use in the immediate vicinity is open water to the west of the project site and a small lake and park land to the east of the study area, with l to 2 story structures further east.

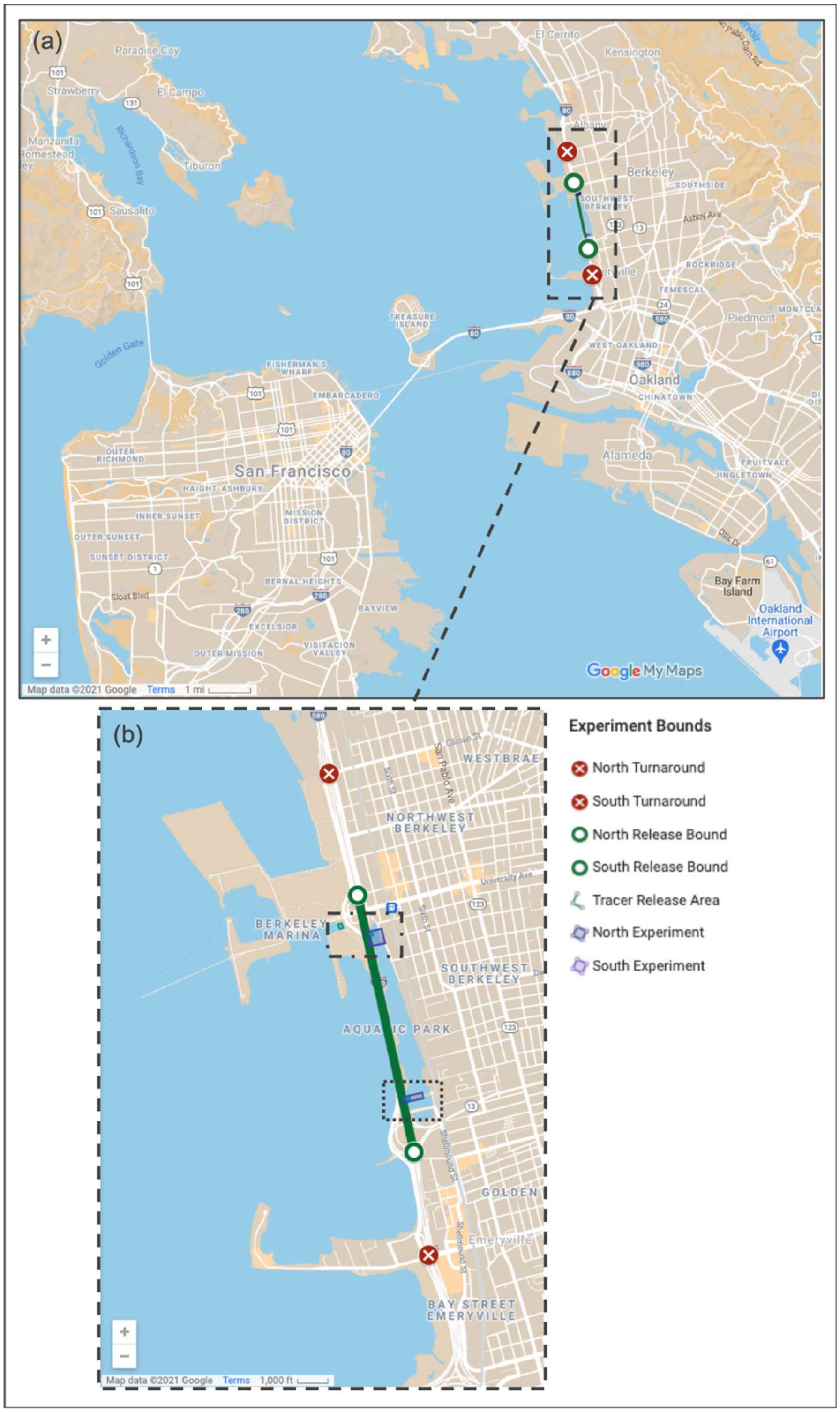

Figure B-1. The study area along I-80 in Berkeley, CA, with greater detail for the boxed area indicated in (a) shown in subpanel (b). The mobile platform turnaround points are marked by the red circles with a white X and the tracer release zone is shown by the green line bounded by the open green circles. The two experimental areas noted by the two boxes in (b) and shown in greater detail below in Figure B-2. Maps are from Google and oriented vertically north-south, with scale noted in the lower left text box of each subplot.

B.5.1 Tracer Data

A total of ten 1-hour tracer experiments were completed using three distinct tracers released from mobile platforms while traveling along the freeway. Each vehicle released a unique tracer. The multiple platforms were used primarily to assess if a sufficient number of vehicles were employed to achieve steady-state concentrations. The emission source strength varied with each experiment. Each hour observation is independent from other hours. Figures B-2 and B-3 show the facility layout for the north and south locations for the tracer experiment. Tracer gas was carefully metered and release using a GPS triggered solenoid for each of the moving vehicles to assure tracer gas was only released at the exact locations along the route. The resulting database of information contains all the detailed measurements from the tracer study needed for air quality modeling this includes:

- Tracer release rate for each of the three tracers used, location for each tracer collection bag (receptor), the tracer emission rate, and duration of each experiment.
- Information on the traffic volumes, truck volumes, and speeds in each direction during the experiment.
- Meteorological conditions during the experiment.

Figure B-2. North Experimental – North end of Aquatic Park, Berkeley, CA. Tracer gas samplers were placed at downwind target distances of (a) 1, (b) 20, and (c) 100 m. All inlet heights were >1 m above ground and were placed away from potential obstacles that would bias sampled air flow.

Figure B-3. In the South Aquatic Park Experimental Area, samplers were placed at downwind target distances of 1, 9, 18, 30, 50, and 100 m. The inlet height at the 1 m distance (a) was extended to a height about 3 m above the sampling area ground level, while the inlets of all other samplers were elevated to about 0.7 m on the backs of folding chairs (b, c). Note the passing mobile tracer release platform in (d) relative to the elevated inlet for the nearest highway sampler also shown in (a).

Complete details of the tracer experiment can be found in the NCHRP 25-55 final report, *Assessment of Regulatory Air Pollution Dispersion Models to Quantify the Impacts of Transportation Sector Emissions*, Appendix B – Tracer Experiment.

B.5.2 Near Road Ambient Air Quality Monitoring Data

In addition to the tracer data collection, co-located measurements of air pollutants as measured by the local air district (Bay Area Air Quality Monitoring District) were also recorded. This site is a near-road SLAMS air quality monitoring trailer that records continuous measurements of: NO, NO2, CO, continuous (hourly) PM2.5, and black carbon. It includes Federal Equivalent Method (FEM) for PM2.5. Continuous monitoring began at the site on July 1, 2016. The distance from the road to probe inlet is 8 meters. Hourly concentrations of CO and PM2.5 were paired with background concentrations as measured nearby middle scale air monitoring station (AIRS ID: 06-013-1004) located to the west of the freeway. This dataset formed the basis for the regulatory model evaluation.

B.5.3 Traffic and Emissions Data

Traffic and emissions information from the backbone of the regulator modeling chain as these determine the emission rate that's a direct input to the air dispersion model. Traffic data was collected to support the regulatory modeling analysis as the emission rate needed to determine based on vehicular traffic. To estimate the emission rates for CO and PM2.5 the team used the California Air Resources Board's (CARB) emission factor model (EMFAC), which is similar to MOVES but based on on-road mobile source regulations unique to California. The traffic data includes vehicle counts collected by a traffic engineering firm that specializes in the collection and analysis of traffic, transportation, transit, and parking data. The traffic volume was extracted from Caltrans Performance Measurement System (PeMS).

The traffic survey was performed over a 24-hour period (midnight-midnight) for vehicular traffic flow in both directions of the I-80 freeway near the Aquatic Park Berkeley. To collect the traffic count information video cameras were installed to collect video footage for a 24-hour period. The recording cameras were located to allow views of vehicle axles. Following the collection of video recordings, the files were reviewed by trained staff which categorized traffic information into 15-minute blocks for each direction. Vehicle traffic volumes were summarized every 15-minutes into the 13 standard FHWA vehicle classification scheme.

Traffic volume (vehicles per hour) for all experiment periods and for all of 2019 were collected from Caltrans' performance measurement systems (PeMS) located at or near Aquatic Park. These roadway embedded sensors report vehicle counts, vehicle miles traveled (VMT), vehicle hours traveled (VHT), flow, speed, truck flow, truck proportion (percent), Q (VMT/VHT), truck VMT, truck VHT on an hourly (and even every 5 min) interval by direction and lane from individual sensors.

The EMFAC2021 model was used to determine the emission factors (g/mile) for CO and PM2.5 for use in regulatory evaluation of the air quality models. EMFAC contains for each California county default fleet age and vehicle mix distributions based on vehicle registration information. The model was run using Alameda County characteristics. The output of the EMFAC2021 model is combined with the combined with the vehicle count data from PeMS to determine the emission rate (g/s) for the regulatory air dispersion modeling.

Further details of the traffic and emissions can be found in the NCHRP 25-55 final report, *Assessment of Regulatory Air Pollution Dispersion Models to Quantify the Impacts of Transportation Sector Emissions*, Appendix C – Dispersion Model Intercomparison.

B.5.4 Meteorological Data

For the tracer modeling study, we used the on-site meteorology data. The tracer dataset included hourly averaged data from May 3, 2021, June 6, 2021, and June 24, 2021, experiment days. The datasets were compiled for a one-hour averaging period consistent with the temporal resolution of the air quality models.

The 2-D sonic anemometer located 200 meters west of the north Aquatic Park location was used for the ambient measure of wind speed and direction for both the north and south Aquatic Park locations as the 2-D sonic anemometer was located at a background location upwind of interference from I-80 traffic. The 2-D sonic recorded data every minute from which hourly averages were determined for the

- horizontal wind speed and direction, and
- standard deviation of the horizontal wind direction.

Additional meteorological and surface characteristic parameters needed by the dispersion models but not measured on-site are:

- surface air temperature,
- cloud cover,
- upper-air temperature,
- Bowen ratio (ratio of sensible heat to latent heat),
- albedo, and
- surface roughness.

Meteorological data from the Oakland Airport Automated Surface Observing System (ASOS) and upper-air radiosondes were used for the first three bullets above. The latter three bullets were derived using land use and land cover information as described in the next section.

For the regulatory modeling we followed EPA and local BAAQMD guidance on how to prepare meteorological data for use in air dispersion models, as would be the case in an EIS/EIR and/or a NEPA assessment. The most representative one-year set of high-quality hourly meteorological data available was from the 2019 Oakland Sewage Treatment Plant (STP) archived by the BAAQMD. The plant is located approximately 4.5-km south of the BAAQMD near-roadway air quality monitoring trailer, with similar orientation to the Bay at key influence in wind direction and speed. However, the observational dataset has frequent gaps in the wind measurements. An automated weather station with a similar land-water exposure is operated by the National Buoy Center (OKXC1) located at Pier 34, which is only 3.8-km southwest of the STP site, and its hourly wind and temperature data were used for substitutions, bringing the completeness up to 99.6%, 99.0%, 99.8%, and 93.1% for each quarter. Only 41 hours out of the 8,760 hours have key data missing for which AERMOD is unable to determine a concentration. This dataset is representative of a complete one-year dataset with available on-site turbulence parameter of standard deviation in wind direction that would be used in a typical regulatory application. No adjustment for friction velocity was needed since we used meteorology data that have on-site measurements of turbulence.

B.5.5 Land Use and Land Cover

For the tracer modeling, the land-cover pre-processor (AERSURFACE[7]) was used in determining surface characteristics (Bowen ratio, albedo, and surface roughness). AERSURFACE uses lookup tables to estimate the surface characteristics based on land cover within a certain distance of the meteorological site (10-km radius for the albedo and Bowen Ratio; 1-km radius for surface roughness). AERSURFACE by default uses inverse-distance

[7] AERSURFACE User Guide for the AERMOD Tool, EPA-454/B-20-008, EPA/OAQPS, February 2020

weighting for calculations of surface roughness ("ZORAD"). We used the latest version of AERSURFACE (20060), which accommodates more recent versions of available land-cover data available from USGS NLCD (years 2001, 2006, 2011, and 2016)—we used the 2011 data, as the Aquatic Park location has not undergone any major changes in land-cover characteristics over the past ten years. We followed the local air district's (BAAQMD) recommendations for seasonal land-use settings consistent with a Mediterranean climate (dry, relatively cool temperatures during the summer and fall and wet winters), with surface moisture conditions characterized using 2019 data derived from Berkeley monthly precipitation data (site 04069) compared with 30-year climatological means for determination of the Bowen Ratio.

Because AERMOD estimates of concentrations are particularly sensitive to the treatment of surface roughness (Long et al. 2004[8]), we used the full spatial resolution permitted by AERMOD for determining values of surface roughness by direction (separate values for each of twelve radial directions, with 30˚ spacing) using the default inverse weighting scheme (ZORAD).

A threshold hourly average horizontal wind speed of 0.2828 m/s was used in the model for "on-site" winds, although data collected by the 2-D sonic with speeds below this level were used in computing the hourly averages and turbulence. The 0.2828 m/s threshold wind speed (determined as (2*0.2*0.2)0.5) is based on AERMOD's default values of 0.2 m/s for the standard deviation in the vertical and the normal direction. AERMOD's AERMET[9] meteorological pre-processor (version 21112) was used to integrate the on-site and off-site data. Because the processors operate on a minimum of 24-hours data, we processed an entire day at a time.

For the regulatory modeling we used EPA's AERSURFACE (v20060) with 2016 land-cover data and the geographical coordinates of the Oakland STP site to determine surface characteristics needed by AERMOD, including twelve radial sectors within 1 km for estimation of surface roughness. AERSURFACE was run separately specifying dry, average, and wet surface moisture, with the results later used to create composite surface characteristics by month. The rainfall data for the 30-year period ending 2019 were gathered and 30-year monthly averages computed for Berkeley, CA (cooperative # 040693). The Oakland Sewer Treatment Plant (STP) onsite data were used for the ONSITE portion with data from the CMAN buoy OKXC (9414776 - Oakland Berth 34) used to fill-in for missing time periods for wind speed and direction. The MODIFY option in AERMET was used to allow Oakland Airport cloud cover data and temperature to be used when needed for the SURFACE portion of the processing

Further details of the traffic and emissions can be found in the NCHRP 25-55 final report, *Assessment of Regulatory Air Pollution Dispersion Models to Quantify the Impacts of Transportation Sector Emissions*, Appendix C – Dispersion Model Intercomparison.

---

[8] Long, G.E.; Cordova, J.F.; Tanrikulu, S. An Analysis of AERMOD Sensitivity to Input Parameters in the San Francisco Bay Area. In Proceedings of the 13th Joint Conference on the Applications of Air Pollution Meteorology with the Air & Waste Management Association; A&WMA: Pittsburgh, PA, 2004; pp 203-206.
[9] AERMET User Guide for the AERMOD Meteorological Preprocessor (AERMET), EPA-454/B-21-004, EPA/OAQPS, April 2021

B.6 MODELING

With EPA's release of AERMOD (version 19191) a new model feature, RLINE, became available to the air quality transportation specialist. RLINE is a new line source algorithm that is based on the numerical integration of a series of point source emissions. This provides a more realistic approximation of a moving line source release. The RLINE release was a beta feature meaning that the model has been vetted through the scientific community and is waiting to be promulgated as a regulatory option. The RLINE accounts for plume meander during low wind speed conditions. This model evaluation assessment primary purpose was undertaken to evaluate AERMOD-RLINE in comparison to the area source algorithm currently in use in regulatory applications.

B.6.1 Dispersion Model Testing (Tracer)

In this section, we summarize the proposed modeling approach for generating the near-road modeled concentrations used in the model tracer evaluation. Table B-1 summarizes the sources of model inputs and monitored concentrations that was used in the tracer model assessment for AERMOD-RLINE application.

Table B-1. Sources and Information used in the AERMOD-RLINE Tracer Evaluation

Pollutants	3 perfluorocarbon tracers (PMCP, PDCB, PMCH), plus the sum of the 3 tracers
Time Period	5 separate hours: 5/3/2021 at 8–9am, 6/6 and 6/24 at 8–9am and 1040–1140 am (modeled as 11am–12pm for simplicity)
Receptors/ Sampling Locations	All time periods have 3 receptors at north end (same coordinates each day) at 1, 20, and 100-m from the freeway. 6/6/2021 and 6/24 have 6 receptors at the south end (at approx. same coordinates each day but not exactly). 5/3 has 12 at south end. The south end receptors for June experiments were located at increasing distances from freeway at approximately 1, 9, 18, 30, 50 and 100 m.
Number of Tracer Concentration from all experiment hours	9 receptors per tracer experiment x (3 individual and one combined) for a total of 4 per experiment x 4 experiment = 144 receptor-concentration outputs in June; 12 receptors per tracer x (3 individual and one combined) for a total of 144 receptor-concentration outputs in May. First tracer experiment had 15 receptors x 4 tracers for a total of 60 receptor-concentration outputs. Total number receptor-concentration outputs 348. All concentrations are hourly averages.
Background	Verified prior to each tracer experiment non-detect of tracer gas.
On-site Meteorology	Met One AIO sonic anemometer (logged 1-minute average winds - calculated sigma theta (standard deviation of horizontal wind direction). A threshold hourly average horizontal wind speed of 0.2828 m/s was used in the model for "on-site" winds. The 0.2828 m/s is based on AERMOD's default values of 0.2 m/s for sigma-v and sigma-w. AERMOD's AERMET meteorological pre-processor (version 21112) was used to integrate the on-site and off-site data
Other Met Needs	Oakland airport for other surface parameters (cloud cover); Oakland twice-daily radiosonde for upper-air.
Terrain	Flat terrain only for AERMOD-RLINE.

Land-Use/Land-cover	AERSURFACE (20060) pre-processor used to determine surface characteristics based on land-cover within 10-km radius for the albedo and Bowen Ratio; a 1-km radius for surface roughness. Land-cover based on land-use data available from USGS NLD data from 2016. Used local air districts (BAAQMD) recommendations for seasonal land-use settings consistent with a Mediterranean climate (dry, cool temperatures during the summer and fall and wet winters). Surface moisture conditions characterized using 2019 data derived from Berkeley monthly precipitation data (site 04069) compared with 30-year climatological normal for determination of the Bowen Ratio. Highest spatial resolution used (every 30 degrees) for determining values of surface roughness by direction using default inverse weighting scheme.
Emissions	Tracer release rates on average were emitted per release period at a rate of 0.08, 0.09, and 0.07 g s-1 for PDCB, PMCP and PMCH, respectively.
Traffic and Facility Type	High volume freeway (10-lane 280,400 AADT) – used Caltrans Performance Measurement System (PeMS) and on-stie traffic data collection survey on the one experiment day when traffic data was collected. Diurnal profiles for trucks and non-trucks were derived from the hourly data.
Models	AERMOD-RLINE (version 21112)
Source Characterization	While the tracer was not directly released on ramps it was assumed some ramp traffic contributed to the well-mixed highway turbulence. We used a number of AERMOD-RLINE configurations to characterize the complex polygons shapes and included limited sections of ramps and merge lanes directly adjacent to the free-flowing highway lanes. Freeway flow lanes were modeled with AERMOD-RLINE. EPA guidance was followed to calculate source release height (average vehicle height × 1.7 ÷ 2, assuming truck average height of 4.0 m and car average height are 1.53 m, with the average vehicle height calculated with weighting based on each observed hourly truck and non-truck traffic mix. Initial vertical dimension was based on estimated average vehicle heights (average vehicle height × 1.7 ÷ 2.15).

The above information was used as model inputs and then model simulations were performed to evaluate AERMOD-RLINE model using the tracer dataset.

B.6.2 Regulatory Modeling Chain Testing

In this section, we summarize the modeling approach for generating the near-road modeled concentrations used in the model regulatory evaluation. Table B-2 summarizes the sources of model inputs and monitored concentrations that was used in the regulatory model assessment for AERMOD-RLINE application.

Table B-2. Sources and Information used in the AERMOD-RLINE Regulatory Evaluation

Pollutants	CO, PM2.5
Time Period	1 year (hourly) (2019)
Receptors and Sampling Locations	Single receptor located at BAAQMD monitoring trailer site
Number of Receptor*Time Outputs (per pollutant)	8,760
Background Monitor	San Pablo (AIRS 06-013-1004 west of Interstate 80.
On-site Meteorology	None, nearest representative site based on consultation with BAAQMD – Oakland Sewage Treatment Plant (STP) and Oakland NOAA National Data Buoy (OKXC1) Oakland Berth 34 for missing data

Pollutants	CO, PM2.5
Other Meteorological Needs	Oakland airport for cloud cover, Oakland twice-daily radiosonde for upper -air.
Terrain	AERMOD: use of USGS terrain data at highest resolution.
Land-Use/Land-cover	Used EPA's AERSURFACE (v20060) with 2016 land-cover data and the geographical coordinates of the Oakland STP site to determine surface characteristics needed by AERMOD, including twelve radial sectors within 1 km for estimation of surface roughness.
Emissions	EMFAC2021 emission factors with corresponding traffic volume data (2019). Stratify PM2.5 exhaust, brake/tire, and road dust.
Traffic and facility type	High volume freeway (10-lane 280,400 AADT) – used annual data from Caltrans Performance Measurement System (PeMS) for 2019
Models	AERMOD (AREAPOLY)
Sources Characterization	Model the two freeway directions as separate sources using AERMOD-AREAPOLY; same for ramps near monitoring locations that may have measurable impacts on the monitor. AERMOD AREAPOLY was used to characterize the complex polygons shapes and included limited sections of ramps and merge lanes directly adjacent to the free-flowing highway lanes. EPA guidance was followed to calculate source release height (average vehicle height × 1.7 ÷ 2, assuming truck average height of 4.0 m and car average height are 1.53 m, with the average vehicle height calculated with weighting based on each observed hourly truck and non-truck traffic mix. Initial vertical dimension was based on estimated average vehicle heights (average vehicle height × 1.7 ÷ 2.15).

B.7 ANALYSIS OF RESULTS

The key statistical tests for model acceptability for regulatory analysis are the robust highest concentration (RHC) and the fractional bias. These two statistics are the most succinct measures of model performance due to the importance of reproducing the upper end of the distribution of highest values in air quality conformity determinations as well as providing assurance the models are finding the correct concentration for the right reason. They two statistics consider both peak concentrations and the spread of the data in the reporting of a value that is comparable across models. The RHC is calculated by factoring in the probability of high concentration values occurring along with the average concentration of these concentrations above a specifically set threshold. This allows for the evaluation of a single number which considers both the value and likelihood of these high concentrations that may otherwise skew or not be represent the reported values thus giving a measure of the accuracy of the model for the upper end of observed concentration distribution. Representation of the highest concentrations is especially important for regulatory applications, where the model should be flexible enough to predict concentrations that mimic the tail end of the distribution. Comparisons are made between the observed and modeled RHC values.

The fractional bias calculates a fraction using the observed and predicted from the highest set of observed values that is used to represent the overall fit of the model to the data with respect to the peak concentrations. This metric is a unitless value ranging between -2.0 and 2.0, meant to predict model fit to the observed data, where a value closer to 0 indicates more accurate predictions of the model. Ideally these performance statistics for regulatory purposes should fall within a factor of two which is a value between 0.66 and -0.66.

The Pearson correlation coefficient or (r-squared) is a statistical tool used to measure the extent of the relationship between observed and predicted values paired in time and space. It measures the strength in the relationship between observed and predicted values. Results vary from a perfect inverse correlation (-1) to no correlation (0) to perfect correlation (1). Ideally air quality models will have an r-squared near 1.

Another measure used in the weight of evidence are Q-Q plots values of both the observed and modeled data. By plotting these values on the same graph, we can evaluate the distributions of both sets of data to determine how well they correspond. If the model distribution is quite different from observed distribution, it indicates that model is not reliable. The ideal Q-Q plot is a 1:1 line from one corner of the graph to the other, this is unlikely to happen due to inherent uncertainties in modeled and observed data.

These performance metrics are compared against tracer data for the dispersion modeling testing and against the near-road ambient air quality monitoring data for the regulatory modeling chain testing. Review of these three-performance metrics are used in a weight of evidence approach to decide if the AERMOD-RLINE model is an improvement over AERMOD-AREAPOLY and provide insight on why the difference(s) exist between the two and potentially how they can be improved.

B.7.1 Dispersion Model Testing Against Tracer Data

Table B-3 shows the statistical measures for AERMOD with AREAPOLY source configuration when compared against observed data for the tracer evaluation. The RHC for both PDCB and PMCP are only slightly above the observed RHC with only PMCH tracer below. The fractional bias is best for PMCH but with all values well within a factor of two showing good overall model performance when emissions rates are well defined.

Table B-3 AERMOD-AREAPOLY Quantitative Metric under Tracer Evaluation

Statistical Measure	PDCB	PMCP	PMCH	Total
R-squared	0.77	0.83	0.74	0.82
Fractional Bias	0.40	0.22	-0.01	0.24
Robust Highest Concentration – Observed	194.8	261.1	114.2	563.8
Robust Highest Concentration – Modeled	204.3	234.2	144.4	585.3

Figure B-4 and Figure B-5 show the scatter and Quantile-Quantile plots of AERMOD-AREAPOLY modeled concentrations versus observed data. Blue dots represent concentrations at near field receptors (less than 50 m from the roadway) while red dots represent concentrations at far field receptors (≥50 m receptors). As shown, the model tends to have better performance for near field receptors as compared to far-field receptors. The scatter shows that for the near field concentrations almost always fall within a factor of two of the observed value. Similarly, for values unpaired in time and space as needed in regulatory applications the model does an excellent job reproducing the concentration distribution for PMCH, PMCP and total.

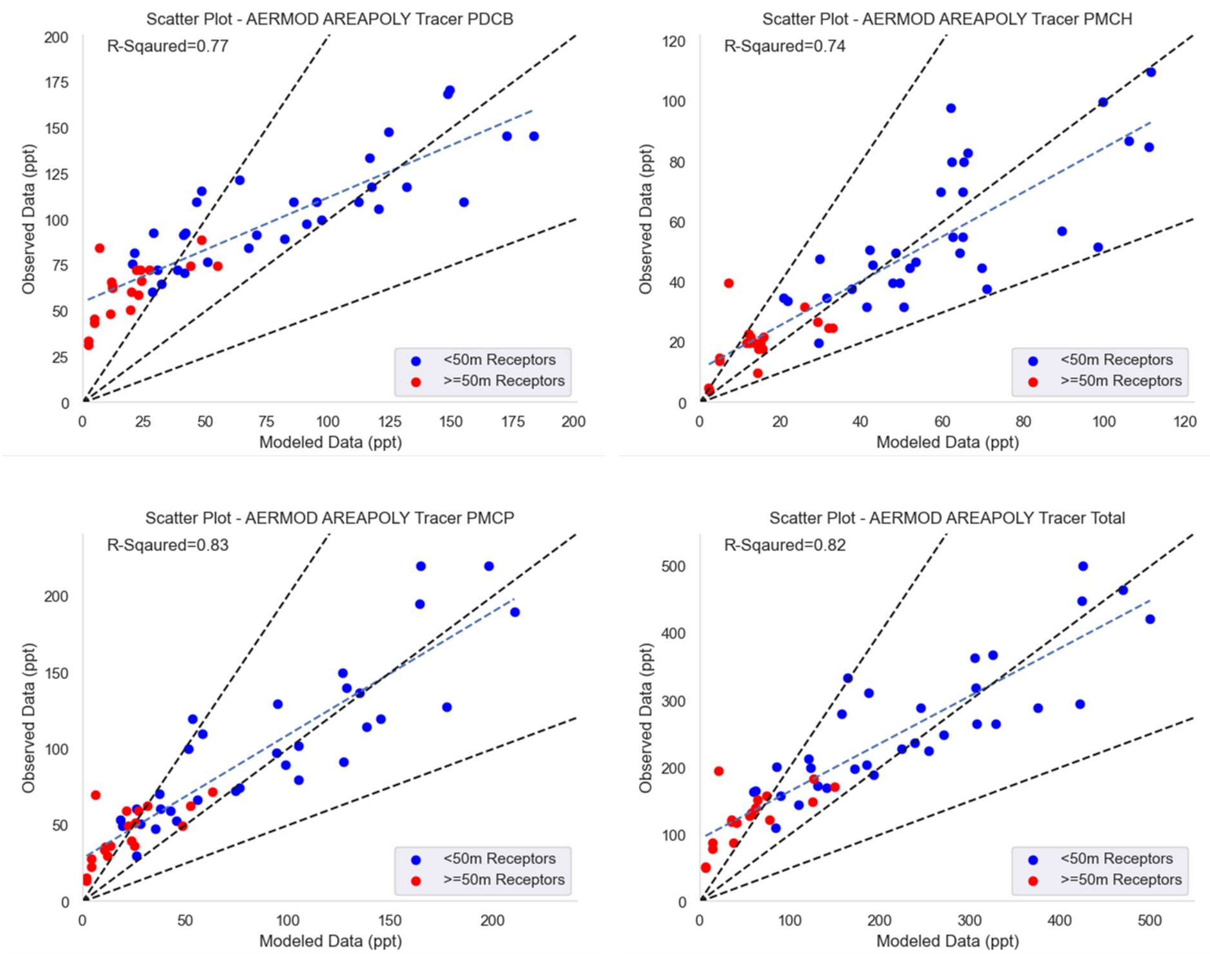

Figure B-4. Scatter Plots of Observed vs. AERMOD Modeled Concentrations (PPT) for Tracer Study

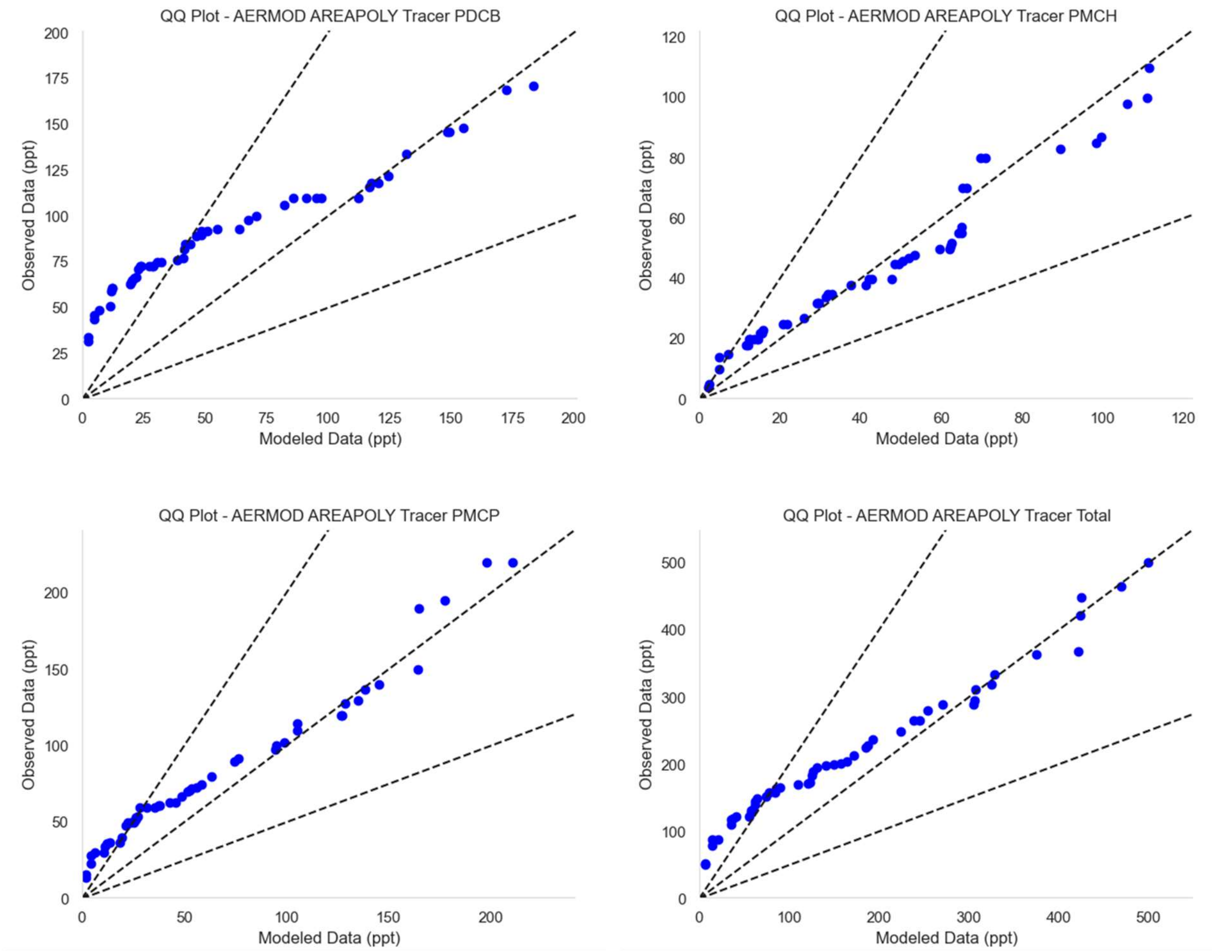

Figure B-5. Q-Q Plots of Observed vs. AERMOD-AREAPOLY Modeled Concentrations (PPT) for Tracer Study (unpaired in time and space)

Table B-4 shows the statistical measures for AERMOD with RLINE source configuration versus observed data for the tracer evaluation. As indicated by the RHC, the AERMOD- RLINE underestimate the concentrations for all three tracers by substantial margins. Similarly, the fractional bias shows the model underpredicting by more than a factor of two. However, the correlation coefficient (r-squared) remains similar to those found using AERMOD AREAPOLY only.

Table B-4. AERMOD-RLINE Quantitative Metric under Tracer Evaluation

Statistical Measure	PDCB	PMCP	PMCH	Total
R-squared	0.78	0.83	0.73	0.82
Fractional Bias	0.86	0.71	0.48	0.72
Robust Highest Concentration – Observed	194.8	261.1	114.2	563.8
Robust Highest Concentration – Modeled	106.9	124.7	72.4	311.6

Figure B-6 and Figure B-7 show the scatter and Quantile-Quantile plots when AERMOD-RLINE modeled concentrations are compared with observed data. Blue dots represent concentrations at near field receptors (less than 50 m from the roadway) while red dots represent concentrations at far field receptors (≥50 m receptors). Similar to AERMOD-AREAPOLY, the model tends to have better performance for near field receptors as compared to far-field receptors.

However, the scatter plots show that for low concentrations and not the high values that would be used in design values a considerable number of predicted vs observed pairs fall outside the factor of two of the observed value particularly for the PDCB tracer and that for most pairings the AERMOD-RLINE underpredicts. Similarly, for values unpaired in time and space as needed in regulatory applications the Q-Q plots show that AERMOD-RLINE has a clear tendency to underpredict. However, the higher concentrations (i.e., nearer the road) are generally within the factor of two.

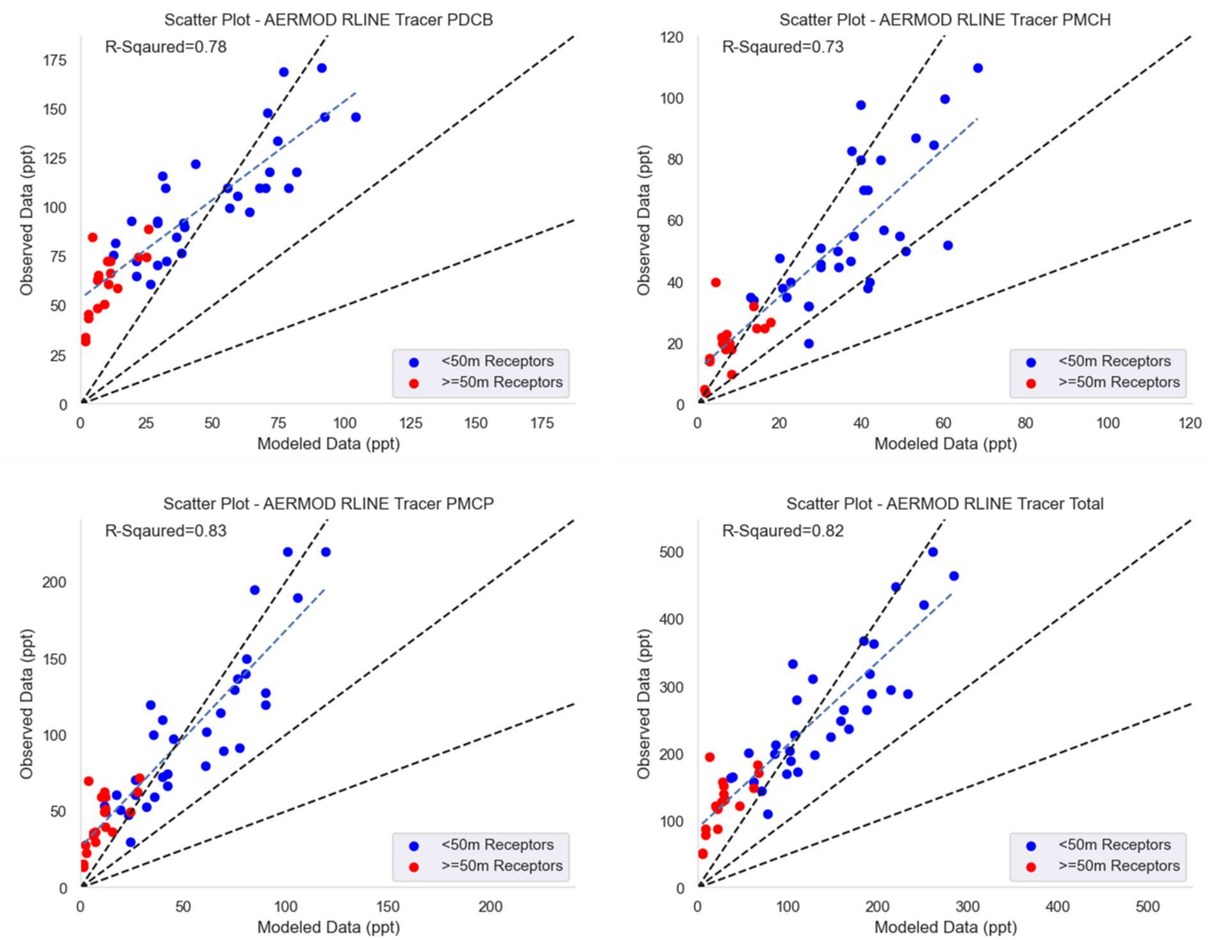

Figure B-6. Scatter Plots of Observed vs. AERMOD-RLINE Modeled Concentrations (PPT) for Tracer Study

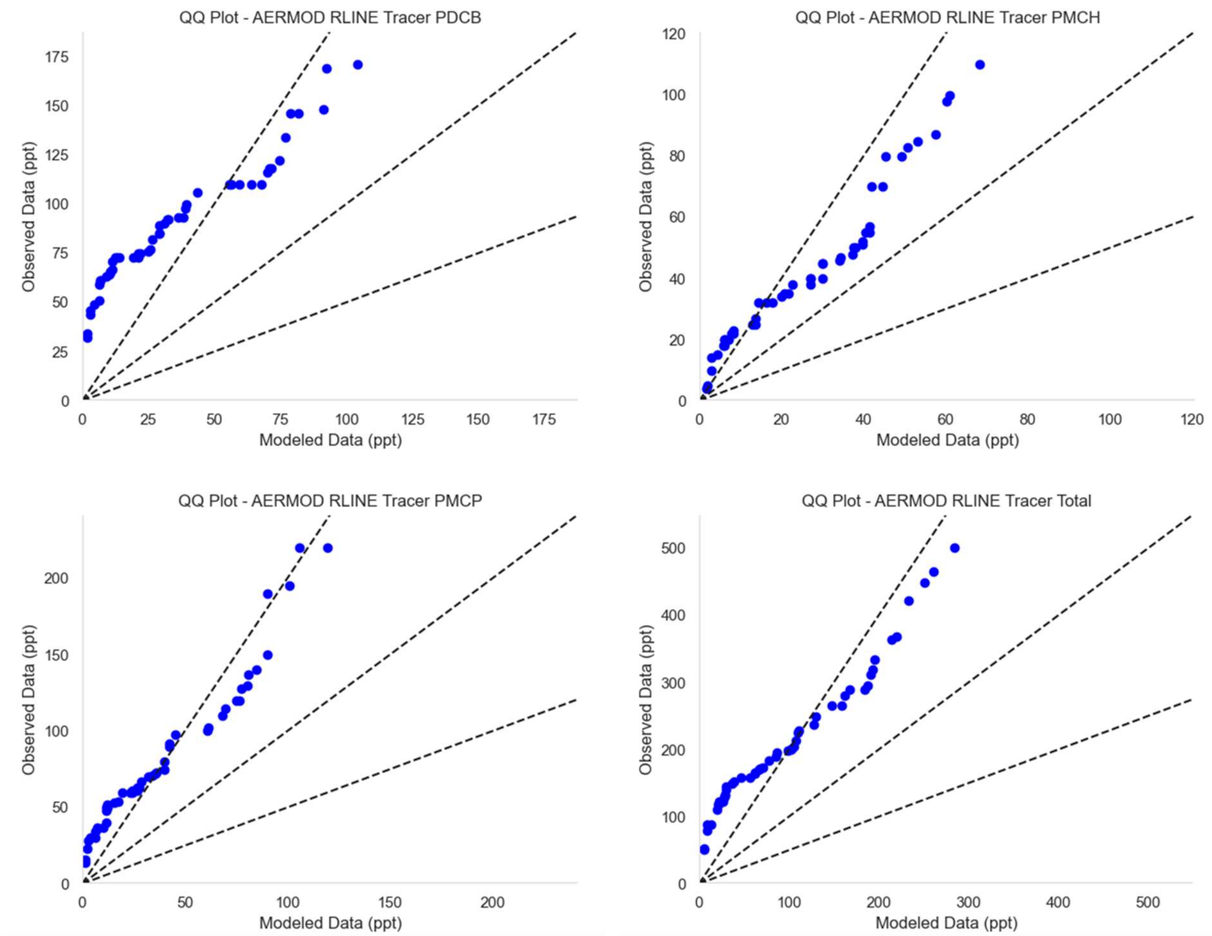

Figure B-7. Q-Q Plots of Observed vs. AERMOD-RLINE Modeled Concentrations (PPT) for Tracer Study (unpaired in time and space)

Other measures as part of the model evaluation process include model parsimony, ease-of-use, run time and uncertainty. Here we discuss how these measures change with the use of AERMOD-RLINE instead of AERMOD-AREAPOLY.

Model parsimony means that the dispersion models and the associated EPA guidance for applications in regulatory settings should be no more complex or resource intensive than needed for the sufficient level of accuracy for the regulatory need.

The model evaluation process has considered the following issues related to model parsimony as it relates to use of AERMOD-RLINE relative to the effort to apply AERMOD-AERAPOLY. Model inputs should be identified based on sensitivity assessments and their associated level of effort to the uncertainty of the output from each model. Based on NCHRP 25-55 research the following level of accuracy of model inputs (Table 1) are needed to maintain model accuracy while retaining parsimony. In general, the same level of accuracy is needed for model inputs when applying AERMOD-AREAPOLY or AERMOD-RLINE. Any differences are noted in Table B-5.

Table B-5. Accuracy Needed in AERMOD Model Inputs to Retain Parsimony When Applying AERMOD-RLINE.

Model input parameter	Information needed	Measurement resolution	Other
Meteorology	Site-specific highly preferrable; 5-yr NWS least preferable	0.1 m/s	Minimum of 0.5 m/s instrument threshold. Available from local air agency or National Weather Service (NWS).
Facility Geometry	Detailed facility layout both spatially and vertically	Nearest meter	GIS – shape file or similar digital design information – may require some additional effort for AERMOD-RLINE expressing the input information as only a line source depending on facility geometry.
Terrain	Highest resolution datasets available	10-meter horizontal resolution	Available from United States Geologic Survey
Land-cover	Recommend most recent available NLCD 2016	30-meter resolution	Available from Multi-Resolution Land Characteristic consortium from National Land Cover Database
Urban/rural setting for urban heat island	Determination based on land-use information is preferrable	Not applicable	This may need to be direction depending on setting
Traffic volume	Traffic volume with fraction as trucks and cars preferably with diurnal profile	Nearest 100 per hour	State DOT will typically have for large facility types

Ease-of-use. The use of AERMOD-RLINE does not place any additional burden on the user over the effort to develop inputs for AERMOD-AREAPOLY. The addition of the AERMOD-RLINE option is a relatively straightforward capability and is easy for users to add at the beginning of the project and does not present a barrier to its use. This does not mean that development of other inputs is easy just that this additional functionality is easy to implement. While the use of an integration of a true line source model as part of the AERMOD does affect model run-time it is generally only a minimal increase in overall runtime.

Quality Assurance and Control (QA/QC): The AERMOD has limited error checking features other than gross error checking of inputs within ranges. However, available commercial interfaces do include additional error checking functionality particularly through data visualization (e.g., facility layout, wind rose, graphical spatial displays and cross-sections of model results as well as some inputs). ADMS-Roads includes much the same visualization tools as the commercial graphical user interfaces (ADMS Mapper) as those that have been developed for AERMOD.

For model results AERMOD does provides information on output in tabular format for US air quality standards but specifying those within the AERMOD input environment can be challenging. However, commercial interface provide additional support to the analyst in assuring outputs are post-processed in a clear and easy to follow manner. ADMS-Roads has some visualization capabilities to review model output using ADMS Mapper but also has links to other software packages such as Surfer and ArcGIS.

Uncertainty *Ideally, sensitivity tests would have been performed for model inputs for which the dispersion model has been identified as most responsive to appropriately characterize the uncertainty associated with the dispersion model. While this effort was not undertaken as part of the NCHRP 25-55, project, key model sensitivity variables have been identified by Wayson and Voigt (2022) for AERMOD-RLINE which included*:

- *Release height*
- *Initial vertical dimension*
- *Winds parallel to roadway*

Sensitivity tests comparing the AERMOD-RLINE with the results from AERMOD-AREAPOLY would provide additional information on how the model behaves with changes to these key variables and if modifications are needed to increase or decrease these sensitivities.

Based on the statistical tests of the RHC and the fractional bias the AERMOD-RLINE shows significantly degraded accuracy in predicting the highest concentrations relative to AERMOD-AERAPOLY. The non-statistical measures for using AERMOD-RLINE show little change relative to using AERMOD-AREAPOLY.

B.7.2 Regulatory Modeling Chain Testing against Near-Road Ambient Air Quality Data

The intended regulatory purpose in this assessment of the modeling chain is a comparison of the accuracy of the design values (DV) for each pollutant against the applicable NAAQS. Table B-6 provides a comparison of the DVs for CO and PM2.5 concentrations modeled with AERMOD-AREAPOLY against the those calculated based on the BAAQMD's near-road monitoring station. The DV is a statistic that describes the air quality status of a given area relative to the level of the National Ambient Air Quality Standards (NAAQS). Here in this study, the 1-hour and 8-hour CO DVs are calculated as the second highest 1-hour, and 8-hour averaged concentrations over the course of the year 2019. For 24-hr PM2.5, the DV is calculated as the 98th percentile of 24-hr average PM2.5 concentration throughout the year. Two sets of comparisons are provided: a) DVs with outliers, and b) DVs without outliers. For the purpose of this study, outliers are defined as any concentration values exceeding average concentrations plus 3 standard deviations. Upon analyzing the observed CO data, we noticed single hour jumps in CO concentrations (5.6 ppm) on November 7th at 2 AM and November 8th at 1 AM. Considering the low traffic volumes at these hours, we speculate that these values were incorrectly measured and thus removed from the analysis. As shown in Table B-6, when outliers were removed, modeled DVs calculated through AERMOD-AREAPOLY were very consistent with DVs calculated from observed data. For 1-hour CO, DVs were different by 0.1 ppm; for 8-hour CO, DVs were the same; for 24-hour PM2.5, the modeled DV was 1 $\mu g/m^3$ lower than observed, and for annual PM2.5, there was 0.5 $\mu g/m^3$ difference.

Table B-6. AERMOD-AREAPOLY Design Values for Regulatory Evaluation[10]

Averaging Period and Pollutant	With Outliers		Without Outliers	
	Modeled	Observed	Modeled	Observed
1-hour CO (ppm)	2.2	5.6	1.2	1.1
8-hour CO (ppm)	1.3	1.2	1.0	1.0
24-hour PM2.5 (μg/m^3)	20	20	18	19
Annual PM2.5 (μg/m^3)	9.0	9.5	8.8	9.3

Figure B-8 and Figure B-9 are scatter and Q-Q plots, respectively, comparing modeled vs. observed 1-hour and 8-hour average CO as well as 24-hour average PM2.5 concentrations11. As shown, the modeled concentrations (with background concentration added) were well correlated with observed data. However, background concentration contributes more than 80 percent (across various averaging periods) to the total CO and PM concentrations at the near-road site. Therefore, when compared, it is not surprising to see relatively low bias between modeled and observed concatenations. The Q-Q plot show modeled CO concentrations tend to be lower than the observations. Because the tracer evaluation should good model performance the underestimation here is more likely due to relatively high uncertainties with modeled emission factors (e.g., emissions deterioration, in-use performance relative to certification driving cycle). Figure 8 shows the rank ordered Q-Q plots (unpaired in time) as would be done as part of a regulatory evaluation of the modeling chain. Here the AERMOD-AREAPOLY results show good model performance particularly for the highest observed concentrations.

[10] Consistent with the form of the NAAQS and reporting requirements, annual PM2.5 are reported to one decimal place 24-hr PM2.5 values are rounded to nearest whole number. The form of the CO NAAQS is also integer values, however this not specified in EPA's CO Hotspot Guidance and U.S. EPA reporting of design values are typically done to one decimal place. We have repeated that here for 1-hr and 8-hr CO DVs.

[11] Only data without outliers are presented in these charts.

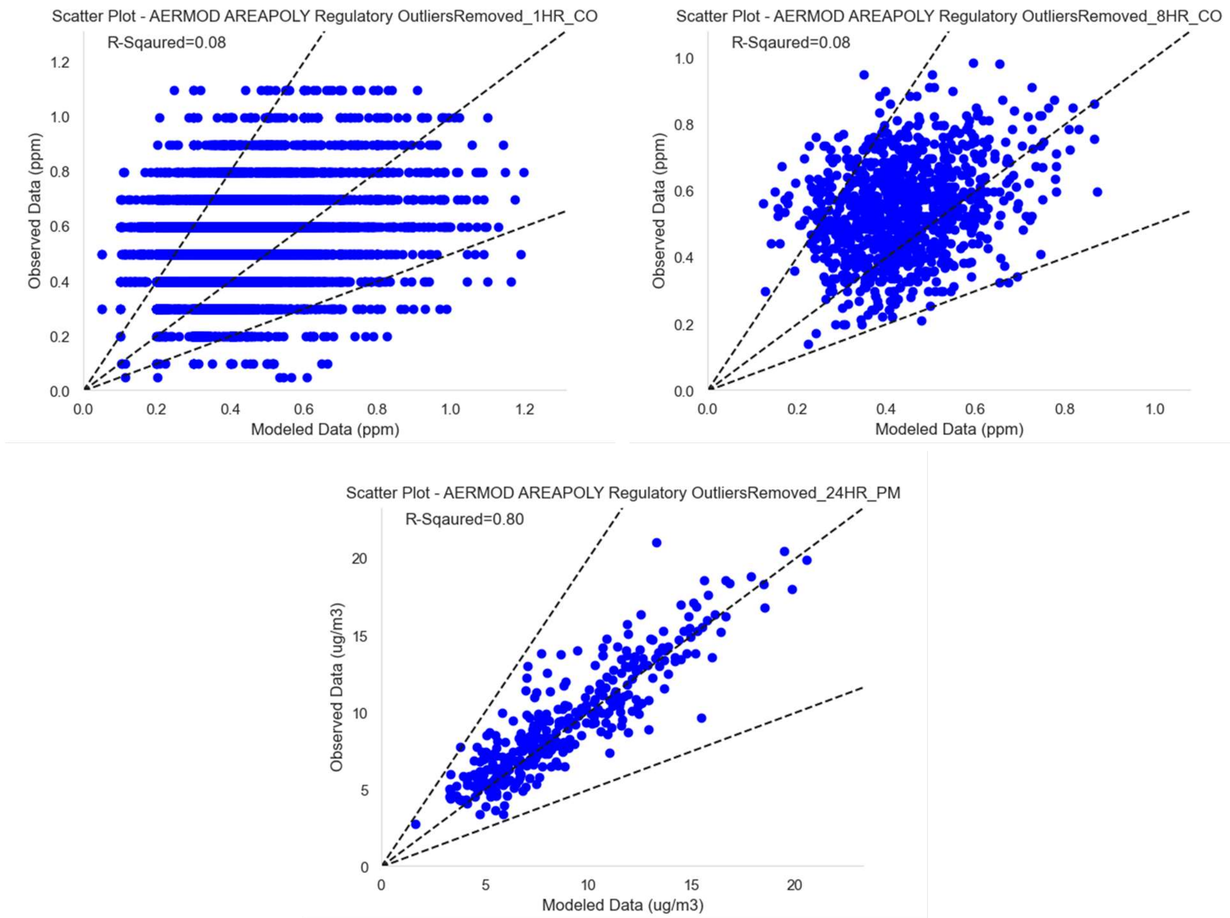

Figure B-8. Scatter Plots of Observed vs. AERMOD-AREAPOLY Modeled Concentrations for Regulatory Evaluation (n=8,760 for 1 hour CO, n=1,095 for 8-hour CO, and n=365 for 24-hr PM2.5)

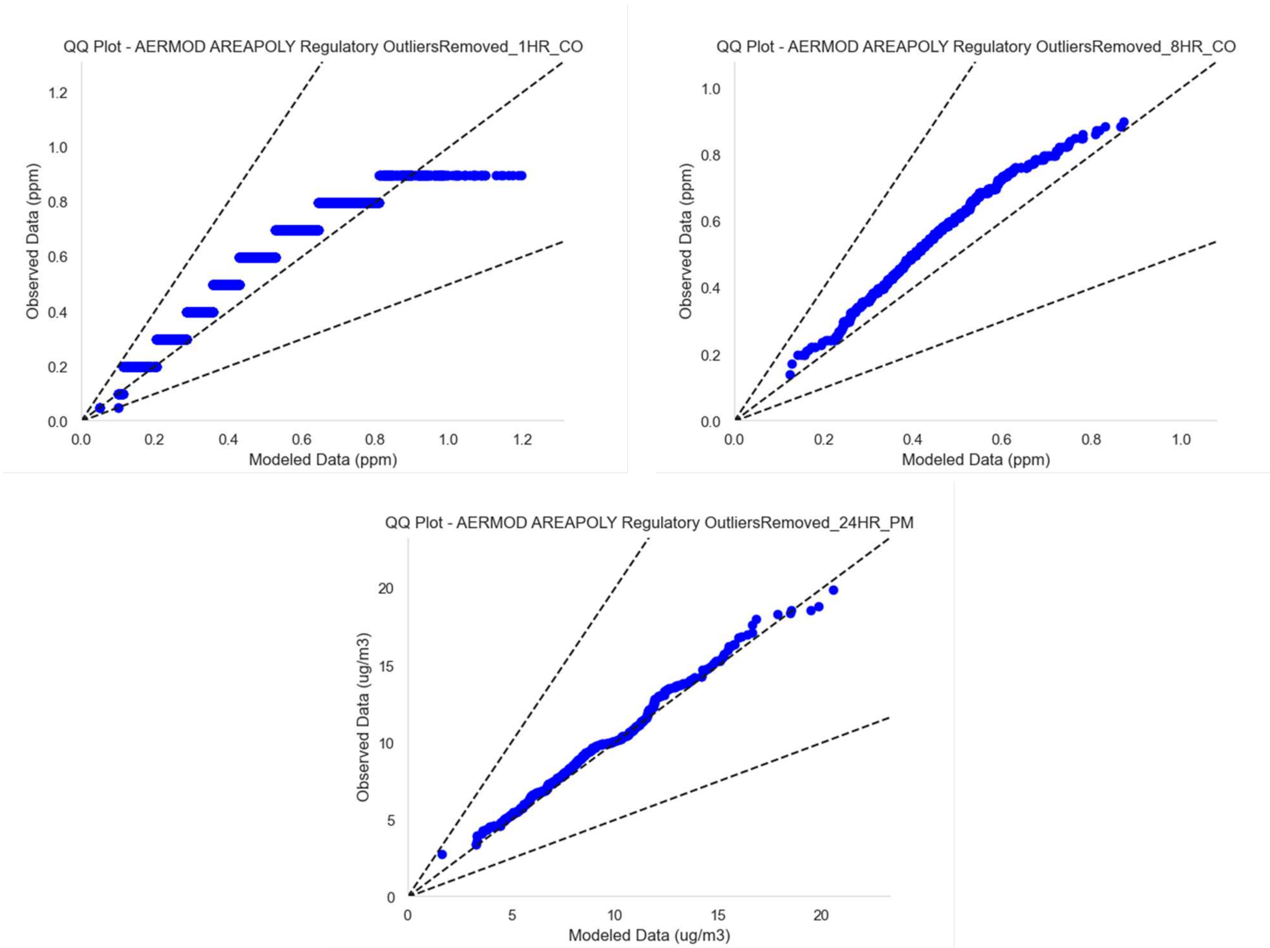

Figure B-9. Q-Q Plots of Observed vs. AERMOD-AREAPOLY Modeled Concentrations for Regulatory Evaluation (n=8,760 for 1 hour CO, n=1,095 for 8-hour CO, and n=365 for 24-hr PM2.5)

Table B-7 provides a comparison of the DVs for CO and PM2.5 concentrations against the DVs calculated based on near-road monitoring station. As described for the AREAPOLY assessment two sets of comparisons are provided: a) DVs with outliers, and b) DVs without outliers. As shown in Table B-7, when outliers are removed, DVs calculated through AERMOD-RLINE are quite similar to design values calculated from observed data, although the AERMOD-RLINE modeled values tended to be slightly lower than observed on average. For 1-hour and 8-hour average CO, DVs are different by 0.1 ppm or less; for 24-hour PM2.5 the modeled design value is 1 µg/m3 higher than observed, and for annual PM2.5 there is 0.4 µg/m3 difference.

Table B-7. AERMOD-RLINE Design Values for Regulatory Evaluation[12]

Averaging Period and Pollutant	With Outliers		Without Outliers	
	Modeled	Observed	Modeled	Observed
1-hour CO (ppm)	2.0	5.6	1.2	1.1
8-hour CO (ppm)	1.3	1.2	0.9	1.0
24-hour PM2.5 (μg/m^3)	22	20	20	19
Annual PM2.5 (μg/m^3)	9.9	9.5	9.7	9.3

Figure B-10 and Figure B-11 are scatter and Q-Q plots, respectively, comparing modeled vs. observed 1-hour and 8-hour averaged CO as well as 24-hour averaged PM2.5 concentrations. Similar to the AERMOD-AREAPOLY evaluation, the Q-Q plot demonstrates that overall, most modeled CO concentrations were lower than compared to the observed data. As noted earlier, model performance here is a combination of the dispersion model, the emissions modeling, and background concentrations.

[12] Consistent with U.S. EPA reporting of design values, the DVs for 1-hr CO, 8-hr CO, and annual PM2.5 are reported with one decimal, while it is rounded to nearest whole number for 24-hr PM2.5.

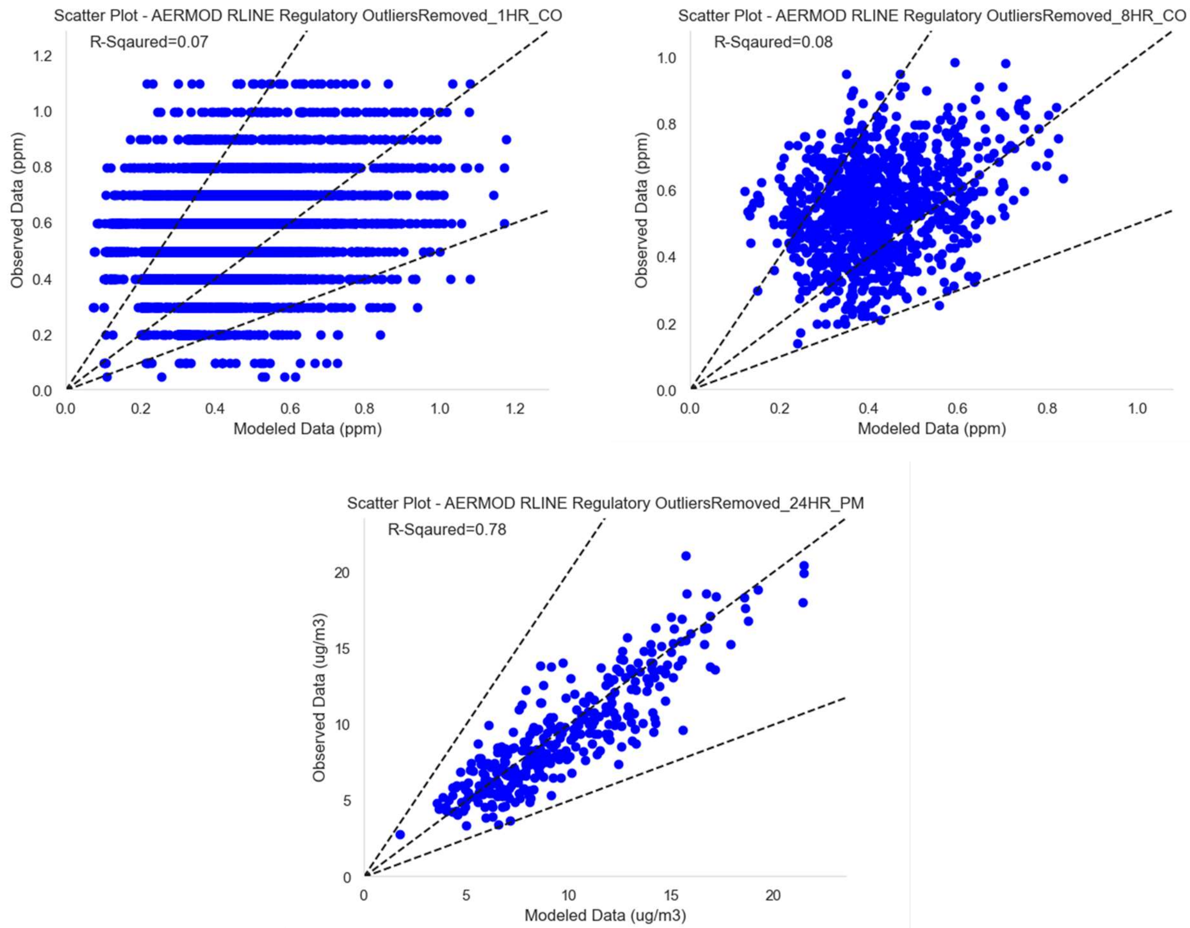

Figure B-10. Scatter Plots of Observed vs. AERMOD-RLINE Modeled Concentrations for Regulatory Evaluation (n=8,760 for 1 hour CO, n=1,095 for 8-hour CO, and n=365 for 24-hr PM2.5)

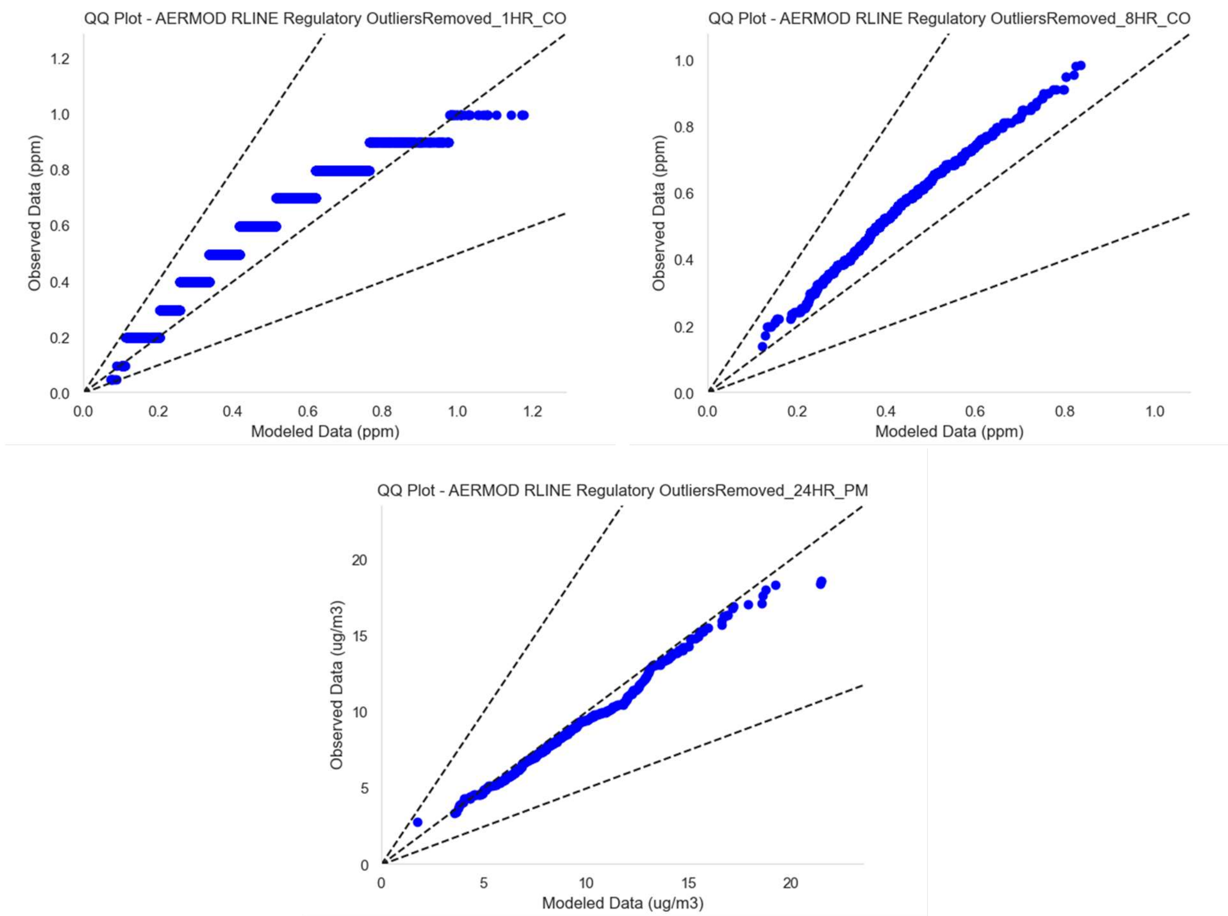

Figure B-11. Q-Q Plots of Observed vs. AERMOD-RLINE Modeled Concentrations for Regulatory Evaluation (n=8,760 for 1 hour CO, n=1,095 for 8-hour CO, and n=365 for 24-hr PM2.5)

B.7.3 Comparison of Dispersion and Regulatory Modeling Chain Testing

The results show that the AERMOD-AREAPOLY and AERMOD-RLINE dispersion models all performed much better against tracer data than did the regulatory modeling chain did against near-road monitoring data for PM2.5 and CO. For example, the r-squared for the tracer experiments for AERMOD-AREAPOLY and AERMOD-RLINE range from 0.73 to 0.83 depending upon the tracer, while for PM2.5 and CO paired in time and space the r-squared ranged from just 0.07 to 0.08. Similarly, the AERMOD-AREAPOLY and AERMOD-RLINE for the PM2.5 and CO paired in time and space had similar low correlations clearly identifying that the source of the uncertainty lies outside the dispersion component of the modeling chain.

The other sources of input to the modeling chain are traffic volume, background concentration and vehicle emission factor. The traffic data for the study were obtained from the on-site PEMS data and so is considered well-characterized and therefore not likely to be a significant source of error in the regulatory modeling chain. The backgrounds concentrations were determined from nearby monitoring station and are unlikely to be a significant source of error in the regulatory modeling chain framework. The most likely cause is the emission modeling step using EMFAC and AP-42 for road dust in the regulatory modeling chain.

Evidence from the 25-55 study shows that the dispersion component of the model chain has high accuracy when emission rates are well known. Further testing of the regulatory modeling chain is needed where emphasis on data collection would include a longer period of air quality monitoring both on-site for ambient background and downwind concentration in the near road environment for both PM2.5 and CO along with on-site vehicle count and fleet mix information. Additionally, to evaluate the emission rates from emission factor models (EMFAC and MOVES) measurements could be made using real-time instrumentation as part of a tunnel study. Tunnel experiments have well defined volumes of air exchanged so that measured differences in PM2.5 and CO concentration between inlet and outlet locations can be determined for comparison with the emission factor model output.

B.7.4 Comparative Analyses

While no formal comparative analysis was conducted under the NCHRP 25-55 study, the following findings and insights were made based on results of the technical analyses conducted during the research project.

- Model comparisons showed better performance for AERMOD when applied using AREAPOLY source type than RLINE. This is likely a result of the current EPA guidance on estimating release height and initial vertical dimension being developed prior to the incorporation of RLINE into the AERMOD model. Key papers used by EPA to recommend plume height and initial vertical dimension from vehicles most notably under stable conditions were reviewed. A key reference was the paper by Gilles et. al (2005) which reported that the initial vehicle height is scaled by a factor of 1.7 to determine release height. However, this was based on a small fleet of nine vehicles with no vehicle higher than 3.3 meters. We suspect that for taller vehicles this 1.7 factor may be an overestimate.

- Additionally, as noted by Gilles that in addition to atmospheric stability, the shape of the vehicle, and the angle of the ambient wind with respect to the direction of vehicle travel play a role in determining this scaling height factor. Further study is needed to better estimate this scaling factor as currently estimated it likely overestimates the release height and initial vertical dispersion so that when this parameterization is coupled with the better source characterization line source algorithm in AERMOD-RLINE (numerically integrating point sources) the model tends to underpredict concentrations.

B.7.5 Determinations of Model Adequacy

This section includes two main categories for our model evaluation example. Adequate with identified limitations and inadequate and is not recommended for use at this time in these current settings for regulatory applications.

<u>ADEQUATE, WITH LIMITATIONS</u>: Cases in which model performance has been determined to be adequate, with any limitations as stated, based on an assessment against standard criteria in an approved model evaluation.

- Mode: Refined AERMOD - AREA, LINE, or AREAPOLY
- Regulatory Test: AERMOD applied following EPA PM hot spot guidance showed acceptable model performance in characterizing near road concentrations

- Setting: urban although much of the upwind direction rural

Facility Types: Freeway (Free-flow and congested)

- Configuration: 10-lane freeway (280,400 AADT)
- Operating Condition: Free-flow and moderately congested (hourly traffic volume range: 850 – 1,460 per lane), truck percentage range: 4.2 – 7.9
- Near-Road Barrier: None
- Nearby Sources (rail, stacks, etc.): Absent (not within 200 meters from roadway)
- Meteorology: Wind speed range (1.0 – 3.0 m/s, wind direction 10 to 64 degrees from parallel to roadway)

NOT SHOWN TO BE ADEQUATE: This includes both cases in which model performance has been determined to NOT be adequate based on an assessment against standard criteria in an approved model evaluation plan, and that an adequate database of a tracer dataset is available for comparison analysis.

The remainder of this section would include all other transportation applications (e.g., intersections, interchanges, arterials, skewed intersections) for which model evaluation has not been done against representative field data. It would include a caveat that it does not mean that they would not work, just that the results may be suspect as the needed model validation has not yet been done.

Dispersion model applications that are only marginally different than NCHRP 25-55 study and can likely be applied without further evaluation only requiring a warning, these include:

- Non-freeway facility types such as high-volume with limited access
- Similar or higher surface roughness setting as used in the NCHRP 25-55 study
- Build/No Build comparisons, application of the model is likely valid as the model is sensitive to changes in facility geometry and traffic volumes which are the features most likely to change in the build/no build scenario.
- NAAQS compliance test, all models performed reasonably well when emission rates are well known, current emission factor models are the best tools available to determine emission rates for input to the dispersion models thus only a warning is needed when applying the dispersion models for NAAQS compliance.

B.8 CURRENT LIMITATIONS WITH DISPERSION MODEL APPLICATION

Potential limitations with the AERMOD model with the RLINE option based (version 22112) include the following:

- Can only be applied using flat terrain
- Sound wall (i.e., barrier) option has not been approved for general use

B.8.1 Identification of Deficiencies in the Model and Possible Fixes

Deficiency: Based on the use of current EPA guidance the AERMOD-RLINE model is not as accurate as AERMOD AREA for high volume freeway applications under a variety of meteorological conditions.

Recommendation: Model comparison shows better performance for AERMOD when applied using AREAPOLY source type. It is likely a result of the current EPA guidance on estimating release height and initial vertical dimension as was developed prior to the incorporation of RLINE into the AERMOD model. The values for the release height and initial vertical dispersion were same for each source as described in section on Dispersion Model Testing. Research is needed to determine the appropriate release height for both cars and trucks by measuring the vertical concentration profile under a variety of meteorological conditions. This should lead to an improved characterization of the release height and initial vertical dimension.

For the tracer gas analysis, AERMOD-RLINE performance compared less favorably than the AERMOD-AREAPOLY approach for high to moderate volume freeway settings. At this time, we recommend using AERMOD-AREAPOLY for high to moderate volume freeway settings and not to use AERMOD-RLINE until additional research is carried out on the parameterization for the appropriate release height for cars and trucks or new guidance is issued.

B.9 CONCLUSIONS

Two version of the AERMOD dispersion model were evaluated (AERMOD-AREAPOLY and AERMOD-RLINE) using field data collected during the NCHRP 25-55 project. The models were compared with both tracer gas and using regulatory guidance methodology for CO and PM2.5. The tracer evaluation showed that both models did a good job pairing observations with predictions (r-squared = 0.82), however AERMOD-RLINE bias was high with a fractional bias (FB) greater than a factor of two at 0.72, while AERMOD-AREAPOLY was well within a factor of two at FB=0.24. Both models tend to have better performance for near field receptors as compared to far-field receptors.

The modeled design values for AERMOD-AREAPOLY were slightly below the observed DVs. While for AERMOD-RLINE the DV's were slightly higher than observed DV's. The largest difference was for the 24-hour PM2.5 at 1.0 $\mu g/m^3$ lower than observed for AERMOD-AREAPOLY, while AERMOD-RLINE was 1.0 $\mu g/m^3$ higher than observed. This would suggest that AERMOD-RLINE configuration is the more conservative model in the case of the design value criteria.

The results show that the AERMOD-AREAPOLY and AERMOD-RLINE dispersion models all performed much better against tracer data than did the regulatory modeling chain did against near-road monitoring data for PM2.5 and CO. For example, the r-squared for the tracer experiments for AERMOD-AREAPOLY and AERMOD-RLINE was 0.82, while for PM2.5 and CO paired in time and space the r-squared ranged from just 0.07 to 0.08. Similarly, the AERMOD-AREAPOLY and AERMOD-RLINE for the PM2.5 and CO paired in time and space had similar low correlations clearly identifying that the source of the uncertainty lies outside the dispersion component of the modeling chain.

The other sources of input to the modeling chain are traffic volume, background concentration and vehicle emission factor. The traffic data and background concentrations were considered well-characterized and not a likely source of significant error in the regulatory modeling chain. The most likely cause is the emission factor model and/or for PM2.5 EPA's AP-42 guidance for road dust.

The AERMOD-RLINE option in AERMOD does not place any additional burden on the user over the effort to develop inputs for AERMOD-AREAPOLY. The AERMOD-RLINE option is a relatively straightforward capability and is easy for users to add at the beginning of the project and does not present a barrier to its use. It should be noted however, the initial setup with AERMOD can be time consuming, especially if not done using commercial software, that greatly facilities developing inputs to the model.

Overall, evidence from the model evaluation shows that the dispersion component of the model chain has high accuracy when emission rates are well known. Further testing of the regulatory modeling chain is needed with additional measurement instrumentation to differentiate the sources of uncertainty and error within the modeling chain.

B.10 RECOMMENDATIONS FOR FUTURE RESEARCH

B.10.1 Lessons Learned

Based on our experience in performing the field experiment in this research we found the following issues to be of particular importance for having a successful experiment

- Careful attention needs to be made to meteorological conditions on experiment days to successfully measure downwind and upwind concentrations above detection limits.
- As tracer vehicle are seen as intermittent sources along the roadway it is critical that the tracer gas measurements be collected on a continuous basis over the entire sampling period.
- In such a complex near roadway environment, requiring moving tracer release vehicles, it is important to not only carefully design the tracer program but that a pilot study be conducted to assure that equipment operates as intended, tracer gas release amounts are sufficient, identify if other sources of tracer gas may be present that will interfere with the experiment, and assure collection equipment is free from cross-contamination.
- Under steady-state meteorological and traffic conditions we found that a single tracer release vehicle can provide sufficient concentration measurements for evaluating air quality dispersion models if the experiment contains a sufficient number of vehicles passes by the receptor network. In this study we found that a pass by frequency of at least once every 7-minutes was sufficient for capturing a representative one-hour average concentrations. Statistical evaluations of model performance were consistent between the use of one, two or three tracer release vehicles.

B.10.2 Dispersion Model Improvement Studies

We identified the need for further assessment for the plume height and initial vertical dimension from moving vehicles most notably under stable conditions. The current approach as recommended by EPA is based on a study by Gilles et. al (2005) which reported that the initial vehicle height is scaled by a factor of 1.7 to determine release height. However, this was based on a small fleet of nine vehicles with no vehicle higher than 3.3 meters. We suspect that for taller vehicles this 1.7 factor may be an overestimate. Additionally, as noted by Gilles that in addition to atmospheric stability, the shape of the vehicle, and the angle of the wind direction with respect to the direction of vehicle travel play a role in determining this scaling height factor. Further study is needed either through literature review or field study to determine an improved estimate for this scaling factor. This could require additional parameterization within the dispersion model if these additional factors play an important role in the scaling factor. We recommend that additional research first be carried out in a field study to better understand the plume rise heights from moving vehicles and then, depending upon the results from those findings, incorporate the appropriate parameterization into the dispersion model. We estimate the cost for a field program to carry out this study cost between $300K and $400K depending upon the number of experiments conducted, the number of vehicles used in each experiment, and the instrumentation used to determine the plume height.

Current EPA plans is to make the AERMOD-RLINE the default option when modeling roadways. Priorities for 2023 are to finish testing and evaluation of the model to include terrain and solid barriers (i.e., sound walls). This includes resolving issues with edge effects from barriers. The details for what will be included in promulgation to Appendix W will be presented at the 13th EPA Modeling Conference to be held in the Fall of 2023 with promulgation of those actions in the Summer of 2024 after considering public comments. We recommend that the June 27th, 2022, public release of AERMOD-RLINE (22112) be evaluated as to it's impact on the 25-55 findings as that version of the AERMOD-RLINE contains extensive number of bug fixes as well as several enhancements including update to the meander algorithm consistent with other source types and introduction of a minimum wind speed near surface. We recommend that further model evaluation studies be conducted that examine additional facility types where terrain and potentially barriers are of concern. In the near-term we recommend that terrain effects be tested and evaluated as this is the most limiting feature in the current AERMOD-RLINE configuration. We estimate the cost for a field program to carry out this study using both tracer and regulatory modeling chain evaluation as between $700K and $900K depending upon the number of experiments conducted, number of vehicles used, pollutants monitored, and number of facilities evaluated. For mid- to long-term needs evaluation studies should include the more complicated layouts of freeway interchanges and intersections.

B.10.2.1 Field Studies for Model Evaluations or Improvements

Evidence from the model evaluation using the NCHRP 25-55 study shows that the dispersion component of the model chain has high accuracy when emission rates are well known. Further testing of the regulatory modeling chain (not including tracer) is needed to validate these findings where emphasis on data collection would include a longer period of air quality monitoring both on-site ambient background and downwind concentration in the near road environment for PM2.5 and CO along with on-site vehicle count, speed and fleet mix information. The effort to conduct this research study would likely range between $300K and $450K depending on the duration of the study, the current level of monitoring (air quality, traffic volume, meteorology) and the subsequent data analysis, model evaluation and reporting requirements.

In addition, to better identify the emission rates, we recommend that a field program be carried out to evaluate the emission rates from emission factor models using real-time instrumentation as part of a tunnel study. Tunnel experiments have well defined volumes of air exchanged so that measured differences in PM2.5 and CO concentration between inlet and outlet locations within the tunnel can be definitively determined for comparison with the emission factor models. Ideally, this would be carried out at the same time as the above study and would ideally be conducted in close proximity. The additional effort to conduct this research study would likely range between $150K and $250K depending on the duration of the study, the logistics of a tunnel measurement setup, the pollutants measured, and the subsequent data analysis and model evaluation.

The NCHRP 25-55 study showed that for today's fleet of vehicles exhaust emissions of PM2.5 are just one component of PM2.5. Brake and tire wear are equally important and at times break wear may be more than twice exhaust PM2.5 emissions. The importance of brake and tire wear emissions will only continue with:

- the introduction of light-duty electric vehicles with no exhaust emissions
- autonomous vehicles which enable higher traffic volumes resulting in increased tire wear emissions
- light-duty electric vehicles which are heavier than light-duty gasoline vehicles leading to increased tire wear

In addition to brake and tire wear, re-entrained road dust will increase with the heavier light-duty battery electric vehicles as well as autonomous vehicles increasing the road carrying capacity. We recommend that a carefully designed field program be developed that will measure all of these important components of PM2.5 as well as PM10 emissions in the near roadway environment. Measurements in real world road conditions are needed that can separate tire wear emissions independently from brake wear and re-suspended road dust.

This is an area of active research and initial efforts would best be carried out with an extensive literature review which would include identifying such information as:

- Latest techniques for measurement methods – in field and in laboratory
- Identification of key variables in tire wear emissions
    - Impact of road pavement conditions on tire emissions
    - Vehicle weight
    - Tire material/composition
- Key variables in brake wear emissions
- Re-suspended road dust
    - Dependent on traffic volume

Depending on findings from the literature review a field study could be outlined and developed to improve the reliability of the emission estimates from tire, brake wear and re-suspended road dust emissions. This is particularly an important issue for re-entrained road dust from high volume roadways. The estimated cost to conduct the literature review and design a field study would be in the range between $100K-$150K depending on the emission types assessed and the level of detail outlined for the field program.

The NCHRP 25-55 study also showed the need for additional model evaluation studies for additional transportation facility types (e.g., high volume interchange and intersections) given the limited model validation studies completed for these facility types from moving vehicles. We would anticipate these studies would be designed in an approach similar to the NCHRP 25-55 study. The effort to conduct this research study would likely range between $250K and $350K for each facility type depending on the duration of the study, the complexity of the interchange and/or intersection, the pollutants measured, and the subsequent data analysis, model evaluation and the reporting requirements.

B.11 REFERENCES

Gillies, J.A., V.Etyemezian, H.Kuhns, D.Nikolic, D.A. Gillette, *Effect of vehicle characteristics on unpaved road dust emissions*, (2005) Atmospheric Environment 39 (2005) 2341–2347.

USEPA, 2016. Technical Support Document (TSD) for Replacement of CALINE3 with AERMOD for Transportation Related Air Quality Analyses, OAQPS, Air Quality Assessment Division, Research Triangle Park, December, EPA-454/B-16-006.

Wayson, R and Voight, C. (2022) *AERMOD Source Types RLINE and RLINEXT Testing*, FHWA Office of Natural Environment and FHWA Virginia Division Office, March 2022, The report is available on the AECOM.com website (link) and will also be posted on the page for MOVES-Matrix modeling on the Georgia Institute of Technology (https://www.gatech.edu/) website as part of the set of three reports that were developed for FHWA addressing the AERMOD RLINE model.

Attachment C: Model Evaluation Databases for Regulatory Air Pollution Dispersion Models from Mobile Sources

To effectively evaluate an air dispersion model for impacts of vehicle traffic and emissions in the near road environment a tracer gas study must be performed using a tracer gas released in a manner to simulate vehicle exhaust with a precisely known emission release rate from a moving mobile source to capture the wake and turbulent mixing effects within the roadway environment. In addition, site-specific meteorological measurements are needed to fully characterize the dispersion of gas once released into the atmosphere. These requirements currently limit the number of tracer databases suitable for dispersion model evaluation.

In Phase 1 of this research study, we identified historical tracer dataset that could potentially be used to assess air dispersion models. We reviewed the historical tracer datasets identified in Phase 1 and the requirements just discussed needed for a model evaluation study. Table C-1 summarizes those historical tracer datasets that meet the information requirements along with the NCHRP 25-55 tracer study (the "I-80 Freeway Experiment") and the current Caltrans tracer study.

This dataset can serve as a database for model development testing and evaluation. This dataset, (with the exception of the Caltrans – Riverside not yet completed) will be delivered to NCHRP as part of the final deliverables for the project.

Table C-1. Existing and new Tracer Dataset Studies for Model Comparison and Evaluation

Study Name and Location	Facility Type	Comment
Berkeley Freeway Experiment, I-80 at Aquatic Park, Berkeley, CA Tracer data	High-volume freeway	Mobile platform release of 3 unique perfluorocarbon tracers (10-lane ~12,700 VPH; 280,400 AADT)
GM Sulfate Experiment, Milford, MI	Low-volume freeway	SF6 tracer release from a fleet of moving vehicles (4-lane) at near steady state speed of 50 mph and traffic volume (~ 5,550 VPH).
Berkeley Freeway Experiment, I-80 at Aquatic Park, Berkeley, CA Near Road air quality monitoring data	High-volume freeway	Hourly ambient air quality measurements of CO, $PM_{2.5}$, and black carbon both upwind and downwind. high volume freeway (10-lane ~12,700 VPH; 280,400 AADT).
Caltrans Highway 99 Tracer Study, Sacramento, CA	Low-volume freeway	SF6 tracer release, rural location along US Hwy 99, 4-lane divided highway, 14-m median, 35,000 AADT, receptor in median, closest receptor outside mixing zone 50-m from highway; used in a number of past model evaluation studies.

Study Name and Location	Facility Type	Comment
UCR – Caltrans, Riverside, CA (Full study to be available 2023) preliminary research findings from the 2019 tracer study can be found at https://link.springer.com/article/10.1007/s11869-021-01104-9	Freeways – with and without barriers	SF6 tracer release from eight vehicles also measure CO, CO_2, and some limited black carbon sampling in post 2019 experiments