[DRAFT] MEMORANDUM

SUBJECT: Guidance on Significant Impact Levels for Ozone and Fine Particles in the Prevention of Significant Deterioration Permitting Program

FROM: Stephen D. Page, Director
Office of Air Quality Planning and Standards

TO: Regional Air Division Directors, 1-10

The purpose of this memorandum is to provide guidance on compliance demonstration tools for use with ozone and fine particles ($PM_{2.5}$) in the Prevention of Significant Deterioration (PSD) permitting program. The Environmental Protection Agency has developed a new analytical approach and has used it to identify a significant impact level (SIL) for each ozone and $PM_{2.5}$ National Ambient Air Quality Standard (NAAQS) and the $PM_{2.5}$ PSD increments. We recommend that permitting authorities[1] consider using these values to help determine whether a proposed PSD source[2] causes or contributes to a violation of the corresponding NAAQS or PSD increments. The supporting technical document[3] provides a detailed discussion of the technical analysis used to develop these values. The supporting legal memorandum provides further detail on a legal basis that permitting authorities may choose to adopt to support using SILs to show that requirements for obtaining a PSD permit are satisfied.[4] This memorandum provides the results of the technical analysis and information on the particular points in the PSD air quality analysis at which permitting authorities may decide to use these values on a case-by-case basis in

[1] Permitting authorities include the EPA, state, local and tribal permitting authorities.
[2] As used in this memorandum, "PSD source" means a construction or modification of a major stationary source triggering PSD permitting requirements.
[3] "Technical Basis for the EPA's Development of Significant Impact Thresholds for $PM_{2.5}$ and Ozone"; EPA/XXX-X-XX-XXX, [DATE]
[4] Legal Support Memorandum: Application of Significant Impact Levels in the Air Quality Demonstration for Prevention of Significant Deterioration Permitting under the Clean Air Act," [DATE]

the review of PSD permit applications.[5] This memorandum and the supporting documents are not final agency actions and do not create any binding requirements on permitting authorities, permit applicants or the public.

I. INTRODUCTION

When a PSD permit applicant has shown through air quality modeling that the projected impact from a proposed source is less than a SIL value for a particular pollutant, the EPA believes there is a valid analytical and legal basis for the permitting authority to conclude that this showing is sufficient to demonstrate that the proposed source will not cause or contribute to a violation of a NAAQS or PSD increment for that pollutant. Permitting authorities may elect to use the SILs discussed below, and the EPA has provided policy, technical and legal analyses that permitting authorities may choose to adopt or adapt in supporting their use of the SILs in particular PSD permitting actions. The use of SILs can help satisfy PSD requirements while conserving resources for applicants and permitting authorities.

The EPA has previously issued guidance describing particular uses of SILs.[6,7,8,9] Permitting authorities have long had the discretion to apply SILs on a case-by-case basis in the review of individual permit applications, provided such use was justified in the permitting record.[10] In an effort to reduce the need for case-by-case justification by permitting authorities, in 2010, the EPA finalized a rule to codify particular $PM_{2.5}$ SIL values and specific applications of those values,[11] but in subsequent litigation the EPA found an inconsistency between the preamble and

---

[5] The term "case-by-case basis" is used in this memorandum to refer to a permitting authority's use of a SIL value in a particular air quality analysis in an individual PSD permitting action when the SIL value has not been adopted in the state's EPA-approved PSD SIP rules (or the federal PSD rules, as applicable) pursuant to section 165(a)(3) of the Clean Air Act (CAA). When the SIL value has not been adopted into the applicable PSD rules, the permitting authority's record of each PSD permitting action in which a SIL is used must contain a justification demonstrating that the particular level and use of the SIL value is consistent with the CAA and applicable PSD rules. The permitting authority's justification may make use of the policy, legal and technical analysis documents developed by the EPA. We note that in a broader sense, all PSD permit reviews are "case-by-case" under section 165(a) of the CAA; in this memorandum, for clarity we refer to the case-specific nature of PSD permit reviews as "permit-specific" when not discussing the use of a SIL value by a permitting authority on a case-by-case basis.

[6] Memorandum from Stephen D. Page, EPA OAQPS, to EPA Regional Air Division Directors, "Guidance Concerning the Implementation of the 1-hour SO_2 NAAQS for the Prevention of Significant Deterioration Program," August 23, 2010.

[7] Memorandum from Stephen D. Page, EPA OAQPS, to EPA Regional Air Division Directors, "Guidance Concerning the Implementation of the 1-hour NO_2 NAAQS for the Prevention of Significant Deterioration Program," June 29, 2010.

[8] Memorandum from Stephen D. Page, EPA OAQPS, to OAQPS Personnel and EPA Regional Modelers, "Modeling Procedures for Demonstrating Compliance with $PM_{2.5}$ NAAQS," March 23, 2010.

[9] Memorandum from Gerald A. Emison, EPA OAQPS, to Thomas J. Maslany, EPA Air Management Division EPA Region 3, "Air Quality Analysis for Prevention of Significant Deterioration (PSD)," July 5, 1988.

[10] Order Responding to Petitioner's Request that the Administrator Object to Issuance of a State Operating Permit, In the Matter of CF&I Steel, L.P. dba EVRAZ Rocky Mountain Steel, Petition Number VIII-2011-01, at 15-17 (May 31, 2012) ("Rocky Mountain Steel Order"); In re: Mississippi Lime Company, 15 E.A.D. 349, 375-379 (EAB 2011).

[11] 75 FR 84864 (October 20, 2010).

regulatory text, and the court granted the EPA's request to vacate and remand the inconsistent regulatory text.[12]

Following the litigation, the EPA initially began developing a new rule to address the inconsistencies identified in the 2010 rulemaking.[13] However, after further evaluation and the identification of a revised set of SIL values based on the technical and legal analyses described below, the EPA believes it should first obtain experience with the application of these values in the permitting program before establishing a generally applicable rule.[14] In addition, permit applicants and permitting authorities have communicated a need for the EPA to develop SIL values for ozone on an expedited basis. As a result, the EPA intends at this point to take a two-step approach.

First, the EPA is providing non-binding guidance so that we may gain valuable experience and information as permitting authorities use their discretion to apply and justify the application of the SIL values identified below on a case-by-case basis in the context of individual permitting decisions. We will be seeking to learn generally about permitting agencies' experiences in applying SILs in particular PSD permitting decisions. We will also be seeking more specific information, including how often and in what types of settings the application of a SIL at the single-source assessment and cumulative assessment stages of the PSD air quality analysis has made a critical difference in whether a conclusion was reached that the proposed source will not cause or contribute to a NAAQS or PSD increment violation. The EPA intends to obtain this information through its own PSD permitting activities in states that do not have SIP-approved PSD programs, regular discussions between our regional offices and air agencies, regular conference calls with the permitting committees of national organizations of air agencies, and technical conferences of air quality modelers and others interested in permitting activities.

Second, the EPA will use this experience and information to assess, refine and, as appropriate, codify SIL values and specific applications of those values in a future, potentially binding rulemaking. [15] During this second step, to assess whether it is appropriate to codify the particular SIL values derived using EPA's technical methodology or to codify revised values, the EPA will consider what SIL values are suitable in all locations and circumstances to show that an increase in air quality concentration below the corresponding SIL value does not cause or contribute to a violation of the NAAQS or PSD increments. Until the EPA conducts a rulemaking, permitting authorities retain discretion to use or not to use the EPA-derived SILs in particular PSD

---

[12] *Sierra Club v. EPA*, 705 F.3d 458 (D.C. Cir. 2013). In its litigation brief at n. 10, the EPA stated an intent to issue guidance in the near future concerning $PM_{2.5}$ SIL values remaining in 40 CFR 51.165(b). The EPA issued such guidance in May 2014. Memorandum from Stephen D. Page, EPA OAQPS, to EPA Regional Air Division Directors, "Guidance for $PM_{2.5}$ Permit Modeling," May 20, 2014.

[13] Fall 2015 Regulatory Agenda, USEPA, 80 FR 78024, December 15, 2015. Ozone and Fine Particulate Matter ($PM_{2.5}$) Significant Impact Levels (SILs) for Prevention of Significant Deterioration (PSD), RIN: 2060-AR28. *http://www.reginfo.gov/public/do/eAgendaViewRule?pubId=201510&RIN=2060-AR28.*

[14] *See SEC v. Chenery Corp.*, 332 U.S. 194, 199-203 (1947) (recognizing that some principles may warrant further development before they are ready to be codified in a rule of general applicability).

[15] The EPA does not at present have a schedule for a future rulemaking on ozone and $PM_{2.5}$ SILs, but we will review the status from time to time. This rulemaking will continue to appear in the EPA's regulatory agendas under longer-term actions until we develop a specific schedule.

permitting actions. If a permitting authority chooses to use these or other SIL values on a case-by-case basis, it must justify the values and their use in the administrative record for the permitting action.

Since the 2010 rulemaking, the EPA has examined the legal basis for using SIL values in PSD air quality impact analyses. In addition, the EPA has sought to develop an improved technical methodology for deriving SIL values. This memorandum and supporting documents are the products of this effort. They identify specific SIL values for ozone and $PM_{2.5}$ and provide a supporting justification that permitting authorities may choose to apply on a case-by-case basis. The values and supporting justification are designed so that permitting authorities can choose to apply the SIL values at any location to demonstrate that a proposed source does not cause or contribute to a violation of air quality standards. In contrast to the 2010 rulemaking, we have developed separate SIL values for the $PM_{2.5}$ NAAQS and PSD increments, and we have developed SILs for the ozone NAAQS. Since there are no PSD increments for ozone, the EPA has not developed SILs for ozone.

The EPA believes that the application of these SILs in the manner described below would be sufficient in most situations for a permitting authority to conclude that a proposed source will not cause or contribute to a violation of an ozone or $PM_{2.5}$ NAAQS or $PM_{2.5}$ PSD increment. However, this guidance is not a final agency action and does not reflect a final determination by the EPA that any particular proposed source, or class of proposed sources, does not cause or contribute to a violation or may obtain a PSD permit. A determination that a proposed source does not cause or contribute to a violation can only be made by a permitting authority on a permit-specific basis after consideration of the permit record. This guidance is not legally binding and does not affect the rights or obligations of permit applicants, permitting authorities, or others. The SIL values identified by the EPA have no practical effect unless and until permitting authorities decide to use those values in particular permitting actions. The experience of permitting authorities in using these SILs on a case-by-case basis, or in choosing to limit or forego their use in specific situations, will be valuable information for the EPA to consider in a future rulemaking. Permitting authorities retain the discretion to apply and justify different approaches and to require additional information from the permit applicant to make the required air quality impact demonstration, consistent with the relevant PSD permitting requirements.

II. BACKGROUND

A PSD permit applicant must demonstrate that "emissions from construction or operation of such facility will not cause, or contribute to, air pollution in excess of any" NAAQS or PSD increment.[16] The EPA has reflected this requirement in its PSD regulations.[17] The CAA does not specify how a permit applicant or permitting authority is to make this demonstration, but section 165(e) authorizes the EPA to determine how the analysis is to be conducted, including the use of air quality models. In accordance with this authority, the EPA has promulgated regulations that

[16] Section 165(a)(3) of the CAA. The EPA interprets the phrase "in excess of" to mean a violation, not the exceedance described in 40 CFR 50.1(l).
[17] 40 CFR 51.166(k); 40 CFR 52.21(k).

identify such models and the conditions under which they may be used in the PSD program to make the demonstration required under the Act.[18]

Using the models identified in EPA regulations, there are two basic ways that a PSD permit applicant can demonstrate that the proposed source's emissions will not cause or contribute to a violation of any NAAQS or PSD increment. One way is to demonstrate that no such violation is occurring or projected to occur in the area affected by the emissions from the proposed source.[19] A second way is to demonstrate that the emissions from the proposed source do not cause or contribute to any identified violation of the NAAQS or PSD increments.[20]

The Act does not define "cause" or "contribute." Reading these terms in context, the EPA has historically interpreted this provision in section 165(a)(3) of the CAA and associated regulations to mean that a source must have a "significant impact" on ambient air quality in order to cause or contribute to a violation.[21] Thus, the EPA and other permitting authorities have concluded that a proposed source may meet the requirements in CAA section 165(a)(3) and the EPA's PSD regulations by showing that its projected impact on air quality at the site of a modeled violation is below a level of air quality impact considered to be significant.[22]

Historic Use of SILs

In the context of section 165(a)(3) of the CAA, the EPA has historically used pollutant-specific concentration levels known as "significant impact levels" to identify the degree of air quality impact that "causes, or contributes to" a violation of a NAAQS or PSD increment.[23] Consistent with EPA guidance, proposed sources have met the requirement to demonstrate that they do not cause or contribute to a violation by showing that the ambient air quality impacts resulting from the proposed source's emissions would be below these concentration levels.[24] The SIL values have served as a compliance demonstration tool to make the required demonstration in the PSD program. They have helped to reduce the burden on permitting authorities and permit applicants to conduct often time-consuming and resource-intensive air dispersion modeling where such modeling was unnecessary to demonstrate that a permit applicant meets the requirements of section 165(a)(3), consistent with the procedures set forth originally in 1977 in the "Guidelines for Air Quality Maintenance Planning and Analysis, Vol 10 (Revised) and Procedures for Evaluating Air Quality Impact of New Stationary Sources."[25]

Recent Status of SILs for Ozone and $PM_{2.5}$

[18] 40 CFR 51.166(l); 40 CFR 52.21(l); 40 CFR part 51, Appendix W (Guideline on Air Quality Models).
[19] 1990 Draft NSR Workshop Manual at C.51.
[20] 40 CFR part 51, App. W, § 10.2.3.2(a); 1990 Draft NSR Workshop Manual at C.52.
[21] In re Prairie State Generating Co., 13 E.A.D. 1, 105 (EAB 2006). This EAB opinion includes a long discussion of EPA's prior guidance with other examples.
[22] 1990 Draft NSR Workshop Manual at C.52.
[23] 61 FR 38250, 38293 (July 23, 1996); 72 Fed. Reg. 54112, at 54139 (September 21, 2007).
[24] 1990 Draft NSR Workshop Manual at C.51-C.52.
[25] October 1977, U.S. EPA, Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711. Specific applications of how SILs have been used in the PSD program are discussed later in this memorandum.

Stakeholders have long sought compliance demonstration tools for ozone and secondarily-formed $PM_{2.5}$. In July 2010, Sierra Club petitioned the EPA to designate computer models to use in determining if major proposed sources of air pollution cause or contribute to violations of the ozone or $PM_{2.5}$ NAAQS. In January 2012, the EPA granted the petition and committed to engage in rulemaking to evaluate whether updates to Appendix W are warranted and, as appropriate, incorporate new analytical techniques or models for ozone and secondarily-formed $PM_{2.5}$. In granting the petition, the EPA explained that the "complex chemistry of ozone and secondary formation of $PM_{2.5}$ are well-documented and have historically presented significant challenges to the designation of particular models for assessing the impacts of individual stationary sources on the formation of these air pollutants."[26] Because of these considerations, the EPA's past judgment had been that it was not technically sound to designate with particularity specific models that must be used to assess the impacts of a single source on ozone and secondarily-formed $PM_{2.5}$ concentrations. Instead, the EPA established a consultation process with permitting authorities for determining (on a permit-specific basis) the analytical techniques that should be used for single-source analyses for both ozone and secondarily-formed $PM_{2.5}$.

The EPA has responded to the Sierra Club petition by proposing revisions to Appendix W.[27] As discussed in the Appendix W proposed language, recent technical advances have made it reasonable for the EPA to provide more specific guidelines that identify appropriate analytical techniques or models that may be used in compliance demonstrations for the ozone and $PM_{2.5}$ NAAQS. The EPA expects that the final Appendix W revisions will include criteria and process steps for choosing single-source analytical techniques or models to estimate ozone impacts from precursor nitrogen oxide and volatile organic compound emissions. The ozone SIL value recommended in this guidance is intended to complement the Appendix W updates by providing a threshold that may be used to determine whether an impact predicted by the chosen technique or model causes or contributes to a violation. With respect to $PM_{2.5}$, the EPA expects the final Appendix W revisions will include criteria and process steps for choosing single-source analytical techniques or models to assess concentrations of direct and secondarily-formed $PM_{2.5}$.

In the 2010 $PM_{2.5}$ SILs rule, the EPA established SIL values for $PM_{2.5}$ in paragraph (k)(2) of 40 CFR 51.166 and 52.21 of the PSD regulations. In January 2013, the U.S. Court of Appeals for the District of Columbia Circuit granted the EPA's request to vacate and remand the paragraph (k)(2) provision in both PSD regulations so the EPA could correct them.[28] Paragraph (k)(2) as promulgated in 2010 included numerical values of $PM_{2.5}$ SILs and statements about their role in completing an air quality impact analysis with regard to the $PM_{2.5}$ NAAQS and PSD increments. Specifically, the 52.21(k)(2) rule text stated that if the impact of a proposed source seeking a federal PSD permit were below the relevant SIL value(s), then the proposed source would be deemed to not cause or contribute to a violation. The 51.166(k)(2) rule text stated that a state's PSD rules could contain a similar provision. The EPA asked the court to vacate and remand the (k)(2) paragraphs of both PSD regulations so that the EPA could correct an inconsistency

[26] Letter from Gina McCarthy, Assistant Administrator, EPA Office of Air and Radiation, to Robert Ukeiley, Sierra Club, January 4, 2012.
[27] 80 FR 45340 (July 29, 2015).
[28] *Sierra Club v. EPA*, 294 F.3d 155, 160 (D.C. Cir. 2002).

between (1) that rule text, which left no discretion for the permitting authority, and (2) our statements in the preamble to the 2010 $PM_{2.5}$ SILs rule, which identified circumstances where it may not be appropriate for a permitting authority to rely solely on the $PM_{2.5}$ SILs as a basis for concluding that a proposed source does not cause or contribute to a violation.[29]

The court left intact the $PM_{2.5}$ NAAQS SIL values contained in 40 CFR 51.165(b)(2), because the regulatory text therein did not say that a showing that a proposed source has an impact equal to or less than the SIL value is always deemed to not cause or contribute. The regulatory text contained at 51.165(b)(2) says that, at a minimum, an impact greater than the listed SIL must be considered significant, but does not compel the opposite conclusion for impacts equal to or below that value.[30]

III. RECOMMENDED SIL VALUES FOR USE IN AIR QUALITY IMPACT DEMONSTRATION REQUIRED TO OBTAIN A PSD PERMIT

As discussed above, the EPA has interpreted the phrase "cause, or contribute to" in section 165(a)(3) of the Act to mean that a proposed source is prevented from obtaining a permit if the proposed source will have a "significant impact" on air pollutant concentrations that violate the standards. In this context, the EPA believes permitting authorities may read the phrase "cause, or contribute to" to be inapplicable to an air quality impact that is insignificant. This interpretation is more fully explained in the legal support memorandum. In this context, the EPA believes an insignificant impact is an impact on air quality concentrations that is small and not meaningful. (The EPA has often described such an impact as "trivial" or "*de minimis*".)

The term "contribute," as used in the context of section 165(a)(3), is ambiguous. In the absence of specific language in section 165(a)(3) regarding the degree of contribution that is required (such as the term "significantly"), the EPA has the discretion under this provision to exercise its judgment to determine the degree of impact that "contributes" to adverse air quality conditions based on the particular context in which the term "contribute" is used. The EPA may also identify criteria or factors that may be used to determine whether something "contributes,"

---

[29] These preamble statements were the following: "[N]otwithstanding the existence of a SIL, permitting authorities should determine when it may be appropriate to conclude that even a *de minimis* impact will 'cause or contribute' to an air quality problem and to seek remedial action from the proposed new source or modification." *See* 75 FR 64864 at 64892. "[T]he use of a SIL may not be appropriate when a substantial portion of any NAAQS or increment is known to be consumed." *See* 75 FR 64864 at 64894. "[W]e earlier provided an example of when it might be appropriate to require a modified source to mitigate its contribution to a violation of a NAAQS or increment even when the predicted ambient impact of the proposed emissions increase would result in what is normally considered to be *de minimis*." *See* 75 FR 64864 at 64894.

[30] Section 165(b)(2) is phrased such that an impact equal to the listed value is treated the same as impacts below the listed value. This contrasts to the approach in (k)(2), and in this guidance, that an impact equal to the SIL is treated the same as impacts above the SIL.

including qualitative or quantitative criteria that are appropriate to the particular context.[31] For purposes of implementing section 165(a)(3) of the Act, the EPA has found it more expedient and practical to use a quantitative threshold (expressed as a level of change in air quality concentration) to determine whether increased emissions from proposed construction or modification of a source will contribute to air quality concentrations in violation of applicable standards. The EPA believes that the permitting process can be streamlined without compromising air quality, if the EPA and permitting authorities are able to identify a quantitative threshold or dividing line between an insignificant and significant impact on air pollutant concentration. Using a quantified threshold for this purpose is permissible as long as the EPA or the appropriate permitting authority provides a reasoned explanation for why impacts below that value do not constitute a contribution to a violation in this context.

To determine what is (and is not) a significant impact in the context of section 165(a)(3) of the Act, the EPA has generally supported using the values in 40 CFR 51.165(b).[32] The EPA has described these levels as "significance levels."[33] Section 51.165(b)(2) was originally promulgated by the EPA in 1987 as part of an offset program that permitting authorities could apply after it was determined that construction at a stationary source was predicted to cause or contribute to a violation of the NAAQS.[34] This regulation provides that a proposed source planning to locate in an attainment area will be considered to "cause or contribute to" a violation of the NAAQS if its impact would exceed specific values identified in the regulations. For example, section 51.165(b) states that a proposed source impact any larger than 5 micrograms per cubic meter ($\mu g/m^3$) for the 24-hour SO_2 NAAQS causes or contributes to a violation of that NAAQS. The section refers to these values as "significance levels." Values are not provided for every NAAQS, and in particular not for ozone (and until 2010 not for $PM_{2.5}$), but for those NAAQS covered in this regulation, the application is the same. Over time, these air quality concentration significance levels in section 51.165(b) have become known as "significant *impact*

[31] See *Catawba County, N.C. v. EPA*, 571 F.3d 20, 39 (D.C. Cir. 2009). In this case interpreting the term "contributes" in section 107(d) of the CAA, the court held that the EPA is not required to establish a quantitative or objective, bright-line test to define a contribution by sources to adverse air quality conditions in a nearby area in the context of designations with respect to attainment of a NAAQS. The court recognized that the EPA has the discretion to use a totality-of-the-circumstances test if the agency defines and explains the criteria that it is applying. While this opinion said that a quantified threshold is not required to define contribution in the context of section 107(d), the court's reasoning does not preclude PSD permitting authorities from choosing to use a quantitative level of impact to represent a contribution to a violation of the NAAQS or PSD increment when implementing section 165(a)(3) of the CAA.

[32] Emison Memo at footnote 5 references 40 CFR 51.165(b), which defines "significant," and the NSR Workshop Manual at C.26-C.28 lists values from 40 CFR 51.165(b)(2) for the purpose of defining the area of "significant ambient impact."

[33] The EPA initially promulgated these same concentration values in 1979 as the "significance levels" under which a source locating in the "clean" portion of a nonattainment area may be exempt from the preconstruction review requirements in Appendix S to Part 51, 44 FR 3274, 3283 (January 16, 1979).

[34] 52 FR 24672, 24713 (July 1, 1987).

levels"[35] [emphasis added] in order to distinguish them from the significant *emissions rates* reflected in the definition of the term "significant," which serve a different function in the PSD program.[36] The EPA has also issued guidance memoranda that have provided recommended SIL values for the 1-hour NO_2 and SO_2 NAAQS, to be used for the purpose of determining what are (and are not) significant impacts for these pollutants.[37] The EPA has also observed that permitting authorities have discretion to develop their own SIL values, provided that such values are properly supported in permitting authority actions or decisions in which the values are used to make the required showing.[38]

The EPA's basis for the values in section 51.165(b)(2) of its regulations has generally been a percentage of the applicable PSD increments for each pollutant. The EPA used a similar approach in 2010 to add $PM_{2.5}$ values to section 51.165(b)(2) and establish $PM_{2.5}$ values in sections 51.166(k)(2) and 52.21(k)(2). However, given limitations in the rationale supporting them, the EPA recognized in the preamble to the 2010 $PM_{2.5}$ SILs rule that a permitting authority may not be able to apply the SIL values derived through this approach in every situation to show that proposed construction does not cause or contribute to a violation of standards. The EPA acknowledged that "the use of a SIL may not be appropriate when a substantial portion of any NAAQS or increment is known to be consumed." The EPA also said that "notwithstanding the existence of a SIL, permitting authorities should determine when it may be appropriate to conclude that even a *de minimis* impact will 'cause or contribute' to an air quality problem and to seek remedial action from the proposed new source or modification."[39] To guard against the improper use of the 2010 SILs for $PM_{2.5}$ in such circumstances, the EPA later recommended that permitting authorities use those SILs only where they could establish that the difference between background concentrations in a particular area and the NAAQS was greater than those SIL values.[40] This approach was intended to guard against misuse of the SILs that were based on a percentage of the $PM_{2.5}$ PSD increments.

Since that $PM_{2.5}$ modeling guidance was issued, the EPA has developed a new technical method for determining a concentration level that can be considered an insignificant impact on air pollutant concentrations for ozone and $PM_{2.5}$ in the context of PSD permitting. This technical method, referred to as the air quality variability approach, is described in the supporting technical docuement. Given the improvements reflected in this method, the EPA does not see a need for permitting authorities to show that the difference between background concentrations and the

---

[35] The first reference to "significant impact levels" is in the 1980 NSR Workshop Manual, which the EPA subsequently updated in the 1990 draft. It is worth noting that the 1977 comments to the proposed Appendix W rule (45 FR 58543) addressed whether a single-source screening technique should be used to determine if a cumulative modeling analysis would be required in a preconstruction review; industry and state agency comments indicated both groups favored some use of a tool to alleviate resource burden.

[36] Section 52.21(b)(23) also uses the term "significance" and applies discrete values for determining if a proposed source is significant. This regulation states that significance is any net emissions increase equal to or exceeding 40 tons per year (TPY) for ozone, and, for direct emissions of $PM_{2.5}$, 10 TPY (40 TPY for SO_2 and 40 TPY NO_2 unless demonstrated not to be a $PM_{2.5}$ precursor).

[37] Page memoranda at footnotes 5 and 6.

[38] 77 FR 37038 (June 20, 2010); 14 E.A.D. 723 (EAB 2010).

[39] 75 FR 64864, 64892.

[40] Memorandum from Stephen D. Page, EPA OAQPS, to EPA Regional Air Division Directors, "Guidance for $PM_{2.5}$ Permit Modeling," May 20, 2014.

relevant NAAQS is greater than the SIL value before applying one of the recommended $PM_{2.5}$ SIL values, as previously stated. The EPA's intention with this new method is to derive SIL values that are more universally applicable to a range of conditions, including those where a substantial portion of the NAAQS or PSD increment is known to be consumed. The EPA does not consider its qualifying statements from the preamble of the 2010 rule (quoted in the prior paragraph) to be applicable to the $PM_{2.5}$ SIL values derived with this new method; however, permitting authorities retain discretion to decide to apply or not to apply SILs as a general matter, or in particular permitting actions based on information in the administrative record.

In order for a concentration level to be used to show that the air quality impact of a proposed source does not cause or contribute to a violation of the NAAQS or PSD increment, the concentration value must represent a level of impact on ambient air quality that is insignificant or not meaningful. An insignificant impact on air pollutant concentrations can be identified and quantified based on an assessment of the variability of air quality, using data from the U.S. ambient $PM_{2.5}$ and ozone monitoring network. Due to fluctuating meteorological conditions and changes in day-to-day source operations, there is an inherent variability in the air quality in the area of a monitoring site. This variability can be characterized through the application of a well-established statistical framework for quantifying uncertainty in population statistics. The analysis described in the supporting technical document quantifies the fluctuations in pollutant concentrations (as measured by design values) and, for each NAAQS, determines a value for a concentration difference that is meaningful in the context of inherent variability. Changes of less than this magnitude may be considered to be in the "noise" of observed design values. This technical analysis provides a basis for a permitting authority to conclude that concentration increases below this SIL do not cause or contribute to violations of the relevant NAAQS or PSD increments.

SILs for NAAQS

Using this air quality variability approach, the EPA derived SIL values for the 8-hour ozone NAAQS and each $PM_{2.5}$ NAAQS averaging period, which are applicable to attainment and unclassifiable areas. The SIL values for the NAAQS are listed in Table 1. Each SIL value is based on the level, averaging period and statistical form of its corresponding NAAQS. For example, for ozone the recommended SIL value is based on the 4^{th} highest daily maximum 8-hour concentration, averaged over 3 years. The derived value from the air quality variability analysis is 1.0 parts per billion (ppb), and we recommend the case-by-case application of this value as the SIL for the 8-hour ozone NAAQS.

For the 24-hour $PM_{2.5}$ NAAQS, the SIL value we recommend is 1.2 $\mu g/m^3$. The derived value from the air quality variability analysis is 1.3 $\mu g/m^3$ and is based on an analysis of the 98^{th} percentile 24-hour concentrations averaged over 3 years; however, 40 CFR 51.165(b)(2) still lists 1.2 $\mu g/m^3$ as the SIL value for the 24-hour $PM_{2.5}$ NAAQS, and, pending further evaluation by the EPA, we recommend it for maintaining consistency with the rule. In the 2010 $PM_{2.5}$ SILs rulemaking, the EPA determined that an impact above this value will be considered to cause or contribute to a violation of the 24-hour $PM_{2.5}$ NAAQS at any location that does not meet this standard. In the same rule, the EPA also sought to establish that an impact below this value would not cause or contribute to a violation of this NAAQS but acknowledged that there could be circumstances where this conclusion was not always valid. Even though the ambient air quality variability approach indicates that an impact below 1.3 µg/m^3 is not significant, 51.165(b)(2) remains in the EPA's regulations and the agency is presently bound by its prior conclusion (that an impact above 1.2 µg/m^3 is significant and will cause or contribute to a violation of the 24-hour $PM_{2.5}$ NAAQS). Thus, the EPA cannot conclude at this time that an impact between 1.2 µg/m^3 and 1.3 µg/m^3 is an insignificant impact or an impact that will not cause or contribute to a violation of the NAAQS. However, based on the ambient air quality variability approach, the EPA is able to conclude that impacts below 1.2 µg/m^3 are insignificant at any location and will not cause or contribute to a violation of the NAAQS.[41] The case-by-case use of this recommended SIL value should be justified in the record for each permit.

For the annual $PM_{2.5}$ NAAQS, we recommend 0.2 µg/m^3 as the SIL value, which is the value derived from the air quality variability analysis and is based on a 3-year average of annual average concentrations. The case-by-case use of this recommended SIL value should be justified in the record for each permit. This value is lower than the value of 0.3 µg/m^3 listed in 51.165(b)(2). Since section 51.165(b)(2) does not address whether an impact below 0.3 µg/m^3 causes or contributes to a violation of the NAAQS, permitting authorities retain the discretion under this provision to determine on a case-by-case basis whether an impact between 0.2 µg/m^3 and 0.3 µg/m^3 will cause or contribute to a violation of the annual $PM_{2.5}$ NAAQS. Based on the ambient air quality variability approach, the EPA's judgment is that an impact below 0.2 µg/m^3 is insignificant and should be considered to not cause or contribute to any violation of the annual $PM_{2.5}$ NAAQS that is identified.

Table 1. Recommended SIL Values for Ozone and $PM_{2.5}$ NAAQS

Criteria Pollutant (NAAQS level)	NAAQS SIL concentration
Ozone 8-hour (70 ppb)	1.0 ppb
$PM_{2.5}$ 24-hour (35 µg/m^3)	1.2 µg/m^3*
$PM_{2.5}$ annual (12 µg/m^3 or 15 µg/m^3)	0.2 µg/m^3

* The table takes into account the SIL value for the 24-hour $PM_{2.5}$ NAAQS that is in section 51.165(b)(2). Refer to the guidance discussion for details.

[41] 40 CFR 165(b)(2) provides that a source impact higher than one of the listed significance levels is to be considered significant. A source impact exactly equal to a significance level need not be considered significant. In contrast, in this memorandum, consistent with past guidance, we are recommending that a value exactly equal to a recommended SIL be considered significant. Thus, these two approaches treat a value equal to the stated level differently. In practice, we do not expect this to be a practical difference because it will be very unusual for a source's impact to exactly equal one of the recommended SIL values.

We recommend that these SIL values apply everywhere, regardless of the class of the airshed.[42] For $PM_{2.5}$, this recommendation is different than what was provided in the vacated (k)(2) paragraphs, where the SIL value that would be used for NAAQS purposes was different for Class I areas than for Class II and III areas. The EPA recognizes that, historically, Congress has provided special protections to Class I areas, via PSD increments. The EPA believes that because each ozone and $PM_{2.5}$ NAAQS is uniform throughout the class areas, no class-specific protection via SILs is necessary when assessing whether a source causes or contributes to a violation of the NAAQS.

SILs for PSD Increment

There are no PSD increments established for ozone, and, thus, no ozone SIL values are needed for PSD increment compliance purposes. We used the air quality variability approach to develop increment SILs for the $PM_{2.5}$ PSD increments (Table 2), but in an indirect way. The SIL values for the $PM_{2.5}$ PSD increments are derived from the NAAQS SIL values and reflect that, under the PSD regulations, the allowable PSD increment values are different for Class I, II and III areas. For Class II areas (which comprise most of the U.S.) and Class III areas (of which there are currently none), we recommend that the values of the NAAQS SILs also be used for PSD increment SILs. For Class I areas, we are recommending annual and 24-hour PSD increment SIL values that are lower than the NAAQS SIL values. The EPA recognizes that Class I areas have historically been provided special protection.[43] To achieve this additional protection, we applied the ratios of the Class I and Class II allowable PSD increments to the NAAQS SIL values derived in our technical analysis.[44] The EPA believes these values for Class I areas will continue to reflect this higher level of protection through the PSD increment SILs.

Table 2. Recommended SIL Values for $PM_{2.5}$ Increment

Criteria Pollutant (averaging period)	PSD increment SIL concentration		
	Class I	Class II	Class III
$PM_{2.5}$ (24-hour)	0.27 μg/m^3	1.2 μg/m^3	1.2 μg/m^3
$PM_{2.5}$ (annual)	0.05 μg/m^3	0.2 μg/m^3	0.2 μg/m^3

IV. APPLICATION OF SILS

[42] When Congress established the PSD program requirements under the 1977 CAA Amendments, it included specific numerical increment levels for SO_2 and particulate matter (expressed at that time as “total suspended particulate”) for Class I, II and III areas. Congress designated Class I areas (including certain national parks and wilderness areas) as areas of special national concern, where the need to prevent deterioration of air quality is the greatest. Consequently, the PSD increments are the smallest in Class I areas. The increments of Class II areas are larger than those of Class I areas and allow for a moderate degree of emissions growth. Class III areas have the largest increments, but to date no Class III areas have been designated. The EPA subsequently defined Class I, II and III increments for NO_2 and PM_{10}, and $PM_{2.5}$ in multiple rulemakings.

[43] The CAA section 169A declares a national goal of preventing future and remedying any existing impairment of visibility in Class I areas.

[44] The Class I PSD increment SIL value starts with the NAAQS SIL value as the base number and is further constrained by the ratio of the associated Class I and II PSD increments. For the annual $PM_{2.5}$ NAAQS, the NAAQS SIL value is reduced by the ratio of 1:4, because the Class I PSD increment is 1 μg/m^3 and the Class II PSD increment is 4 μg/m^3. The ratio of 2:9 is used for the 24-hour $PM_{2.5}$ NAAQS. For the 24-hour NAAQS, we are using the 51.165(b)(2) value of 1.2 μg/m^3 as our base number.

The EPA recommends that permitting authorities consider using these SIL values for $PM_{2.5}$ and ozone on a case-by-case basis at the same points in the PSD air quality analysis as SIL values historically have been used in the PSD program, as described below, with one exception regarding defining the spatial extent for modeling.

First, permitting authorities may elect to use the SIL values reflected in this memorandum in a preliminary (single-source) analysis that considers only the impact of the proposed source in the permit application on air quality to determine whether a full (or cumulative) impact analysis is necessary before reaching a conclusion as to whether the proposed source would (or would not) cause or contribute to a violation.[45] A model result predicting that a proposed source's maximum impact will be below the corresponding SIL value recommended above generally may be considered to be a sufficient demonstration that the proposed source will not cause or contribute to a violation of the applicable NAAQS or PSD increment. If the single-source analysis shows that a proposed source will not have a significant impact on air quality, permitting authorities may generally conclude there is no need to conduct a cumulative impact analysis to assess whether there will be any violations of the NAAQS or PSD increment. However, upon considering the permit record in an individual case, if a permitting authority has a basis for concern that a demonstration that a proposed source's impact is below the relevant SIL value at all locations is not sufficient to demonstrate that the proposed source will not cause or contribute to a violation, then the permitting authority should require additional information from the permit applicant to make the required air quality impact demonstration.

Second, where the preliminary analysis described in the prior paragraph is not sufficient, permitting authorities may choose to use the recommended SIL values in a cumulative impact analysis for a NAAQS, which, in addition to the proposed source, includes the impact of existing sources (on and offsite), and the appropriate background concentration. The EPA has described this application of a SIL as a "culpability analysis."[46] Where a cumulative impact analysis predicts a NAAQS violation, the permitting authority may further evaluate whether the proposed source will cause or contribute to the violation by comparing the proposed source's modeled contribution to that violation to the corresponding SIL value. If the modeled impact is below the SIL value at the violating receptor during the violation, the EPA believes this will be sufficient in most cases for a permitting authority to conclude that the source does not cause or contribute to (is not culpable for) the predicted violation; thus, allowing the permit to be issued. If the proposed source's modeled impact is higher than or equal to the SIL value at the violating receptor during a violation, then a permit should not be issued unless (1) further modifications are made to the proposed source to reduce the proposed source's impact to an insignificant level at the affected receptor during the violation, or (2) the proposed source obtains sufficient emissions reductions from other sources to compensate for its contribution to the violation.[47]

[45] 1990 Draft NSR Workshop Manual at C.24-C.25, C.51.
[46] Prairie State, 13 E.A.D. at 100; Mississippi Lime, 15 E.A.D. at 374.
[47] 1990 Draft NSR Workshop Manual at C.52-C.53; this latter alternative is referred to as a PSD offset, and state implementation plans may include on offset program based on federal regulations at 40 CFR 51.165(b).

Third, permitting authorities may decide to use the SIL values recommended above in a cumulative impact analysis for a PSD increment. According to 40 CFR 51.166(c)(1) and 52.21(c), an allowable PSD increment based on an annual average may not be exceeded and the allowable PSD increment for any other time period may be exceeded during one such period per year at any one location. In either case, the PSD increment SILs recommended above may be used to determine if the proposed source will cause or contribute to that exceedance. If the cumulative impact analysis shows an annual average $PM_{2.5}$ PSD increment exceedance or a 24-hour PSD increment exceedance at a location, then the comparison of the proposed source's impact at that location during the exceedance to the corresponding SIL value may be used to determine whether the proposed source will cause or contribute to the exceedance(s) at that receptor. If the modeled impact is below the SIL and all other PSD requirements are met, then the permitting authority may conclude that the source does not cause or contribute to a violation of the PSD increment.

Finally, SILs have been used in defining the spatial extent of the modeling domain for a cumulative impact analysis. Because an impact from a proposed source below a SIL value is considered not to cause or contribute to a violation, the EPA has previously recognized that there was no informational value in placing modeling receptors farther from the proposed source than the most distant point at which the proposed source's impact is equal to or greater than the applicable SIL value. Streamlining the modeling demonstration to reduce the number of receptors to those of value in determining if the proposed source will cause or contribute to a violation of the applicable NAAQS or PSD increment has enabled permit applicants and reviewers to complete the required modeling with a reasonable effort. As discussed earlier, the EPA recently proposed updates to its Guideline on Air Quality Models. The revisions include providing an appropriate, revised basis for determining the modeling domain for NAAQS and PSD increment assessments. Once finalized, the revised Appendix W will be the appropriate resource to use when considering the extent of the modeling domain.

The SILs identified in this memorandum should not influence Air Quality Related Values analyses, which are independent reviews by the Federal Land Managers during the application review process.

Before a rulemaking is conducted and subject to limitations described in this memorandum, we recommend that permitting authorities consider using the values in the above tables on a case-by-case basis to support air quality analyses and demonstrations required for issuance of PSD permits. Permitting authorities that implement the PSD program under an EPA-approved implementation plan may also choose to use these recommended SILs. Since this memorandum is neither a final determination nor a binding regulation, permitting authorities retain the discretion not to use SILs as described here, either in specific cases or programmatically.

To ensure an adequate record, any PSD permitting decision that is based on the guidance in this memorandum should incorporate the information contained in this memorandum and the supporting technical and legal supporting documents. The permitting authority should also consider any additional information in the record that is relevant to making the required demonstration.

The permitting authorities also retain the discretion to use other values that may be justified separately from this memorandum as levels of insignificant impact, subject to one limitation for the $PM_{2.5}$ NAAQS. Since the EPA has established by regulation that a $PM_{2.5}$ impact greater than certain values will cause or contribute to a violation of the relevant NAAQS, permitting authorities may not use a value higher than 1.2 $\mu g/m^3$ for the 24-hour $PM_{2.5}$ NAAQS or a value higher than 0.3 $\mu g/m^3$ for the annual $PM_{2.5}$ NAAQS. Because ozone is not addressed in section 51.165(b)(2), permitting authorities are not precluded from developing a higher ozone NAAQS SIL value than recommended in this guidance. Likewise, section 51.165(b)(2) does not address PSD increments and, thus, does not constrain the discretion of a permitting authority to use a higher SIL value that a permitting authority may develop for increment purposes. Permitting authorities are also not precluded from developing and using lower SIL values than recommended in this guidance. The case-by-case use of a SIL value should be supported by a comparable record in each instance that shows that the value represents a level below which a proposed source does not cause or contribute to a violation of the NAAQS or PSD increment.

Please inform your permitting authorities of the guidance provided by this memorandum. If you have questions regarding policy or general implementation, please contact Raj Rao at *rao.raj@epa.gov* or (919) 541-5344. For questions regarding the supporting technical document, please contact Tyler Fox at *fox.tyler@epa.gov* or (919) 541-5562. For questions regarding the supporting legal document, please contact Brian Doster at *doster.brian@epa.gov* or (202) 564-1932.

Legal Support Memorandum
Application of Significant Impact Levels in the Air Quality Demonstration for Prevention of Significant Deterioration Permitting under the Clean Air Act

Introduction

Under section 165(a)(3) of the Clean Air Act (CAA or Act), an applicant for a pre-construction permit under the Prevention of Significant Deterioration (PSD) program must "demonstrate … that emissions from construction or operation of such facility will not cause, or contribute to, air pollution in excess of any" National Ambient Air Quality Standards (NAAQS) or PSD increment. 42 U.S.C. § 7475(a)(3). The law is clear that such a demonstration must be made to obtain a PSD permit. *Sierra Club v. EPA*, 705 F.3d 458, 465 (D.C. Cir. 2013). However, the CAA does not specify how a PSD permit applicant or permitting authority is to determine whether a new or modified source will (or will not) cause or contribute to a violation of a NAAQS or applicable PSD increment. *Id.* Considering the relevant terms of the CAA and other factors discussed below, permitting authorities may elect to read section 165(a)(3) of the Act to be satisfied when a permit applicant demonstrates that the increased emissions from the proposed new or modified source will not have a significant or meaningful impact on ambient air quality at any location where a violation of the NAAQS or PSD increment is occurring or may be projected to occur. This reading may be based solely on the EPA's historic interpretation of the phrase "cause, or contribute to," as specifically used in the context of section 165(a)(3) of the CAA, without relying on the inherent authority to establish exemptions for *de minimis* circumstances.

Background

Congress gave the EPA responsibility in the CAA for determining the methods to be used by PSD permit applicants to show that proposed construction does not cause or contribute to a

NAAQS or PSD increment violation.[48] Section 165(e) requires an analysis of "ambient air quality at the proposed site and in areas which may be affected by emissions from such facility" and directs the EPA to issue regulations that define the nature of this analysis. 42 U.S.C. § 7475(e). The regulations must "specify with reasonable particularity each air quality model or models to be used under specified sets of conditions" for purposes of the PSD program. In accordance with this authority, the EPA has promulgated regulations which identify such models and the conditions under which they may be used in the PSD program to make the demonstration required under section 165(a)(3) of the Act. 40 C.F.R. § 51.166(l); 40 C.F.R. § 52.21(l); 40 C.F.R. part 51, Appendix W (Guideline on Air Quality Models).

Using the models identified in the EPA regulations, there are two basic ways that a PSD permit applicant can demonstrate that the proposed source's emissions will not cause or contribute to a violation of any NAAQS or PSD increment. One way is to demonstrate that no such violation is occurring or projected to occur in the area potentially affected by the emissions from the proposed source. A second way is to demonstrate that the emissions from the proposed source do not cause or contribute to any violation of the NAAQS or PSD increments that is identified.

Analysis

Together, two aspects of the CAA reflect congressional intent to leave a gap for the EPA to fill in determining the precise meaning of the phrase "cause, or contribute to" in the context of section 165(a)(3) of the Act. First, as discussed above, section 165(e) of the Act directs the EPA to define the nature of the analysis that is necessary to make the demonstration required under section 165(a)(3) of the Act. Second, the phrase "cause, or contribute to" and the included terms "cause" and "contribute" are not defined in section 169, section 302 or any other part of the

[48] Section 165(a)(3) of the Act requires a showing that the applicant will not cause or contribute to air pollution "in excess of" the applicable NAAQS. The NAAQS are written using specific statistical forms, such as averages and/or percentile values across days, months and/or years. As a result, a set of air quality concentrations over a certain period is not considered "in excess" of a NAAQS unless the applicable statistical criterion for not meeting the NAAQS is satisfied. In order to distinguish a situation in which a set of air quality concentrations is "in excess" of the NAAQS from a single measurement or prediction that might exceed the numerical level of the NAAQS, the EPA typically uses the term "violation" to describe a period of air quality that is "in excess of" the standard, considering the statistical form of the standard. The term "exceedance" refers to a single measurement or prediction above the level of the NAAQS.

CAA. The EPA and other PSD permitting authorities may reasonably infer that Congress's silence "is meant to convey nothing more than a refusal to tie the agency's hands" as to the degree of air quality impact necessary to "cause, or contribute to" air pollution in excess of air quality standards under section 165(a)(3) of the CAA. *See Entergy Corp. v. Riverkeeper, Inc.*, 556 U.S. 208, 222 (2009).

The United States Court of Appeals for the District of Columbia Circuit has observed that the term "contribute" is ambiguous. *Catawba County, N.C. v. EPA*, 571 F.3d 20, 38-39 (D.C. Cir. 2009). In this case, the court considered the use of this term in section 107(d) of the CAA, which governs EPA actions to designate specific areas as in attainment or nonattainment with the NAAQS. Under this provision, a nonattainment area must include any area that does not meet the NAAQS or "that contributes to ambient air quality in a nearby area that does not meet" the NAAQS. The Petitioners argued that the EPA was required to interpret the word "contribute" in this context to require a "significant causal relationship" in order to include a nearby area in a nonattainment area. The Petitioners also argued that the EPA must establish a quantified amount of impact that qualifies as a contribution before the EPA could include a nearby area in a nonattainment area. *Id.* The court held that "section 107(d) is ambiguous as to how the EPA should measure contribution and what degree of contribution is sufficient to deem an area nonattainment." Consequently, the Court held that the EPA was not compelled to apply the Petitioners' preferred meaning of the term "contribute" in the context of section 107(d). The court recognized that the EPA had the discretion to interpret the term "contribute" in section 107(d) of the Act to mean "sufficiently contribute" and that the EPA could use a multi-factor test, rather than a quantified threshold, to determine when a nearby area contributed to nonattainment.

Similar to section 107(d) of the Act, section 165(a)(3) is ambiguous with regard to the degree of air quality impact that is necessary to conclude that increased emissions from an individual source will "contribute to" a violation of a NAAQS or PSD increment. In the absence of specific language in section 165(a)(3) regarding the degree of contribution that is required (such as the term "significantly"), the reasoning of the *Catawba County* opinion supports the view that the EPA has the discretion under this provision to exercise its judgment to determine the degree of impact that "contributes" to adverse air quality conditions based on the particular context in which the term "contribute" is used. *See,* 571 F.3d at 39. Furthermore, this opinion supports EPA's discretion in implementing section 165(a)(3) to identify criteria or factors that may be used to determine whether something "contributes" (including qualitative or quantitative criteria), as long as the Agency provides a reasoned basis to justify using such criteria to represent a "contribution."

In the particular context where contribute is used in the PSD permitting program, this part of the CAA does not prohibit all proposed construction that increases emissions. Rather, the program contemplates that increased emissions resulting from construction or modification of major stationary sources may be authorized after verifying that the proposed construction will incorporate state-of-the-art pollution controls and that the operation of the new or modified major source will not result in unhealthy levels of air pollution (or significantly increase air pollutant concentrations) in the affected area. The PSD program required by Congress is specifically designed to prevent "significant" deterioration of air quality, not all deterioration of air quality. Further, one goal of the PSD program is to "insure that economic growth will occur in a manner consistent with the preservation of existing clean air resources." 42 U.S.C. § 7470(3). Thus, the PSD program strikes a balance that allows construction and modification of major stationary sources that will result in increased emissions, but only after appropriate safeguards are in place to prevent significant deterioration of existing clean air resources.

In light of these considerations, the inclusion of the phrase "cause, or contribute to" in section 165(a)(3) of the Act indicates that Congress intended for the reviewing authority to exercise some judgment in the course of reviewing a permit application. Section 165(a)(3) of the Act does not say a source must show it has "no impact" on a predicted violation. Instead, this provision says the source must show it does not "cause, or contribute to" a violation. This choice by Congress militates against reading section 165(a)(3) to mean that any degree of projected impact (no matter how small) must be considered to contribute to a predicted violation of the NAAQS or PSD increment. Under such a reading, the permitting authority need not exercise any judgment. A source could only qualify for a permit by showing that there would be no violation of the NAAQS or PSD increment in the area affected by the source or that emissions from the source have no projected impact whatsoever on any area where the NAAQS or PSD increment is already or predicted to be exceeded. If Congress had intended in section 165(a)(3) to preclude permitting authorities from exercising discretion to determine the degree of impact that equals a contribution, it would have used a less ambiguous term or specified that no degree of impact on a predicted violation is permissible.

In addition, Congress explicitly recognized that air quality models would be needed to make the showing required under section 165(a)(3) to obtain a PSD permit, and directed the EPA to specify such models in regulations. 42 U.S.C. § 7575(e)(3). Given their mathematical nature, models used for this purpose under the PSD program are capable of predicting small increases in air pollutant concentrations. In order for the "cause or contribute" language in section 165(a)(3) to be implementable as a practical matter in permitting, there must be some point at which a projected air quality impact from a proposed new or modified source becomes so small that PSD permitting authorities may reasonably conclude that such an impact does not cause, or contribute to, an existing or predicted violation of air quality standards.

Furthermore, the PSD permitting requirements in part C of Title I of the Act are one of many required elements of a State Implementation Plan (SIP) under section 110 of the Act. *See generally* 42 U.S.C. 7410(a)(2). The PSD permitting requirements are specifically incorporated as one of these elements under section 110(a)(2)(C) of the Act. The focus of the PSD program is on controlling increased emissions from the construction and modification of large stationary sources, while other provisions under section 110(a)(2) require states to target emissions from existing sources. Where air quality concentrations are high in a specific area because of sources already in operation, section 110 and other provisions of the Act provide tools for addressing this existing pollution through a SIP. In this context, where existing sources have already caused air quality to very nearly approach or even exceed a NAAQS, it is not necessary to construe the PSD provisions to require a permit applicant to show that increased emissions will have absolutely no effect on air quality concentrations. The goals of the PSD program are achieved by demonstrating that increased emissions from construction or modification of the source will be controlled to the point that these emissions will not have a meaningful impact on air quality in the affected area, while looking to other aspects of a SIP to address emissions from existing sources that bear responsibility for high levels of air pollution in the area.

Recognizing this, the EPA has previously supported the use of concentration values called "significant impact levels" (SILs) to represent the point below which the impact of increased emissions from a new or modified major source on ambient air quality does not cause or contribute to a violation of the NAAQS or PSD increment. 61 Fed. Reg. 38250, 38293 (July 23, 1996).[49] At the same time, where a violation is nevertheless predicted in the course of the PSD permitting process, EPA has emphasized the need to address the source of such air pollution problem through a SIP under section 110 of the Act, rather than preventing construction that will not meaningfully add to the adverse conditions. *See* Memorandum from Gerald A. Emison, EPA OAQPS, to Thomas J. Maslany, EPA Air Management Division, EPA Region 3, "Air Quality Analysis for Prevention of Significant Deterioration (PSD)" (July 5, 1988).

This practice in the PSD program has been based, in part, on an interpretation by EPA that the phrase "cause, or contribute to" in section 165(a)(3) does not to apply to an "insignificant" impact. In this context, EPA has used the term "insignificant" to describe a degree of impact that is "trivial" or "*de minimis*" in nature. Conversely, in this context, the EPA has described an impact that is greater than "trivial" or "de minimis" as a "significant impact," which the EPA has represented quantitatively using the values called "significant impact levels." As expressed by the EPA's Environmental Appeals Board (EAB), "EPA has long interpreted the phrase 'cause, or contribute to' to refer to significant, or non-de minimis, emission contributions." *In re Prairie State Generating Co.*, 13 E.A.D. 1, 105 (EAB 2006). Based on a review of the plain terms of the CAA in context, the EAB reasoned in this case that "the requirement of an owner or operator to demonstrate that emissions from a proposed facility will not 'cause, or contribute to' air pollution in excess of a NAAQS standard must mean that some non-zero emission of a NAAQS parameter is permissible." *Id.* at 104. The EAB also illustrated

[49] The historic use of a quantified threshold for this purpose in the PSD program differs from the EPA's practice of using a multi-factor test to define "contribution" in the context of designations under section 107(d) of the CAA. *See Catawba County, N.C. v. EPA*, 571 F.3d 20, 38-39 (D.C. Cir. 2009). While this case held that a quantified threshold is not required to define contribution in the context of section 107(d), the court's reasoning does not preclude PSD permitting authorities from choosing to use a quantitative level of impact to represent a contribution to a violation of the NAAQS or PSD increment when implementing section 165(a)(3) of the Act. For purposes of implementing section 165(a)(3) of the Act, the EPA has found it more expedient and practical to use a quantitative threshold (expressed as a level of change in air quality concentration) to determine whether increased emissions from proposed construction or modification of a source will contribute to air quality concentrations in excess of applicable standards. Under the reasoning of *Catawba County*, using a quantified threshold for this purpose is permissible as long as the EPA or the appropriate permitting authority provides a reasoned explanation for why impacts below that threshold do not constitute a contribution to a violation in this context.

how this historic interpretation of section 165(a)(3) of the Act "is reflected in both applicable EPA regulations and in long-standing EPA guidance." Id.

One example of such an EPA regulation was section 10.2.3.2(a) of the EPA's Guideline on Air Quality Models (40 C.F.R. Part 51, Appendix W). This provision of Appendix W addressed proposed sources "predicted to have a significant ambient impact" and called for permitting authorities, in evaluating whether the source will cause or contribute to an air quality violation, to consider "the significance of the spatial and temporal contribution to any modeled violation." The EPA has recently proposed to revise and reorganize the Guideline on Air Quality Models, and an examination of whether a proposed source has a "significant ambient impact" is reflected in several sections of the proposed Guideline. 80 Fed. Reg. 45340 (July 29, 2015) (*see* sections 4.2(c), 8.1.2(a), and 9.2.3(a)).

In a 1988 guidance memorandum, the EPA said that "a PSD source will not be considered to cause or contribute to a predicted NAAQS or PSD increment violation if the source's estimated air quality impact is insignificant (i.e. at or below defined de minimis levels)." Memorandum from Gerald A. Emison, EPA OAQPS, to Thomas J. Maslany, EPA Air Management Division, EPA Region 3, "Air Quality Analysis for Prevention of Significant Deterioration (PSD)" (July 5, 1988). Extending this logic, in 1990, the EPA also said that a permit applicant may demonstrate that it will not cause or contribute to air pollution in violation of any NAAQS or PSD increment by showing that the "proposed source will not result in a significant ambient impact anywhere." 1990 NSR Workshop Manual, C.51 (Oct. 1990). More specifically, the EPA has generally considered it sufficient for an applicant to demonstrate that the source's emissions alone have an insignificant impact on air quality in the area outside a facility fence line that is defined as "ambient air." *See* In the Matter of Hibbing Taconite Co., 2 E.A.D. 838 (Adm'r 1989); NSR Workshop Manual at C.42, C.52.

In this context, the EPA has often equated an insignificant impact with one that is trivial or *de minimis* in nature. In some instances, the intent of such statements by the EPA has been to justify an exemption to the requirement in section 165(a)(3) of the CAA based on the agency's inherent authority to exempt *de minimis* circumstances from regulation. *See Alabama Power v. Costle*, 636 F.2d 323, 361-63 (D.C. Cir. 1980). After initially proposing in 1996 to add SILs to its PSD regulations but not taking final action on that regulation proposal, the EPA proposed such a regulation in 2007 for only the $PM_{2.5}$ pollutant and finalized that rule in 2010. 75 Fed. Reg. 64864 (Oct. 10, 2010).[50] In that rule, the EPA said that "the concept of a SIL is grounded on the de minimis principles described by the court in Alabama Power." *Id.* at 64891. The EPA repeated this statement in a subsequent administrative order where the EPA also said that the Agency "has interpreted the de minimis doctrine to generally support use of the SILs … for purpose of determining whether a proposed source or modification contributes to predicted violation of a NAAQS." Order Responding to Petitioner's Request that the Administrator Object to Issuance of a State Operating Permit, In the Matter of CF&I Steel, L.P. dba EVRAZ Rocky Mountain Steel, Petition Number VIII-2011-01, at 15-17 (May 31, 2012) ("Rocky Mountain Steel Order"). This order referenced two prior opinions of the EAB that referenced the discussion of the *de minimis* doctrine in the D.C. Circuit's opinion in *Alabama Power*. In the first of these opinions, the EAB observed that "Courts have long recognized that EPA has discretion under the Clean Air Act to exempt from review some emissions increases on the grounds of de minimis or administrative necessity." *Prairie State*, 13 E.A.D. at 104 (internal quotations omitted). However, as discussed above, in this same opinion, the EAB also described how the EPA has interpreted the phrase "cause, or contribute to" to refer to significant emission contributions. *Id.* at 105.

Considering EPA's longstanding and permissible interpretation of the phrase "cause, or contribute to" in section 165(a)(3) and the intended role and function of SILs, it was unnecessary for the EPA to reference its inherent *de minimis* exemption authority in these actions to justify the conclusion that an insignificant impact does not cause or contribute to a violation of the NAAQS or PSD increment within the meaning of section 165(a)(3) of the Act. As historically used on a permit-by-permit basis prior to the 2010 rule, the air quality concentration levels that

[50] In response to a challenge to the 2010 $PM_{2.5}$ SILs regulation in the District of Columbia Circuit, EPA requested that the court remand and vacate two of the EPA's SILs regulations for $PM_{2.5}$ so that EPA could correct an inconsistency between the inflexible terms of the regulation and EPA's exhortation in the record that permitting authorities should exercise discretion before using these values in some circumstances to justify the conclusion that a source does not cause or contribute to a violation of the NAAQS. *Sierra Club*, 705 F.3d at 463-64. The court then vacated these two $PM_{2.5}$ SIL provisions adopted in 2010 "because they allow permitting authorities to automatically exempt sources with projected impacts below the SILs from having to make the demonstration required under 42 U.S.C. §7475(a)(3) even in situations where the demonstration may require a more comprehensive air quality analysis." Id. at 465. The court said that "[o]n remand, the EPA may promulgate regulations that do not include SILs or do include SILs that do not allow the construction or modification of a source to evade the requirement of the Act as do the SILs in the current rule."

the EPA has identified as SILs have not functioned to exempt a source from making the demonstration required by section 165(a)(3) of the Act. Rather, these concentration levels have been used by PSD permit applicants and permitting authorities as a means of making the air quality impact demonstration required by section 165(a)(3). To determine that its increased emissions will not exceed these concentration values, a new or modified source must conduct air quality modeling to determine the degree of impact the source will have on air pollutant concentrations. If the applicant thereby shows that its increased emissions do not have a significant impact on air pollutant concentrations, EPA and other permitting authorities have concluded that the applicant has made a demonstration that its increased emissions will not cause or contribute to any air pollutant concentrations that exceed the relevant NAAQS or PSD increment.

The EPA has previously communicated this view that the statutory requirement in section 165(a)(3) of the Act may be satisfied by showing that a source does not have a significant impact on air pollutant concentrations. In its 2007 proposal of the $PM_{2.5}$ SILs, the EPA said that when "a source can show that its emission alone will not increase ambient concentrations by more than the SILs, the EPA considers this to be a sufficient demonstration that a source will not cause or contribute to a violation of the NAAQS or increment." 72 Fed. Reg. 54112, 54139 (Sept. 21, 2007). The EPA has subsequently expressed similar thoughts in a guidance memorandum. *See e.g.*, Memorandum from Acting Director of Air Quality Policy Division to Regional Air Division Directors, *General Guidance for Implementing the 1-hour NO_2 National Ambient Air Quality Standards in Prevention of Significant Deterioration Permits, Including an Interim 1-hour NO_2 Significant Impact Level*, at 11 (June 28, 2010) ("2010 NO_2 Guidance"). In the 2012 Rocky Mountain Steel Order described above, the EPA observed that a "SIL was a means of demonstrating through modeling that the source's impact at the time and place of the predicted violation will be sufficiently low that such impact will not contribute to that violation."

Although the EPA also referenced its inherent authority to establish a *de minimis* exception to a statutory requirement in these same documents, it was unnecessary for the agency to do so because the phrase "cause, or contribute to" in section 165(a)(3) of the Act is reasonably read not to apply to insignificant impacts on air quality. Likewise, in order to show that a particular degree of change in concentration is insignificant in this context, it is not necessary to make the showing required to establish a *de minimis* exception from a statutory requirement – that the burdens of regulation yield a gain of trivial or no value. Rather, when a concentration value (which may be described as a SIL) is used to quantify the point below which a new or modified source does not cause, or contribute to, a violation of the NAAQS or PSD increment, it is sufficient for the EPA or a state permitting authority to justify the value as a level below which an impact on air quality may be regarded as not significant or meaningful. In general terms, a trivial or *de minimis* impact on air quality may be considered "insignificant," but the use of a SIL to identify such a level in the PSD program need not be based on inherent agency authority to establish a *de minimis* exception to section 165(a)(3) of the Act. The statutory language in this provision is reasonably construed in context not to apply to an insignificant impact on air quality.

While use of a SIL in PSD permitting need not be based on an agency's inherent authority to establish a *de minimis* exception to a statutory requirement, any value used as a SIL must be supported by an appropriate record showing that impacts below that level will not cause, or contribute to, a violation. Thus, in the context of a case-by-case decision by a permitting authority to issue a PSD permit and to use a specific SIL value in making the demonstration required in section 165(a)(3) of the Act, such permit must be supported by a record showing that the SIL value is representative of a level below which the projected impact of a proposed new or modified stationary source is insignificant. *See* Rocky Mountain Steel Order at 18; 2010 NO_2 Guidance at 11.

Technical Basis for the EPA's Development of Significant Impact Thresholds for PM2.5 and Ozone

EPA-454/D-16-001a

July 2016

Technical Basis for the EPA's Development of Significant Impact Thresholds for PM2.5 and Ozone

U.S. Environmental Protection Agency

Office of Air Quality Planning and Standards

Air Quality Analysis Division

Air Quality Modeling Group

Research Triangle Park, NC

Contents

1.0 Introduction

Under the Clean Air Act (CAA), in a variety of contexts, the EPA evaluates the extent to which individual sources or collections of sources in a particular geographic area "contribute" or "contribute significantly" to degradation of air quality. In order to understand the nature of air quality, the EPA statistically estimates the distribution of pollutants contributing to ambient air quality and the variation in that air quality. The statistical methods and analysis detailed in this report focus on using the conceptual framework of *statistical significance* to identify levels of change in air quality concentrations that the EPA considers to be a "significant impact" or an "insignificant impact" contribution to air quality degradation. *Statistical significance* is a well-established concept with a basis in commonly accepted scientific and mathematical theory. The statistical methods and data reflected in this analysis may be applicable for multiple regulatory applications where EPA seeks to identify a level of impact on air quality that is either a "significant impact" or "insignificant impact" by considering a range of values for which the *statistical significance* is examined.

While this technical analysis may have utility in several contexts, one of the primary purposes here is to quantify the degree of air quality impact that can be considered an "insignificant impact" for the Prevention of Significant Deterioration (PSD) program. In order to obtain a preconstruction permit under the PSD program, an applicant must demonstrate that the increased emissions from its proposed modification or construction will not "cause or contribute to" a violation of any National Ambient Air Quality Standard (NAAQS) or PSD increment (*i.e.*, the source will not have a significant impact on ambient air quality at any location where an exceedance of the NAAQS or PSD increment is occurring or may be projected to occur).[51] Compliance with the NAAQS is determined by comparing the measured "design value" (DV) at an air quality monitor to the level of the NAAQS for the relevant pollutant.[52] A design value is a statistic or summary metric based on the most recent one or three years (depending on the specific standard) of monitored data that describes the air quality status of a given location relative to the level of the NAAQS.[53]

The EPA believes that an "insignificant impact" level of change in ambient air quality can be defined and quantified based on characterizing the observed variability of ambient air quality levels. Since the cause or contribute test is applied to the NAAQS, this analysis has been designed to take into account the ambient data used to determine DVs and the form of the relevant NAAQS. The EPA's technical

[51] Code of Federal Regulations; Title 40(Protection of Environment); Part 51;Sections 51.166 and 52.21

[52] A design value is a statistic that describes the air quality status of a given location relative to the level of the NAAQS. More information may be found at: http://www3.epa.gov/airtrends/values.html.

[53] In order to differentiate the usage of 'significant' between the contextual application in the PSD program and as a mathematical assessment, we have adopted the following convention throughout the document: a "significant impact" (quotes) refers the analysis of the ambient impacts from a facility in the context of the "causes, or contributes to" clause in the evaluation of a violation of the applicable NAAQS or PSD increment, whereas we use *significant* (italics) to refer to a mathematical assessment of probabilistic properties.

approach, referred to as the "Air Quality Variability" approach, relies upon the fact that there is inherent variability in the observed ambient data, which is in part due to the intrinsic variability of the emissions and meteorology controlling transport and formation of pollutants, and uses statistical theory and methods to model that intrinsic variability in order to facilitate identification of a level of change in DVs that is acceptably similar to the original DV, thereby representing an *insignificant* change in air quality.[54] The DVs and background ambient concentrations that are used in the PSD compliance demonstrations are obtained through the U.S. ambient monitoring network with measured data being archived for analysis in the EPA's Air Quality System (AQS).[55]

Based on these observed ambient data, the EPA's technical analysis has estimated the distribution of the air quality levels of ozone and $PM_{2.5}$ through applying a well-established statistical approach known as bootstrapping. Bootstrapping is a method that allows one to construct measures to quantify the uncertainty of sample statistics (*e.g.*, mean, percentiles) for a population of data.[56,57] The bootstrap approach applied here uses a non-parametric, random resampling with replacement on the sample dataset (*e.g.*, in this case, the ambient air quality concentration data underlying the DVs), resulting in many resampled datasets. This approach allows measures of uncertainty for sample statistics when the underlying distribution of the sample statistic is unknown and/or the derivation of the corresponding estimates is computationally unfeasible or intractable.[7] Bootstrapping is also commonly utilized to overcome issues that can occur when quantifying uncertainty in samples with correlated measurements. Bootstrapping has been used across a variety of scientific disciplines and in a wide range of applications

---

[54] This approach is applied here strictly for the purpose of section 165(a)(3) and not other parts of the Clean Air Act.

[55] The Air Quality System (AQS) contains ambient air pollution data collected by EPA, state, local, and tribal air pollution control agencies from over thousands of monitors. These data are used to assess air quality, assist in attainment/nonattainment designations, evaluate State Implementation Plans for nonattainment Areas, perform modeling for permit review analysis, and other air quality management functions. More information may be found at: http://www.epa.gov/aqs.

[56] Efron, B. (1979); "Bootstrap methods: Another look at the jackknife". The Annals of Statistics 7 (1): 1–26. doi:10.1214/aos/1176344552.

[57] Efron, B. (2003); Second Thoughts on the Bootstrap. Stat. Sci., 18, 135-140.

within the environmental sciences.[58,59,60,61] For example, bootstrapping has been used to evaluate the economic value of clinical health analyses[62] and environmental policies,[63] evaluations of environmental monitoring programs,[64] and determining uncertainty in emissions inventories.[65] Additionally, the EPA has used bootstrapping techniques as a key component in evaluating air quality model performance for use in our nation's air quality management system.[66,67]

The bootstrap technique, as applied in this analysis, quantifies the degree of air quality variability at an ambient monitoring site and allows one to determine confidence intervals (CIs), *i.e.*, statistical measures of the variability associated with the monitor-based DVs, to inform the degree of air quality change that can be considered "insignificant impact" for PSD applications. This approach for quantifying an "insignificant" air quality impact is fundamentally based on the idea that an anthropogenic perturbation of air quality that is within a specified range may be considered indistinguishable from the inherent variability in the measured atmospheric concentrations and is, from a statistical standpoint, *insignificant* at the given confidence level. Specifically, the analysis uses 15 years (2000-2014) of nationwide ambient ozone and $PM_{2.5}$ measurement data from the AQS database to generate a large number of resampled datasets for ozone and $PM_{2.5}$ DVs at each monitor. These resampled datasets are used to determine CIs that provide a measure of the inherent variability in air quality at the monitor location. This variability may be driven by the frequency of various types of meteorological and/or emissions conditions impacting a particular location. The analysis estimates a range of CIs for each monitor; the 50% CI was

58 Schuenemeyer , J., Drew, L. (2010); Statistics for Earth and Environmental Scientists, John Wiley & Sons, Inc. http://dx.doi.org/10.1002/9780470650707.ch3

59 Park, Lek, Baehr, Jørgensen, eds. (2015); Advanced Modelling Techniques Studying Global Changes in Environmental Sciences, 1st Edition, Elsevier. ISBN 9780444635365.

60 Chandler, R., Scott, M. (2011); Statistical Methods for Trend Detection and Analysis in the Environmental Sciences, John Wiley & Sons, Inc. ISBN: 978-0-470-01543-8

61 Mudelsee, M. & Alkio, M. (2007); Quantifying effects in two-sample environmental experiments using bootstrap confidence intervals, Env. Mod. & Software, 22, 84-96.

62 Campbell, M., & Torgerson, D. (1999); Bootstrapping: Estimating Confidence Intervals for Cost-effectiveness Ratios, Q. J. of Med., 92, 177-182.

63 Kochi, I., Hubbell, B., & Kramer, R. (2006); An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis, Env. & Resource Econ., 34, 385-406.

64 Levine, C., et al (2014); Evaluating the efficiency of environmental monitoring programs, Ecol. Ind., 39, 94-101.

65 Tong, L., et al (2012); Quantifying uncertainty of emission estimates in National Greenhouse Gas Inventories using bootstrap confidence intervals, Atm. Env., 56, 80-87.

66 Hannah, S. (1989); Confidence limits for air quality model evaluations, as estimated by bootstrap and jackknife resampling methods, Atm. Env., 6, 1385-1398.

67 Cox, W. & J. Tikvart (1980); A statistical procedure for determining the best performing air quality simulation model, Atm. Env., 9, 2387-2395.

selected to quantify the bounds of air quality levels that represent a *statistically insignificant* deviation from the inherent variability in air quality, from which the change that can be considered an "insignificant impact" for the purposes of meeting requirements under the PSD program can be determined.

This technical basis document explains the analysis design and results that are applicable to Significant Impact Levels (SILs) in the PSD program. The second section of this document provides an overview of EPA's Air Quality Variability approach, including details on the ambient monitoring network, the ambient ozone and $PM_{2.5}$ data from AQS that are used to derive monitor-specific DVs, a general review of *statistical significance* and confidence intervals, and a description of the bootstrap technique as applied to characterize air quality variability. The third section presents the measures of air quality variability determined from applying the bootstrap technique to the AQS data for ozone and $PM_{2.5}$. The last section provides an analysis of confidence intervals for the ozone and $PM_{2.5}$ DVs and then recommends specific values of the change in air quality that can serve as "significance impact" levels for the ozone NAAQS and the annual and 24-hour $PM_{2.5}$ NAAQS.

2.0 Background on Air Quality Variability Approach

This section provides details on the ambient monitoring data for ozone and $PM_{2.5}$ that were used in the EPA's Air Quality Variability approach and the statistical methods that form the technical basis for the EPA's Air Quality Variability approach.

2.1 U.S. Ambient Monitoring Data

The EPA's understanding of the nation's air quality is based on an extensive ambient monitoring network, which is used to determine the compliance with the various NAAQS. In addition to providing data for use in determining compliance with the NAAQS, the monitoring network is designed to inform the public about the status of air quality across the nation and to support air pollution research, particularly in the evaluation and development of updated NAAQS. The general requirements of the monitoring network are given in 40 CFR Appendix D to Part 58 (Network Design Criteria for Ambient Air Quality Monitoring). These general requirements and choices made by the state and local air agencies that operate the monitoring stations have resulted in monitoring sites across the nation with a variety of characteristics in terms of location, monitoring equipment, and operating schedule.

NAAQS compliance is determined by comparing the DV derived from a monitor's data to the level of the NAAQS for the relevant pollutant. The DV is a particular statistic determined from the distribution of data from each monitor and is consistent with the averaging period and statistical form of the relevant NAAQS. The DVs from an area's monitoring network are used to determine attainment status for that area. The DVs for $PM_{2.5}$ and ozone are determined as follows:

- For the primary ozone NAAQS, the DV is the 3-year average of the annual 4th-highest daily maximum 8-hr average (MDA8) ozone concentration.[68] A monitor is in compliance if the DV is less than or equal to the level of the standard, which was recently revised to be 0.070 ppm (70 ppb.)[69]
- For the primary annual $PM_{2.5}$ NAAQS, the DV is the 3-year average of the $PM_{2.5}$ annual mean mass concentrations.[70] The annual mean is defined as the mean of the data in each of the 4

[68] Appendix U to Part 50 - Interpretation of the Primary and Secondary National Ambient Air Quality Standards for Ozone

[69] National Ambient Air Quality Standards for Ozone, 80 Fed. Reg. 65292 – 65468 (Oct. 26, 2015)

[70] Appendix N to Part 50—Interpretation of the National Ambient Air Quality Standards for $PM_{2.5}$

quarters of the year. A monitor is in compliance with the 2012 annual primary $PM_{2.5}$ standard if the DV is less than or equal to 12.0 μg/m^3.[71]

- For the 24-hr $PM_{2.5}$ NAAQS, the DV is the 3-year average of the annual 98th percentile 24-hr average $PM_{2.5}$ mass concentration. A monitor is in compliance with the 24-hr $PM_{2.5}$ standard if the DV is less than or equal to 35 μg/m^3.

2.1.1 Ozone Monitoring Network

The ozone monitoring network consists of only one type of monitor, Federal Equivalent Method (FEM) monitors.[72] The FEM for ozone uses ultraviolet (UV) light to determine ozone concentrations at high temporal resolutions, on the order of seconds to minutes, although only hourly averages are typically recorded. Unlike $PM_{2.5}$ monitors, most ozone monitors are not required to operate year-round, and are instead required to operate only during the "ozone season." The ozone season is the time of year that high ozone concentrations (which may potentially violate the NAAQS) can be expected at a particular location. The ozone season varies widely by location, but is generally focused on the summer months, with a typical season spanning March through October. During the period of 2000 through 2014, a total of 1,708 ozone monitors reported data, with the locations of the ozone monitors shown in Figure 1 along with the average number of days sampled each year that the monitor was active.

[71] There is a secondary $PM_{2.5}$ NAAQS, with a level of 15 μg/m^3. The work here focuses only on the primary NAAQS at 15 μg/m^3, since compliance with the primary standard explicitly implies compliance with the

secondary standard as well.

[72] FEM monitors are approved on an individual basis. The list of approved monitors and the accompanying CFR references can be found at http://www3.epa.gov/ttn/amtic/criteria.html.

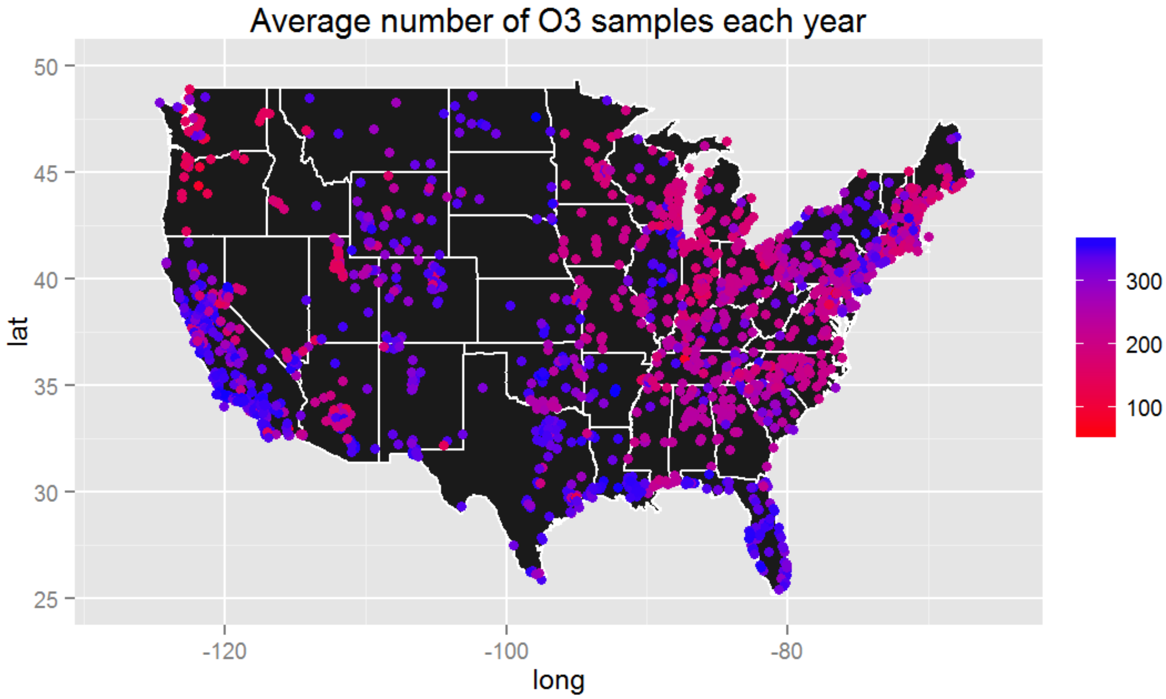


Figure 1 - Location and average number of monitored ozone days each year from the ozone sampling network for the years 2000-2014.

2.1.2 $PM_{2.5}$ Monitoring Network

The $PM_{2.5}$ monitoring network consists of two types of monitors: Federal Reference Method (FRM)[73] and FEM[72] monitors. FRM monitors use a filter-based system, passing a low volume of air through a filter over a period of 24 hours (midnight to midnight) to determine 24-hr average concentrations. All monitors operate year-round, but not all monitors operate every day throughout the year. Although some FRM sites operate every day (*i.e.*, 1:1 monitors), most operate every third day (1:3 monitors), while a smaller number of monitors operate only every sixth day (1:6 monitors), according to a common schedule provided by EPA. Newer FEM monitors are "continuous" monitors that can provide hourly (or shorter) $PM_{2.5}$ measurements. FEM monitors operate on a 1:1 schedule and daily averages from FEM monitors are determined by averaging the 24 hourly measurements collected throughout the day. FEM monitors are slowly replacing FRM monitors, so monitoring sites with a long data record may have data derived from either an FEM, FRM, or combination of both types of monitors. During the period of 2000

[73] Appendix B to Part 50—Reference Method for the Determination of Suspended Particulate Matter in the Atmosphere (High-Volume Method)

through 2014, a total of 1,773 $PM_{2.5}$ monitors reported data, with the locations of the $PM_{2.5}$ monitors shown in Figure 2 along with the average number of days sampled each year that the monitor was active.

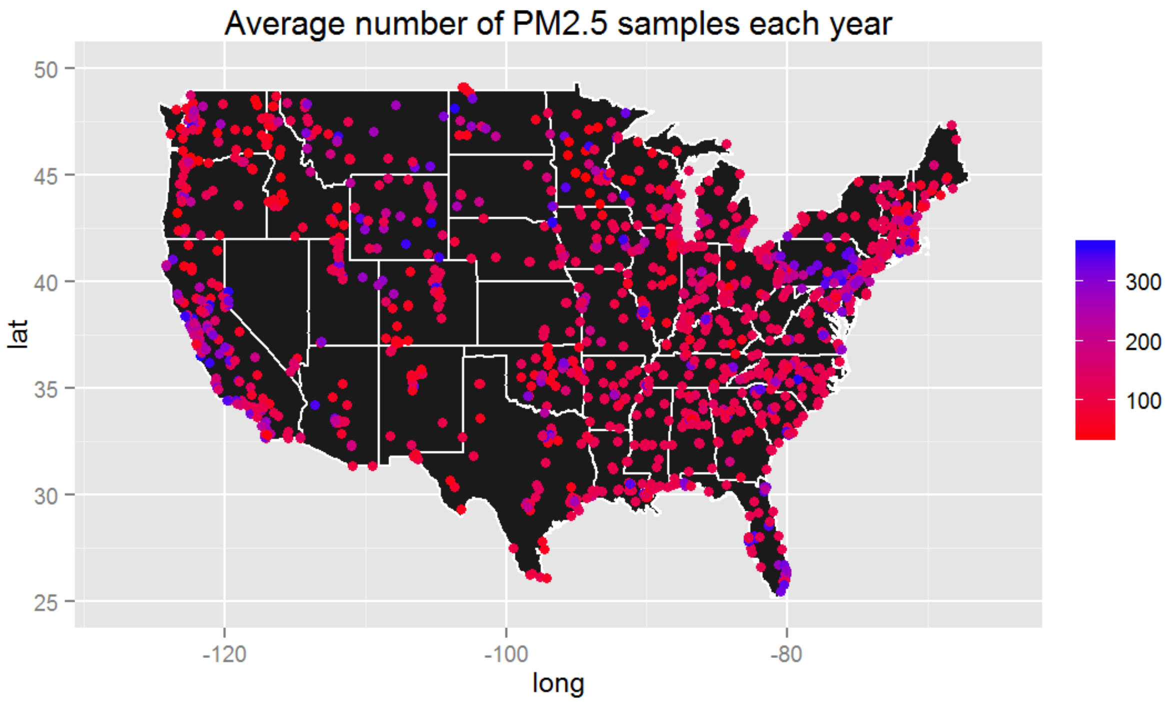


Figure 2 - Location and average number of monitored PM days each year from the $PM_{2.5}$ sampling network for the years 2000-2014.

2.1.3 Monitoring Network Design

The ambient air monitoring network is designed to support several objectives. In consideration of the location and measurement taken, each monitor is assigned a spatial scale. Spatial scales are generally associated with the size of the area that a pollutant monitor is representative of.-The monitor spatial scales are defined in 40 CFR 58 appendix D as:

1. *Microscale*—Defines the concentrations in air volumes associated with area dimensions ranging from several meters up to about 100 meters.
2. *Middle scale*—Defines the concentration typical of areas up to several city blocks in size with dimensions ranging from about 100 meters to 0.5 kilometer.
3. *Neighborhood scale*—Defines concentrations within some extended area of the city that has relatively uniform land use with dimensions in the 0.5 to 4.0 kilometers range. The neighborhood and urban scales listed below have the potential to overlap in applications that concern secondarily formed or homogeneously distributed air pollutants.

4. *Urban scale*—Defines concentrations within an area of city-like dimensions, on the order of 4 to 50 kilometers. Within a city, the geographic placement of sources may result in there being no single site that can be said to represent air quality on an urban scale.
5. *Regional scale*—Defines usually a rural area of reasonably homogeneous geography without large sources, and extends from tens to hundreds of kilometers.
6. *National and global scales*—These measurement scales represent concentrations characterizing the nation and the globe as a whole.

Depending on the distribution and types of sources in an area and the need to determine particular aspects of the air quality, there may be multiple types of monitors placed in an area. For example, a large metropolitan area, due to its size, may require several "urban scale" or “neighborhood” scale monitors to capture the range of air quality in the area. Such an area might also have "microscale" monitors placed in order to assess the impacts from a single source or small group of sources as well as a "regional scale" monitor to establish the background air quality in an area in order to differentiate the impacts from the urban area. Conversely, for a smaller urban area a single "urban scale" monitor may be considered sufficient to fully characterize the local air quality. Thus, there are wide variety of monitors in any area, covering a range of air quality monitoring needs. For ozone, the appropriate spatial scales are neighborhood, urban, and regional scale. For PM2.5, in most cases the appropriate spatial scales are neighborhood, urban, or regional scales; however, in some cases it may be appropriate to monitor at smaller scales, depending on the monitoring objective.

2.1.4 Air Quality System (AQS) Database

The EPA’s AQS database contains ambient air pollution data collected by state, local, and tribal air pollution control agencies, as well as EPA and other federal agencies, from the monitoring stations described above (as well as monitoring stations for other NAAQS).55 AQS also contains meteorological data, descriptive information about each monitoring station, and data quality assurance/quality control information. The Office of Air Quality Planning and Standards (OAQPS), state and local air agencies, tribes, and other AQS users rely upon the system data to assess air quality, assist in attainment/ nonattainment designations, evaluate state implementation plans for nonattainment areas, perform modeling for permit review analysis, and execute other air quality management functions related to the CAA.

2.2 Statistical Methods and Assessing Significance using Confidence Intervals

This section provides a general overview of statistical methods, how air quality variability is characterized for this analysis, and the bootstrapping approach employed to estimate air quality variability.

2.2.1 General Overview of Statistical Methods

Statistics is the application of mathematical and scientific methods used to interpret, analyze and organize collections of data. Most statistical techniques are based on two concepts, a "population" and a "sample." The *population* represents all possible measurements or instances of the entity being studied. The *sample* is a subset of the *population* that is able to be collected or measured. Since the *sample* is only a portion of the *population*, any observations or conclusions made about the *population* based on the *sample* will have uncertainty, *i.e.*, there will be some error in those observations or conclusions due to the fact that only a subset of the population was sampled or measured. Consider the following example:

> As discussed above, the ambient monitoring network is designed to capture a range of ambient impacts from facilities and to characterize both background and local air quality. Suppose we want to determine the average ground-level $PM_{2.5}$ levels in a remote state wilderness area over the course of a year. Since the wilderness area does not have major $PM_{2.5}$ sources and the area is remote (*i.e*., there are no major metropolitan areas upwind), a single, well-placed "regional scale" monitor may be sufficient to capture the nature of $PM_{2.5}$ levels in the area (*i.e.*, the $PM_{2.5}$ levels within the wilderness area are homogenous). Due to the remote nature of the monitor, it is only operated on a 1-in-every-6 days schedule, such that one 24-hr average $PM_{2.5}$ measurement is made every six days. In this case, we may consider the *population* to be the 24-hr average $PM_{2.5}$ concentrations every day (365 potential samples over the whole year) within the wilderness area. The *sample* would be the 1-in-every-6 days 24-hr average $PM_{2.5}$ measurements (60 samples taken over the whole year). From this *sample* of the *population*, a mean 24-hr average $PM_{2.5}$ concentration can be calculated, which can be characterized as representing the mean 24-hr average $PM_{2.5}$ concentration from the *population*, with some amount of error between the *sample* mean and the *population* mean. By using information about the size and distribution of the *sample*, an estimate of the population variability, *i.e.*, the spread of the distribution, can be determined (*e.g.*, the standard deviation).

Significance testing, or determining the *statistical significance* of a particular value as it relates to a *sample*, is a major application of statistics. In formal hypothesis testing, a statement of non-effect or no difference – termed the null hypothesis – is established prior to taking a sample in order to test the effect of interest. A statistical test is then carried out to determine whether a *significant* effect (or difference) is present at the desired level of confidence. Note that not finding a *statistically significant* difference is not a claim of the null hypothesis being true or a claimed probability of the truth of the null hypothesis.[74] *Non-significance* simply shows the data to be compatible with the null hypothesis under the set of assumptions associated with the statistical test.[74] A confidence interval can be used as a mathematically equivalent procedure[74] to a formal hypothesis test for significance. Commonly used statistical techniques employ the size and other characteristics of the *sample* to determine error bars for the mean. These error bars are also referred to as Confidence Intervals (CIs) because they convey the confidence in the *sample* estimate of the *population*. CIs are determined based on the desired confidence level, given the size of and the variability in the sample. This can then be used to determine if the mean is *significantly* different from a particular value of interest, such as zero or some other

[74] Gelman, A. P values and Statistical Practice, *Epidemiology*, 2013, Vol 24, Num 1, pg 70.

threshold for the pollutant, by examining whether the value of interest is within the CI or outside the bounds of the CI.

The most well-known approach to deriving confidence intervals uses the characteristics of sampling distributions and the Central Limit Theorem. The sampling distribution of the mean results from sampling all possible samples of a specified size *n* from the true population and considering the distribution of the resulting means from each sample. The Central Limit Theorem is based on the fact that the sampling distribution of the sample mean will center around the population mean. Regardless of the distribution of the original population, the sampling distribution of the mean will be normally distributed.[75] Additionally, the sampling distribution will have a spread, with a standard deviation that is inversely proportional to the square root of the sample size *n* – *i.e.*, the larger the sample size, the tighter the spread of the sampling distribution of the mean around the true mean of the population. This allows for the derivation of a CI by calculating the estimated mean plus/minus the standard error, which is a function of the sample size, the standard deviation, and the desired level of confidence.

To continue the hypothetical example from above:

> Suppose that the annual mean $PM_{2.5}$ concentration for a given year is 7 μg/m^3, and that based on the Central Limit Theorem utilizing the properties of the sampling distribution, the 95% CI for the annual mean is determined to be 6.4-7.6 μg/m^3 (7 μg/m^3 +/- 0.6 μg/m^3, where 0.6 μg/m^3 has been determined based on the standard error and the desired level of confidence). Since the CI contains the value 7.5 μg/m^3, we may therefore conclude based on this specific sample that the mean of the population is *not significantly* different from 7.5ug/m^3 at the 0.95 confidence level. Conversely, if the 95% CI for the annual mean $PM_{2.5}$ concentration is 6.7-7.3 μg/m^3 (7 μg/m^3 +/- 0.3 μg/m^3) then the CI does not contain 7.5 μg/m^3 and it could be concluded that the mean of the population is *significantly* different from 7.5 μg/m^3 at the 0.95 confidence level.

The Central Limit Theorem also tells us that due to the Gaussian (normal distribution) properties of a sampling distribution, 68/95/99.7 percent of the values in the theoretical sampling distribution will be within 1/2/3 standard deviations of the true population mean respectively. Additionally, in any symmetric distribution such as the Normal Distribution obtained with the theoretical sampling distribution, the mean is equal to the median, where the median is the center value such that 50% of the values are below the median and 50% above. Thus, an alternative approach to deriving a confidence interval directly utilizes these characteristics of the sampling distribution to consider the spread around the sampling distribution mean. For example, a 95% CI would be defined as the lowest value to the highest value of the 95% of the distribution that centers around the sampling distribution mean. This corresponds to the 0.025 and 0.975 quantiles of the sampling distribution. An example of this method of determining CIs is given in Figure 3, which shows a distribution of the mean determined from repeated *samples* from the *population*. Note that in practice the sampling distribution is approximately normal.

---

[75] These are asymptotic properties given that the sample size n is large and that the number of samples (N) drawn from the population is large – in theory, all possible samples of size n are drawn from the population. (Moore and McCabe, 4th Ed, 2003 – p. 262.) In practice, n $\geq$ 30 and N is often 1,000, 10,000, or as determined by convergence of distributional characteristics, and the resulting sampling distribution is approximately normal.

The average of the sample means is 6.98 µg/m^3. In order to determine the 95% CI, the data is first rank-ordered from smallest to the largest concentration value, then the bounds of the 0.025 and 0.975 quantiles are the bounds of the CI (the 50% CI is also shown as an example).

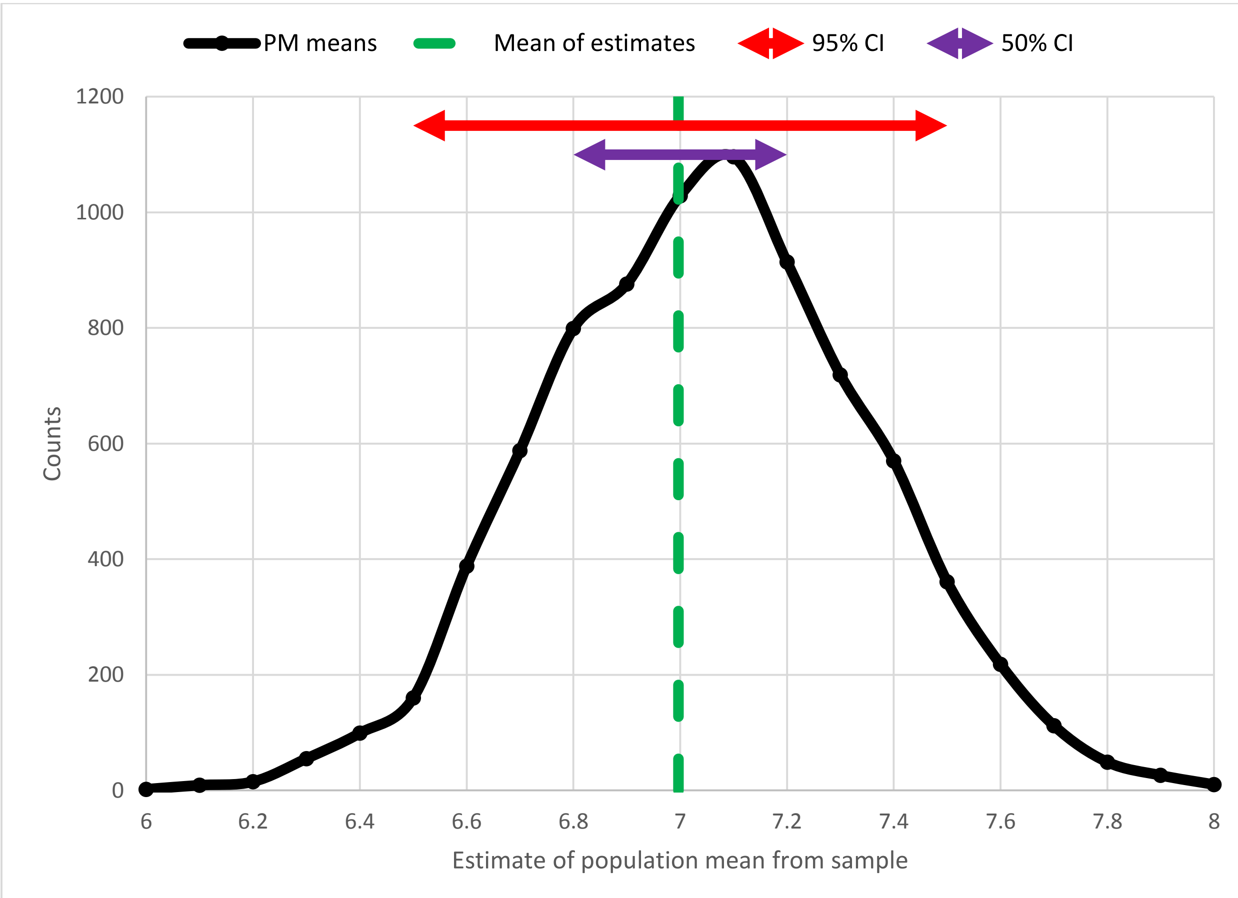


Figure 3- Example of CIs determined from a distribution of sample means.

The techniques utilizing the sampling distribution to make inferences about the population mean can be applied to other statistics as well, such as sample quantiles. Additionally, a statistical technique applied as resampling from one particular drawn sample, known as “bootstrapping”, can be used to generate estimated confidence intervals for any desired statistic. Bootstrapping is further explained in Section 2.2.3.

The CIs for any sample comparison are generally affected by three main factors: the size of the sample, the variability within the sample, and the confidence limits desired for the comparison (*e.g.*, 0.95 level of confidence was used in the example above). Increasing the sample size (taking more measurements or samples) will increase the representativeness of the sample of the population and decrease the variance associated with the calculated measurement, resulting in narrower CIs. Samples from populations with greater inherent variability will have greater uncertainty and result in larger CIs. Finally, increasing the confidence level of the inferred conclusion will necessitate larger CIs, while lower confidence thresholds will result in narrower CIs. There are clearly many complicated aspects of significance testing, many of which require subjective selections by the analyst to insure that the results are appropriate to the application and to reduce the influence of uncontrolled variables on the results and conclusions. These selections are usually made based on convention and standard practice, such as choosing a 95% confidence level. While there are many more applications of statistical techniques and nuances of the principles described above, these basic concepts of the population, sample, CIs (and their relationship to probability) are the fundamental concepts used in the development of "significant impact" thresholds presented here.

2.2.2 Characterizing Air Quality Variability

As discussed in Section 2.1, the DV from a particular monitor is the air quality statistic that is used to describe the air quality in an area (*e.g.*, the annual mean was the statistic from the example above) and is compared to the NAAQS to determine attainment status for that area. Within the conceptual framework discussed in the previous section, the ambient data from a single monitor are a *sample* of a *population* of the air quality and the uncertainty in that sample stems from the inherent variability that occurs in air quality. The inherent variability is driven by a collection of factors, both natural (meteorological) and anthropogenic (emissions), which can be grouped into spatial and temporal categories.

2.2.2.1 Spatial variability

The spatial variability is the change in air quality that is present at any one moment across an area. This variability is driven by the spatial distribution of sources (causing localized increases in ambient concentrations due to their emissions), removal or sinks (causing localized decreases in ambient concentrations due to physical or chemical processes), variations in chemical production for secondarily formed $PM_{2.5}$ and ozone (which do not have direct emissions sources), and meteorology (wind patterns may transport air from areas with higher emissions to areas that typically have lower concentrations due to fewer localized emissions). The spatial variability is directly addressed in the network design, *i.e.*, the spatial scale associated with each monitor and the potential need for multiple monitors to characterize the air quality in an area. One way to estimate the spatial variability is to compare ambient monitors that are in close proximity to one another. Such monitors would likely show similar trends in the ambient concentrations, with some variation due to changes in emissions and meteorology responsible for transporting pollutants and affecting chemical conversion, creation, and removal of atmospheric species that are specific to each individual location.

These spatial variations occur in the population of air quality levels and can be estimated from the existing sample (*i.e.*, data available from the ambient monitoring network). Depending on the intended scale of the monitor, there is some room for interpretation as to the interpretation of the observed sample of air quality data, and this interpretation has implications for the determination of the uncertainty associated with the sample. Given the nature of the variability in air quality, there are three potential populations represented by the sample and the spatial variability between the sample and the population:

1. If the population is considered to be the air quality at the location of the monitor only, then there is no spatial variability.
2. If the population is considered to be the air quality in the immediate vicinity of the monitor, then there will be some spatial variability, the degree of which will depend on nearby sources and sinks and the distance of the location of interest from these sources and sinks. For $PM_{2.5}$, if there is a nearby source of primary $PM_{2.5}$, changes in wind direction and mixing conditions will change where these nearby sources have impacts, such that there would be more spatial variability on this small scale. If there is no nearby source of primary $PM_{2.5}$, then secondary $PM_{2.5}$ would dominate and there would likely be little small-scale spatial variability on this small scale. For ozone, the same is true, in that there will likely be little spatial variability unless there are nearby sources that act as a sink (*i.e.*, major NOx source such as a highway or point source). Without a nearby sink, then the secondary nature of ozone would generally indicate that there is little spatial variability on this small scale.
3. If the population is considered to be the air quality over a larger scale (*e.g.*, a county or Core Based Statistical Area or CBSA), then there is much more spatial variability. As with case 2, the presence and location of sources and sinks will impact how much spatial variability is present, though on such a large scale, there are likely to be many sources and sinks across the area, resulting in more spatial variability.

As discussed in Section 2.2.1, monitoring sites are assigned a spatial scale, which are associated with the size of the area for which a particular monitoring site should be representative of the air quality. For secondarily formed pollutants, Appendix D to Part 58 states that the highest concentration monitors may include urban or regional scale monitors (*i.e.*, 50 to hundreds of km spatial scale). Intuitively, it would be expected that the air quality changes across these distance scales, such that the air quality across such a large area is not identical to the air quality as determined by a single monitor. Indeed, these classifications are supportive of the idea that there are spatial variations, such that multiple monitors are generally needed to adequately characterize the air quality in an urban area. However, in rural areas with few emissions sources, a single monitor may be sufficient to characterize the air quality over hundreds of square km.

2.2.2.2 Temporal Variability

In the example introduced in Section 2.2.1, there may be uncertainty not only from the limited sampling of the population, but also based on changes in the *population* occurring with time.

Temporal variability is the variability in air quality that occurs over time, which is driven by changes in emissions and meteorology over a range of time scales. For shorter time scales, diurnal patterns in both emissions and meteorological processes can impact most atmospheric pollutants. Mobile source emissions, which can substantially contribute to atmospheric pollution, have particularly strong daily (*i.e.*, rush-hour) and weekly (no rush-hour on the weekends) patterns. Day-to-day meteorological variability (*i.e.*, frontal passages and synoptic weather patterns) can also cause temporal variability on the timescale of days to weeks. At intermediate time scales, seasonal changes in weather can have a major impact in transport patterns and chemical reactions. There can be seasonal trends in emission patterns as well, particularly those associated with energy production and mobile source emissions. At longer time scales, there can be longer-term trends in meteorology (*e.g.*, particularly warm or wet years) and emission sources (sources being added or removed or changes in emissions due to emissions controls or economic conditions) that results in long-term air quality variability. Temporal variability is reflected in the form of the standard (*i.e*., compliance with each ozone and $PM_{2.5}$ standard is based on 3 years of data in order to reduce from the impact of temporal variability on NAAQS implementation programs). This variability can be addressed by requiring continuous monitoring in an area, even after air quality levels in an area are below the level of the standard. The long-term temporal variability can be characterized by examining changes in air quality over time at a particular monitor (*e.g.*, trends in DVs or other metrics from the monitor). The shorter-term temporal variability can be described by examining the hourly and daily changes in air quality or by comparing data from periods with similar meteorological conditions (*e.g.*, afternoon, weekdays versus weekends, or summertime concentrations).

Whatever the spatial scale of the monitor, temporal variability will always contribute to the air quality variability, as there will always be day-to-day changes in meteorology and emissions and variability between seasons and years, which may or may not include any trends in emissions and meteorology. The form of the standard (*e.g*., annual average or a ranked daily value), the temporal resolution of the monitoring data (*e.g.*, hourly or 24-hr averaged samples), and the frequency of the sampling (*e.g*., daily samples or samples taken every sixth day) may affect the ability of the monitoring data to fully capture the inherent temporal variability and thus increase the uncertainty in any statistic or DV derived from a particular sample. If a monitor has some missing data, then it is easy to conceptualize that there is some uncertainty caused by temporal variability in that there are days and hours that are not represented by the monitor. On the other hand, if a monitor has a perfect sampling record, then the uncertainty due to reduced sampling frequency is eliminated, but there remains long-term variability. Since the $PM_{2.5}$ and ozone DVs are based on 3 years of data, there is variability between the years that affect the DVs. As noted above, the use of a 3-year DV, rather than a DV derived from 1 or 2 years of data, is geared towards increasing the stability (or reducing the variability) of the DVs. Despite the 3-year DV, from a statistical standpoint, there remains uncertainty in the DVs, as the DVs are statistics derived from samples of the population of air quality.

The importance of temporal variability is perhaps more apparent when the application of the DVs are considered. For area designations purposes, the DVs are always historical (updated DVs for a particular year are published in the following calendar year), such that the DV is just an estimate of the current state of the air quality in an area. Furthermore, in the permitting process, DVs from past years are paired with modeling of past years of meteorology and planned future emissions. Thus, the changes from year-to-year and the uncertainty in estimating future air quality levels are illustrative of important factors affecting temporal variability that impacts regulatory applications and exists regardless of the completeness of the sampling record or the spatial scale defining the population discussed above.

Continuing the example from Section 2.2.1, suppose that after 1 year of sampling, there is some commercial development adjacent to the wilderness area, such that new buildings and larger traffic volumes are present during the second year of the monitor's operation. One might want to assess whether or not the new activity has had a notable impact on the average $PM_{2.5}$ concentrations within the wilderness area. A comparison between the scenarios can be considered, and the idea that the difference between the two may be "notable" can be evaluated by comparing that difference to the estimated CIs created by the bootstrap procedure using the concepts in significance testing (Section 2.2.1).

2.2.2.3 Assessing air quality variability

Based on the description of the population determined above, the DV can be understood to be a statistic determined from a sample of the population. CI's for a particular DV can then be used to compare the DV with another DV or a constant value (*e.g.*, the NAAQS). If the CI of the sample mean contains the value of interest, then mean and the value of interest are statistically indistinguishable from one another, given the sample data available at a particular confidence level. In the context of an air quality analysis, if a CI can be determined for a DV, then it can be concluded that a value within some given amount of variation of a DV (*i.e.*, within a CI for that DV) is *statistically insignificant* with respect to that selected level of confidence. Note that in this context *non-significance* simply shows the data to be compatible with an assumption of no difference between the value and the DV.[74]

2.2.3 Bootstrapping Method

For annual-average standards (*i.e*., averages of many samples during 1 or 3 years), there are standard parametric methods (*e.g*., the standard deviation) that might be used to estimate variability associated with DVs. When a statistic with difficult to estimate variance under parametric assumptions, such as a rank order statistic, is of interest, some other approach must be taken to determine CIs. For non-normal populations, there are some adjustments that can be made to determine CIs of the mean if the data conforms to some standard distribution (*e.g.*, log-normal). For small sample sizes, other non-parametric tests such as the Mann-Whitney[76] test or the Wilcoxon signed-rank test[77] may be used. However, for many statistics (*e.g.*, the 98th percentile), the underlying distribution of the statistic may be complicated or unknown, and thus determination of the CIs for these statistics can be difficult or impossible to determine with traditional metrics.[78] Of the three NAAQS considered here, the annual $PM_{2.5}$ standard is the only NAAQS that is based on a sample mean. However, the calculation of the DV statistic for the

[76] Mann, H. B.; Whitney, D. R. (1947). "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other". Annals of Mathematical Statistics 18 (1): 50–60. doi:10.1214/aoms/1177730491

[77] Wilcoxon, F. (Dec 1945). "Individual comparisons by ranking methods". Biometrics Bulletin 1 (6): 80–83

[78] Woodruff, R. S. (1952); Confidence intervals for medians and other position measure. J. Amer. Stat. Assoc., 47, 635–646, doi:10.1080/01621459.1952.10483443.

annual $PM_{2.5}$ NAAQS is more complicated than merely taking a simple arithmetic average of the 24-hr $PM_{2.5}$ values across 3 years; thus deriving the distribution of the annual $PM_{2.5}$ DV statistic is not straightforward. The CIs for the 24-hr $PM_{2.5}$ and ozone NAAQS are based on rank-order statistics (98th percentile for $PM_{2.5}$ and 4th highest daily maximum 8-hr ozone concentration, see section 2.1) which cannot be easily described using standard statistical techniques. Thus, for the three DV statistics being analyzed here, an alternative technique to determine CIs is needed.

The bootstrapping method mentioned above is a well-established and accepted statistical method that allows one to estimate the underlying distribution of many sample statistics (*e.g.*, mean, percentiles, and correlation coefficients) when the theoretical distribution is complicated or unknown.[56,57,]**Error! Bookmark not defined.** The bootstrap method relies on the underpinnings and characteristics of sampling distributions discussed in Section 2.2. The estimate of the distribution is accomplished by resampling with replacement from the initial dataset many times, resulting in many resampled datasets (bootstrapped samples). The sample statistic of interest is then computed from each resampled dataset, resulting in an empirical estimate of the sampling distribution for the desired statistic. This estimate of the sampling distribution can then be used to determine CIs for the statistic of interest. Bootstrapping does not require any distributional assumptions for the population, nor does it require that there be an established formula for estimating the uncertainty in the statistic.

Meaningful information on the variability associated with the ozone and $PM_{2.5}$ DVs can be derived by using bootstrapping to assess the variability associated with the three DV statistics (*i.e.*, the ozone DV, the annual $PM_{2.5}$ DV, and the 24-hr $PM_{2.5}$ DV).[58] This analysis uses ambient $PM_{2.5}$ and ozone measurement data taken from the EPA's AQS database to determine CIs for each monitor for 3-year DV periods (*i.e.*, the 3 years of ambient data required to compute a DV for these NAAQS). The CIs give a measure of the temporal variability in air quality represented by each monitor. A nationwide analysis of the variability and changes in this variability over time is also conducted. Finally, the results from this analysis are applied to determine appropriate “significant impact” thresholds based on air quality variability.

The dataset used for this technical analysis comes from the AQS database described in Section 2.1 and is the same dataset that would be used for determining the DV at any particular monitor. The ambient $PM_{2.5}$ concentration data used for this analysis consist of 24-hr averaged samples, while the ozone data consist of 8-hr averaged concentrations (*i.e.*, the MDA8’s). This includes data from all of the monitoring sites in the EPA's AQS database from the years of 2000 to 2014.[79]

The bootstrapping estimates used in this analysis were calculated independently for each monitoring site, and the bootstrapping resamples at each site were taken independently within each calendar year.

[79] Raw daily and hourly measurements from Federal Reference Method (FRM) and Federal Equivalent Method (FEM) monitors are aggregated by AQS into a single daily value for each sampling site and NAAQS (annual and 24-hr) according to the procedures described in Appendix N of 40 CFR Part 50. The aggregation procedures in AQS include accounting for multiple monitors at sites, handling of exceptional events (which can be different between the two $PM_{2.5}$ NAAQS), and calculating a 24-hr value from 1-hr measurements. These results reside in the "site_daily_values" table of AQS, which were downloaded for use in the current analysis.

The re-sampling within each year is completed such that the re-sampled year contains the same number of days as the original data. The number of measurements varies by monitoring site and can have important implications for the inherent variability. The variation in the sampling schedule is explored further in Section 3.2.2. The re-sampling and computation of new DVs at each site are conducted to mimic the DV calculation procedures as closely as possible, which differ for each NAAQS.[68,70]

- For the annual $PM_{2.5}$ NAAQS, the data from each year was further subset by quarter (*i.e.*, Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec), such that the re-sampling did not allow for data from one quarter to occur in another quarter. The resulting re-sampled dataset was averaged by quarter; then the quarterly means were averaged to find the annual mean, with the DV being computed as the average of the 3 annual means.
- For the 24-hr $PM_{2.5}$ NAAQS, the data from each year was subset by quarter (*i.e.*, Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec), such that the re-sampling did not allow for data from one quarter to occur in another quarter. The number of days in each quarter was kept equal to the corresponding number in the original dataset. While this isolation of quarters is not a feature of the DV calculation procedure, it was applied as a precaution to avoid changing the seasonal balance in the bootstrapped samples. The resulting re-sampled dataset was then ranked, and the 98th percentile value was selected based on the number of daily measurements in each year, as described in Table 1 of Appendix N. The DVs were then computed as the average of the three annual 98th percentile values.
- For the ozone NAAQS, all available data at each site were used. The ozone monitoring regulations require monitoring for the "ozone season," which typically varies by state. Many states operate a subset of ozone monitors outside of the required monitoring season and when that data is available it is used in determining DVs for regulatory purposes. Therefore, if a monitor operated beyond the required ozone season, all valid data were included in the DV calculation. For example, if the required monitoring season was from April-October, but data from November were also available, then the MDA8 values from April-November were ranked in order to find the 4th highest value. The DVs were then computed as the average of the three annual 4th highest MDA8 values. Though the regulations for processing ozone data to compute a DV do not involve segregation of the data by season, a sensitivity analysis was conducted to determine the impact of applying the same quarterly segregation used for $PM_{2.5}$. The results are summarized in Section A.4 of the Appendix, but the results indicated relatively little sensitivity to this choice for most sites, and thus no quarterly segregation was applied for the final analysis.

For both $PM_{2.5}$ and ozone, each year of data from each site was re-sampled 20,000 times. The distributions derived from the bootstrap analysis did not appear to change after 3,000-4,000 re-samples for several single calendar years. Therefore, 20,000 re-samples were chosen to conservatively ensure that stable results were obtained for all cases. For each 1-year re-sample for each pollutant, the relevant annual statistic was computed (annual mean for $PM_{2.5}$, 98th percentile for $PM_{2.5}$, and 4th highest MDA8), giving 20,000 estimates of the annual statistic for each year. In order to replicate the way in which the standard is calculated, the data from each year is resampled separately from the other years. In order to calculate the bootstrap samples in a manner consistent with the DV calculations, *i.e*., calculating averages and 98th percentile values in each year independently, then averaging the 3 annual values, each of the 20,000 estimates for year 1 were averaged with the corresponding 20,000 estimates for year 2 and year 3, giving 20,000 estimates of the DV. From the 20,000 estimates, the mean, median, standard deviation, maximum, minimum, 25%, 50%, 75% and 95% CIs for the mean,[80] were computed and retained for further analysis. For symmetric distribution such as the Normal Distribution obtained with the sampling distribution, the mean is equal to the median, where the median is the center value such that 50% of the values are below the median and 50% above. Thus a bootstrapped CI for the mean is analogous to a bootstrapped CI for the median and the CIs can be calculated by rank-ordering the bootstrap results and selecting the bounds that contain the corresponding percentage of data. Since data from 2000-2014 were processed, all possible 3-year DVs from 2002-2014 were computed, for a total of 13 DV-years, including five 3-year periods that had non-overlapping years (*i.e.*, 2000-2002, 2003-2005, 2006-2008, 2009-2011, and 2012-2014).[81] As we are defining the CIs as the bounds of the uncertainty and a measure of the air quality variability, we frequently refer to each CI as the uncertainty associated with the actual DV.

The following gives an example of how the CIs are determined utilizing the percentile method[82] for the 24-hr $PM_{2.5}$ DVs from a monitor:

- Consider the dataset X_0, which contains 150 measurements of 24-hr averaged $PM_{2.5}$ monitoring values from year 1. Datasets Y_0 and Z_0 contain data from the same site, but for years 2 and 3 respectively, and contain 250 and 350 days of data respectively.
- From X_0, we calculate the 98th percentile as the 3rd highest value in the dataset. From Y_0, we calculate the 98th percentile as the 5th highest value in the dataset. From Z_0, we calculate the 98th percentile as the 7th highest value in the dataset. The DV for this site is the average of the 98th percentiles from X_0, Y_0, and Z_0.
- From X_0, 20,000 new sample datasets, X_1, X_2, … , $X_{20,000}$, each with 150 measurements of $PM_{2.5}$ are sampled with replacement from the original dataset X_0 Likewise, 20,000 new sample datasets are sampled with replacement from Y_0, and Z_0.
- For each X_i, the 98th percentile value is the 3rd highest value, for each Y_i, the 98th percentile is the 5th highest value, and for each Z_i, the 98th percentile is the 7th highest value. Thus, the DV for each subset, DV_i, is the average of the 3rd high value from X_i, the 5th highest value from Y_i, and the 7th highest value from Z_i. This calculation yields 20,000 different DVs.

[80] Here, and elsewhere in this document, a CI for the median is the interval spanning the data that contains ½ of the CI of the data above the median and ½ of the CI of the data below the median of the re-sampled DV estimates. For example, the 50% CI consists of the 25% of the data above the median and the 25% of the data below the median.

[81] Later in this document, whenever a single year is used to identify a DV, it refers to the last year of the 3-year period.

[82] Efron, B.; Tibshirani, R. (1993); An Introduction to the Bootstrap. Boca Raton, FL: Chapman & Hall/CRC. ISBN 0-412-04231-2.

- To determine the CIs from these 20,000 DVs, the DVs are ranked from low to high. Then the lower bound for the 50% CI is the 5,000th ranked DV, and the upper bound for the 50% CI is the 15,000th ranked DV. That is, the CIs are determined simply by ranking the resulting distribution of DVs and the q% CI for the mean is the bounds of the center of the data that contains q percentage of the results (ie, the lower bound is the $q/2^{th}$ percentile and the upper bound is the $(1\text{-}q/2)^{th}$ percentile).

Section A.1 provides several illustrative examples of the bootstrapping analysis for both the annual and 24-hr $PM_{2.5}$ NAAQS with actual data from 6 different sites.

3.0 Results of the Air Quality Variability Approach

This section provides results on characterizing the variability of air quality for ozone and $PM_{2.5}$ based on EPA's Air Quality Variability approach.

3.1 Ozone results

The results from the bootstrap analysis for the 2012-2014 ozone DVs are shown in Figure 4, which shows the mean, median, minimum, and maximum bootstrap DVs for each monitor, as well as the upper and lower bounds of the 25%, 50%, 75%, and 95% CIs for the median DV calculated from the 20,000 bootstrap samples as a function of the DV determined from the original dataset (top panel), the relative differences between the CI DVs and the actual DVs (middle panel), and box-and-whisker plots of the distribution of the relative difference at each CI (bottom plot). The mean and median of the bootstrap DVs for the ozone NAAQS replicate the actual DV from the original site data fairly well, with some very small deviations (maximum deviation is less than 5%). Even though the ozone NAAQS is based on peak values (similar to the 24-hr $PM_{2.5}$ NAAQS), the magnitude of the relative variability in the ozone bootstrap DVs ranges from 1-5%, with maximums around 25-30%. This is likely due to the nature of ozone formation, *i.e.*, ozone is almost exclusively a secondarily formed pollutant, with precursors typically originating from multiple sources, rather than a single source. There is a component of reaction/formation time, both of which are likely to reduce the spatial variability and temporal variability of the ambient ozone. There is an increase in the absolute variability with an increase in the baseline DVs, but there is not a trend in the relative variability. This indicates that the baseline air quality does not systematically affect the relative amount of variability at a site. This is especially important because it indicates that a central tendency value for the relative variability in the DV for the ozone NAAQS is stable across levels of ozone concentrations. Therefore a representative value can be multiplied by the level of that NAAQS to obtain a value in concentration units (ppb for ozone) that is appropriately used to characterize variability for sites with air quality that "just complies" with the NAAQS.

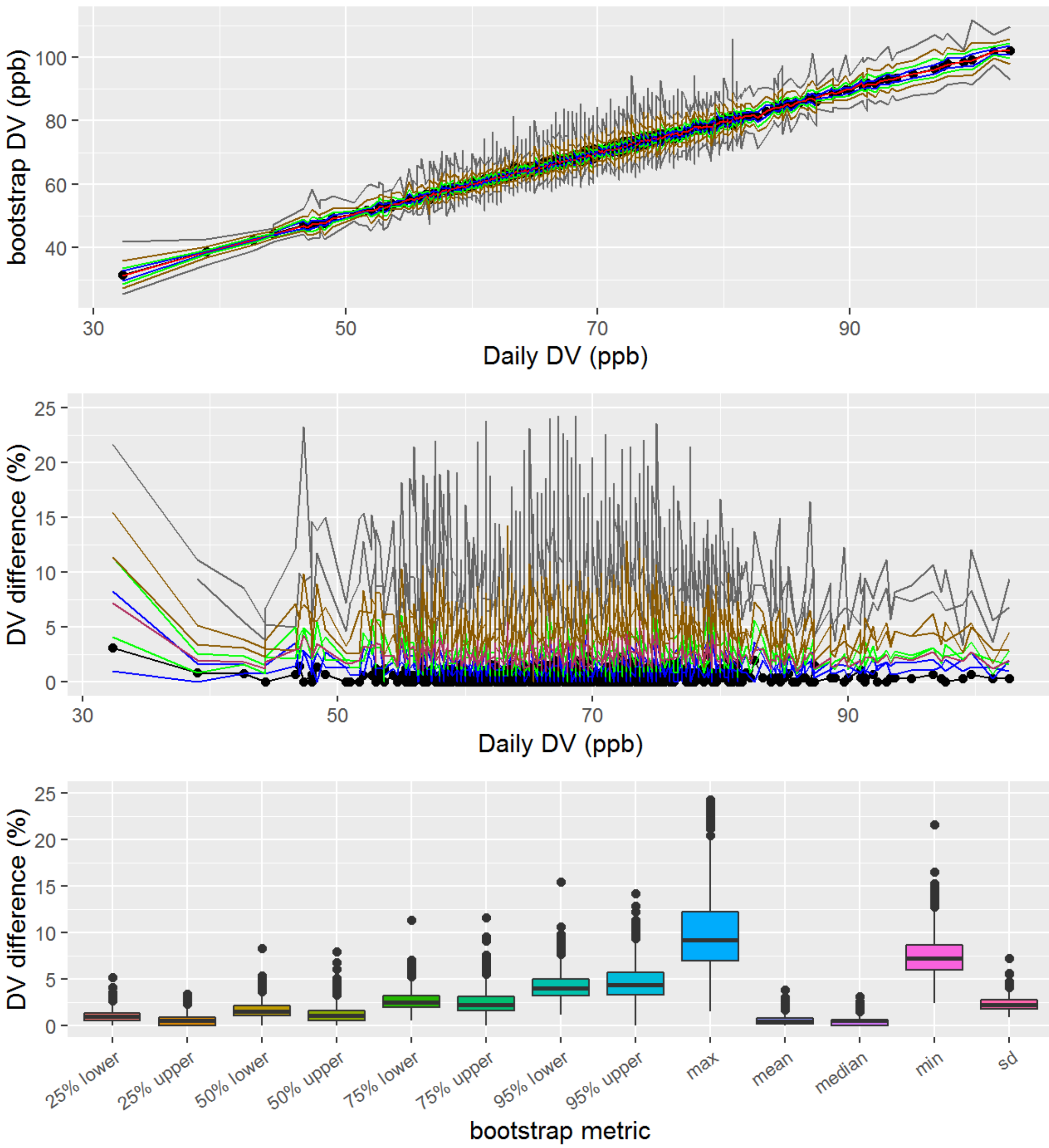


Figure 4 -- Bootstrap results for the ozone 2012-2014 DVs (25%, 50%, 75%, and 95% CIs, along with the mean and median bootstrap DVs) Top panel shows the values for the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

3.2 $PM_{2.5}$ Results (Annual and 24-hr)

The results from the bootstrap analysis for the 2012-2014 DVs are shown in Figures 5 and 6. The top two panels of Figure 5 show the upper and lower limits of the 25%, 50%, 75%, and 95% CIs for the median as well as the mean, median, minimum and maximum DVs calculated from the 20,000 bootstrap samples as a function of the DV determined from the original dataset. Variability is greater for the 24-hr $PM_{2.5}$ NAAQS than the annual $PM_{2.5}$ NAAQS. This is not a surprising result since the mean would be expected to be a more stable statistic than the 98th percentile. Since the $PM_{2.5}$ data distributions tend to be skewed to the right (see examples in the appendix), the presence of a few very high concentration values, or "outliers", in the original dataset for a year would tend to increase the variability associated with any metric based on the highest concentrations (*e.g.*, if the 50th percentile value were determined, it would likely have much less variability than the 98th percentile). The mean and median of the bootstrap DVs for the annual NAAQS almost perfectly replicate the actual DV from the original site data. While some deviations of the mean and median bootstrap DVs from the actual 24-hr NAAQS DV are evident, there are only a few sites where the mean and median bootstrap DVs deviate substantially from the actual DV.

The relative variability (*i.e.*, the difference between the bootstrapping CI value and the actual design value for a single monitoring site, divided by the actual design value for the site) is also shown in Figure 5, with distributions of the relative differences for each CI across monitoring sites shown in Figure 6. Viewing the results on a relative scale allows the display of finer details of the deviations between the bootstrap results and the actual DVs. The relative variability shows that for the annual NAAQS there are relatively small differences in the values corresponding to the 25%, 50%, and 75% CIs compared to the difference between these and the 95% CI. For the 24-hr NAAQS, the values corresponding to the 25% and 50% CIs are fairly close to each other, with greater differences between these and the 75% and 95% CIs. The relative variability shows an important feature: that from a relative sense, the air quality variability is fairly stable as the baseline air quality worsens especially for the 25% and 50% CIs. That is, there is no notable increase in the relative variability of the bootstrap DV as the actual DV increases (While there is no discernible trend with increasing DV for the 75% and 95% CI they display much more variability or noise in the relative CIs). This is important because it indicates that the magnitude of the actual DV does not systematically affect the relative variability in the bootstrap DV at a site and because it indicates that a central tendency value for the relative variability in the DV. Therefore a representative value can be multiplied by the level of that NAAQS to obtain a value in concentration units (μg/m^3 for $PM_{2.5}$) that is appropriately used to characterize variability for sites with air quality that "just complies" with that NAAQS.

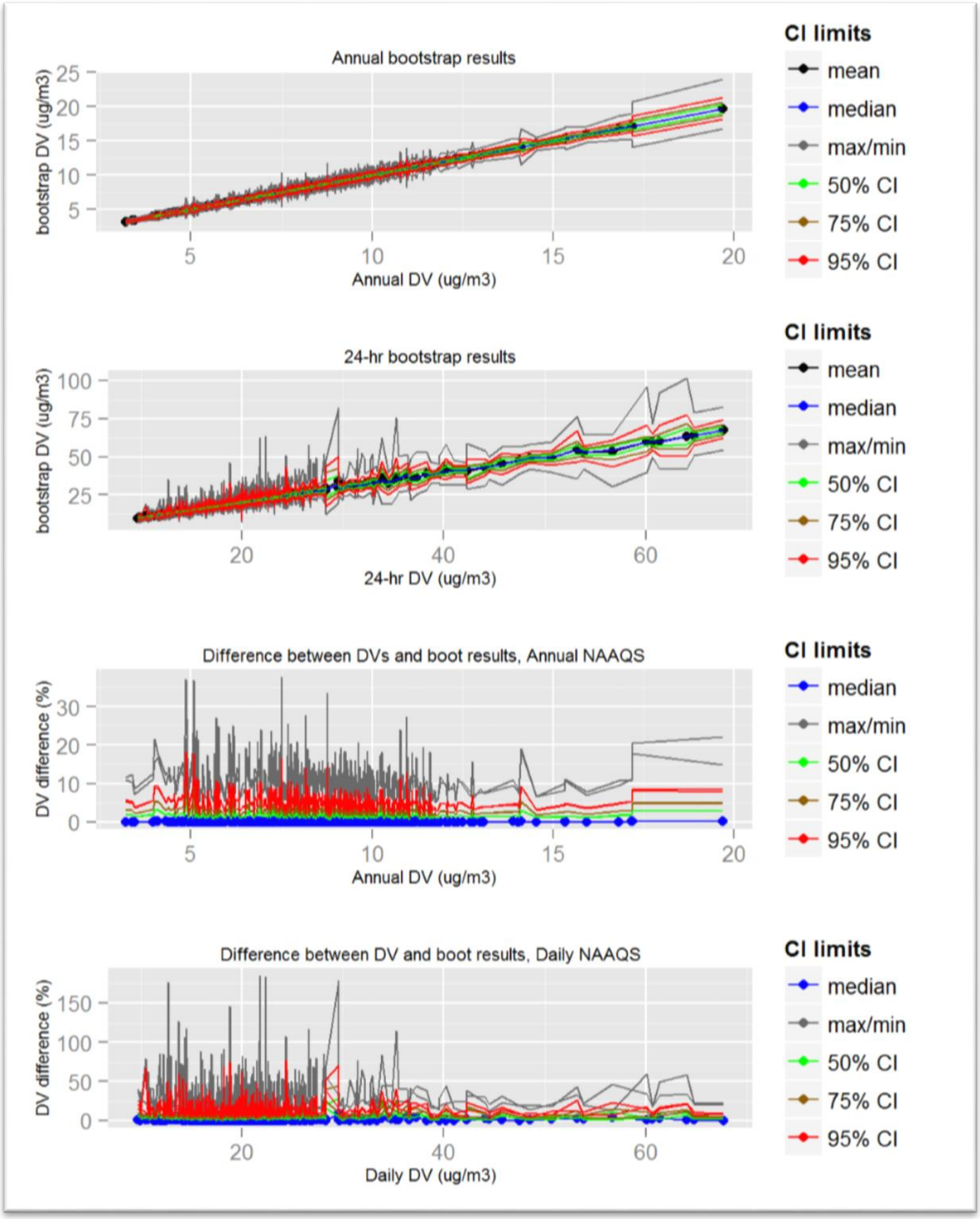


Figure 5 - Bootstrap results for the $PM_{2.5}$ 2012-2014 DVs (25%, 50%, 75%, and 95% CIs, along with the mean and median bootstrap DVs). The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV (defined as abs[CI DV - actual DV]/[actual DV]).

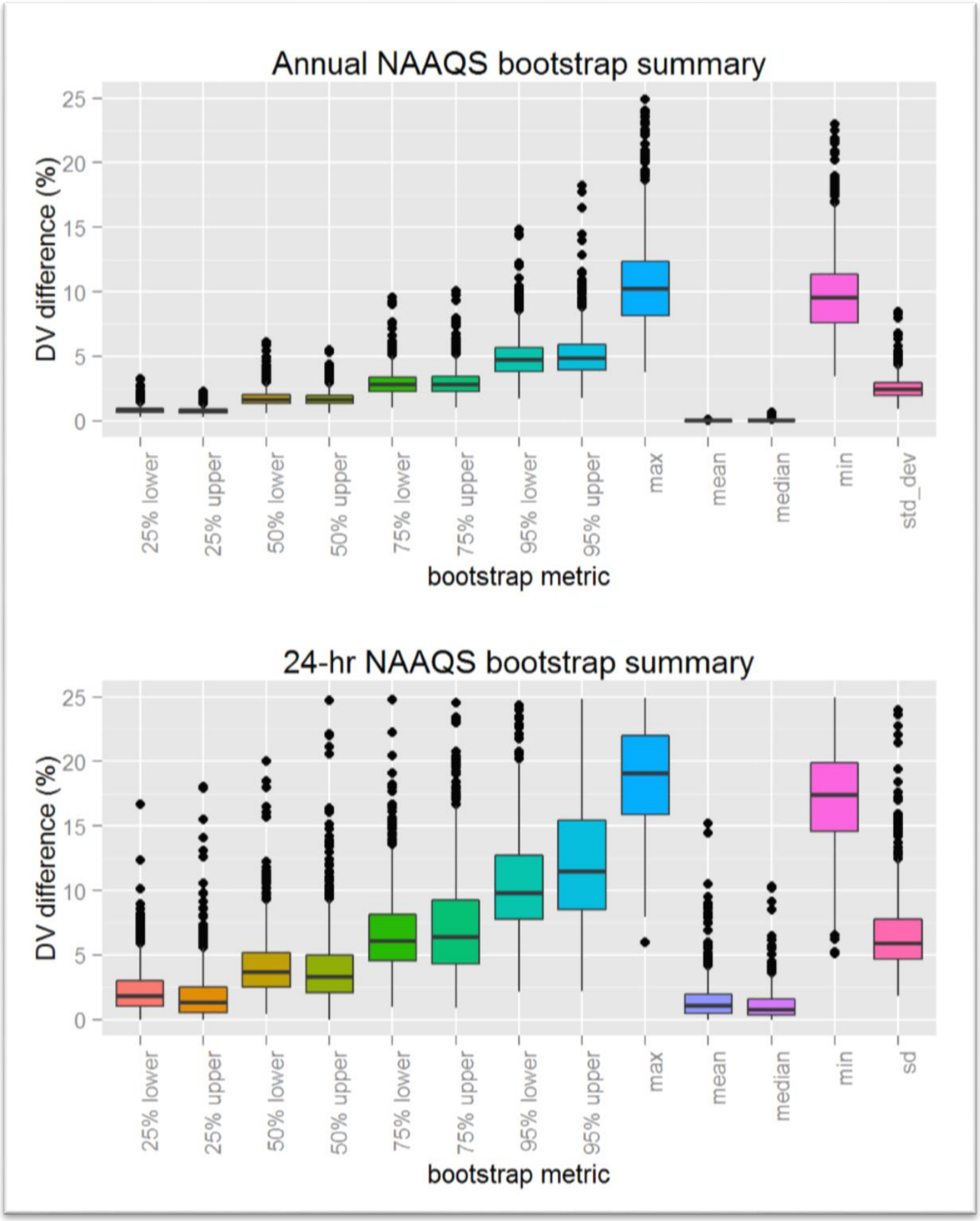


Figure 6 - Bootstrap results for the $PM_{2.5}$ 2012-2014 DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 25%, 50%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

3.2.1 Analysis of $PM_{2.5}$ Spatial Variability

Section 2.1.3 discussed the design of the monitoring network and the spatial scales associated with each monitor. While there may be changes to the area around a monitor after the scale was determined when the monitor was sited, the monitor scale should be somewhat reflective of air quality within the area indicated. This basic need for multiple monitor scales and multiple monitors in an area to assess an area's air quality is due to the fact that there is an inherent spatial variability of air quality. For example, due to the inherent variability in the location of emission sources and changes in meteorological patterns, two "urban scale" monitors located a few blocks from each other would likely record different daily values, resulting in slightly different DVs. The analysis conducted here seeks to quantify that spatial variability by identifying pairs of monitors that are located in proximity to one another to determine the relative difference between the two monitors, as indicated by the DVs. The differences between the DVs are interpreted as a measure of the spatial variability in the area and provide a benchmark to evaluate the variability determined from the Bootstrap analysis.

The analysis was conducted using the 2012-2014 annual and 24-hr $PM_{2.5}$ DVs and focused on pairs of monitors which collected $PM_{2.5}$ samples every day (1:1 monitors) in order to reduce the impact of temporal variability (see Section 4.3.1 for an analysis of the temporal variability). A total of 70 1:1 monitors were identified that were separated by a distance of less than 50 km, with 13 less than 10 km apart. We did not investigate whether -- based on emission sources, winds, and terrain -- any of these sites could reasonably be considered representative for particular locations at which a new source could seek a permit in the future.

The results from the analysis are summarized in Table 1 (monitor pairs within 10 km) and in Figures 7, 8 and 9 (monitor pairs within 50 km). There is a strong correlation between the DVs in the site pairs (top panels in Figure 7), with a slope of 0.8 (r^2 of 0.51) between monitor pairs less than 50 km apart for the annual NAAQS and a slope of 0.87 (r^2 of 0.59) for the 24-hr NAAQS. There are no obvious trends in the differences between the monitors, either the absolute differences or the relative differences (defined as the absolute difference between the DVs from the two monitors divided by the average DV). The relative differences range from 0% to 66%, with a median relative difference of 9% for the annual DVs. For the 24-hr DVs, the relative differences range from 0% to 67%, with a median relative difference of 6%. When the subset of monitors within 10 km are considered, the slope between paired monitors is similar for the annual NAAQS, though the r^2 increases to 0.82, while the slope for the 24-hr NAAQS increases to 0.97 and the r^2 increases to 0.94. For this subset, the maximum relative differences drop to 23% and 16% for the annual and 24-hr DVs, respectively, and the median relative differences drop to 5% and 4% respectively.

These results are interesting and seem to somewhat contradict the results from the bootstrap analysis. This comparison suggests that there is more spatial variability associated with the annual NAAQS, while the bootstrap results show that there is less variability in the annual NAAQS. Conversely, this comparison suggests that there is less spatial variability associated with the 24-hr NAAQS, while the bootstrap results show that there is more variability in the 24-hr NAAQS. Despite this apparent contradiction, these results make sense in the context of secondary pollutants, particularly $PM_{2.5}$. In general, the highest concentrations associated with pollutants that have a substantial portion due to secondary formation occur in widespread "events". These events are an important aspect of the air quality in an area and are associated with unique meteorological conditions, which can either transport air from polluted upwind regions, increasing the background concentrations, or trap local pollutants and facilitate in-situ production. Events are also associated with unique emissions episodes, such as dust storms or biomass burning events that emit large quantities of primary and precursor pollutants. Because of the nature of $PM_{2.5}$ events, there would tend to be a stronger correlation of the higher concentrations across larger spatial scales. The average air quality (annual NAAQS), on the other hand, would not be as heavily impacted by the unique (and wide-spread events) and instead be more heavily affected by local emissions and production. As such, the prevailing meteorological conditions and the prevalent local emission sources would have the most impact on the annual DVs. In this case, localized differences in emissions could cause monitors to have greater differences in the annual DVs than is seen at a number of site pairs.

The result from the spatial variability analysis also suggests an important link to temporal variability. The occurrence of these transport and emissions events is infrequent with varying intensity, such that they may not occur in every year and their frequency and duration would vary. Even when these events do occur, the intensity and impact on regional and local air quality would vary and also be difficult to predict. Since the bootstrap results show that 24-hr NAAQS has the most variability, this seems to imply that temporal variability is the most important component of the 24-hr NAAQS variability, while the spatial variability may be the most important component of the annual NAAQS variability.

Table 1 - Summary of results from $PM_{2.5}$ spatial variability analysis for monitor pairs within 10 km of one another.

State	City	Dist (km)	Monitor 1 ID	Annual DV 1 (µg/m^3)	Monitor 2 ID	Annual DV 2 (µg/m^3)	Delta (%)[83]
Minnesota	Washington	1.0	271630447	8.1	271630448	8.8	8%
Hawaii	Honolulu	1.7	150031001	4.9	150031004	5.6	14%
Pennsylvania	Philadelphia	2.6	421010047	10.3	421010057	10.9	5%
Pennsylvania	Philadelphia	3.1	421010055	11.6	421010047	10.3	12%
Louisiana	East Baton Rouge	5.4	220330009	9.0	221210001	9.2	3%
Nevada	Washoe	5.5	320310016	7.9	320311005	10.0	23%
Pennsylvania	Northampton	5.7	420950025	10.5	420950027	10.1	4%
Rhode Island	Providence	5.9	440070022	7.1	440071010	7.4	3%
Iowa	Clinton	6.4	190450019	10.6	190450021	9.4	11%
Utah	Salt Lake	7.3	490353006	9.2	490353010	9.7	5%
New Mexico	Bernalillo	7.9	350010023	6.5	350010024	6.3	3%
Indiana	Marion	8.9	180970078	11.1	180970081	11.8	6%
Indiana	Clark	9.3	180190006	11.8	211110067	11.3	4%
State	**City**	**Dist (km)**	**Monitor 1 ID**	**24-hr DV 1 (µg/m^3)**	**Monitor 2 ID**	**24-hr DV 2 (µg/m3)**	**Delta (%)[83]**
Minnesota	Washington	1.0	271630447	20.6	271630448	21.1	3%
Hawaii	Honolulu	1.7	150031001	10.9	150031004	11.4	5%
Pennsylvania	Philadelphia	2.6	421010047	24.3	421010057	25.2	4%
Pennsylvania	Philadelphia	3.1	421010055	26.4	421010047	24.3	8%
Louisiana	East Baton Rouge	5.4	220330009	19.7	221210001	19.4	2%
Nevada	Washoe	5.5	320310016	26.8	320311005	31.5	16%
Pennsylvania	Northampton	5.7	420950025	27.2	420950027	28.3	4%
Rhode Island	Providence	5.9	440070022	18.3	440071010	18.6	2%
Iowa	Clinton	6.4	190450019	24.7	190450021	22.8	8%
Utah	Salt Lake	7.3	490353006	42.3	490353010	41.0	3%
New Mexico	Bernalillo	7.9	350010023	15.4	350010024	15.1	2%
Indiana	Marion	8.9	180970078	25.0	180970081	26.4	5%
Indiana	Clark	9.3	180190006	24.2	211110067	22.8	6%

[83] Defined as the difference between the two monitored DVs divided by the mean DV of the two monitors.

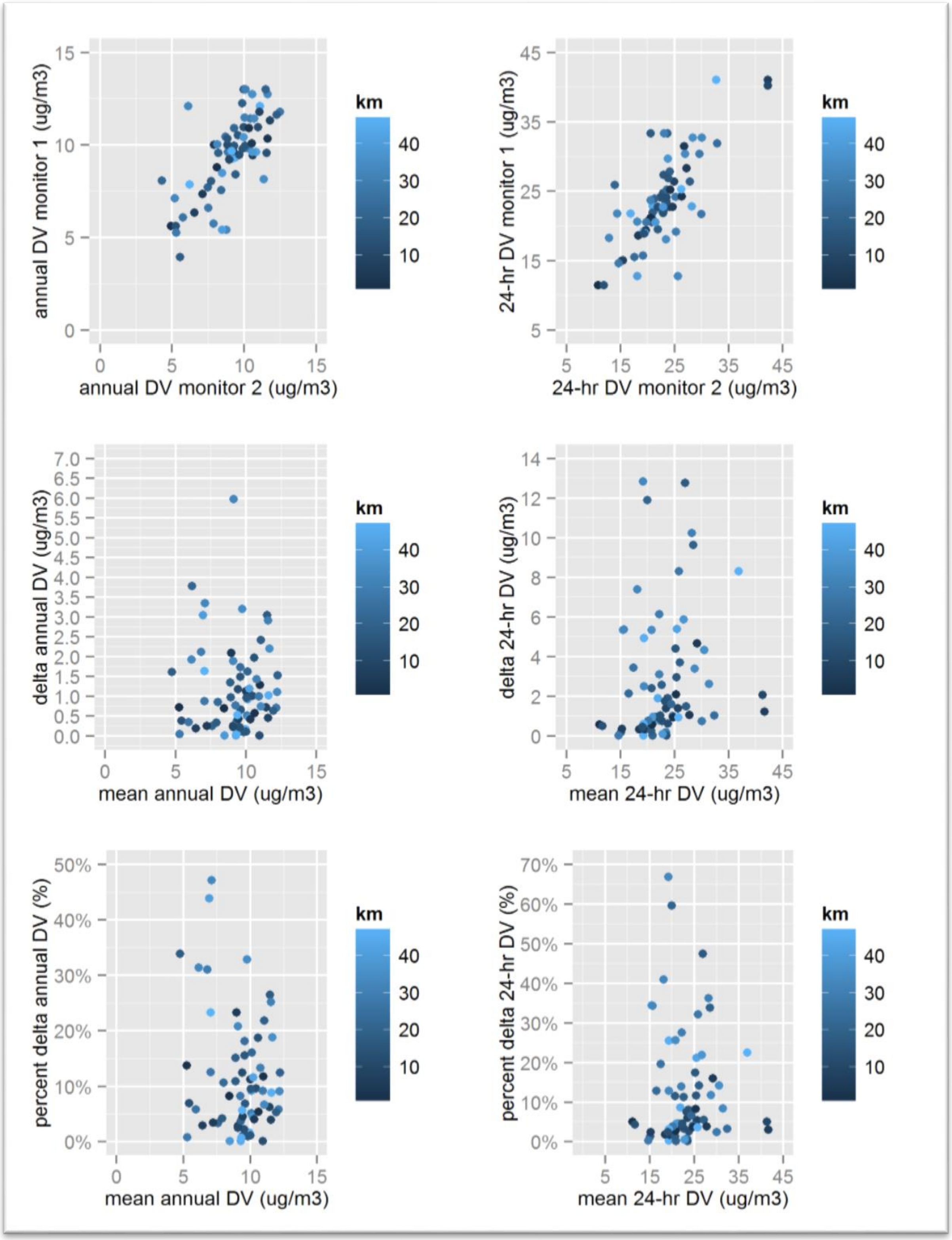


Figure 7 - Results from the analysis of spatial variability. Left column shows results for annual $PM_{2.5}$ NAAQS and the right column shows the results for the 24-hr $PM_{2.5}$ NAAQS.

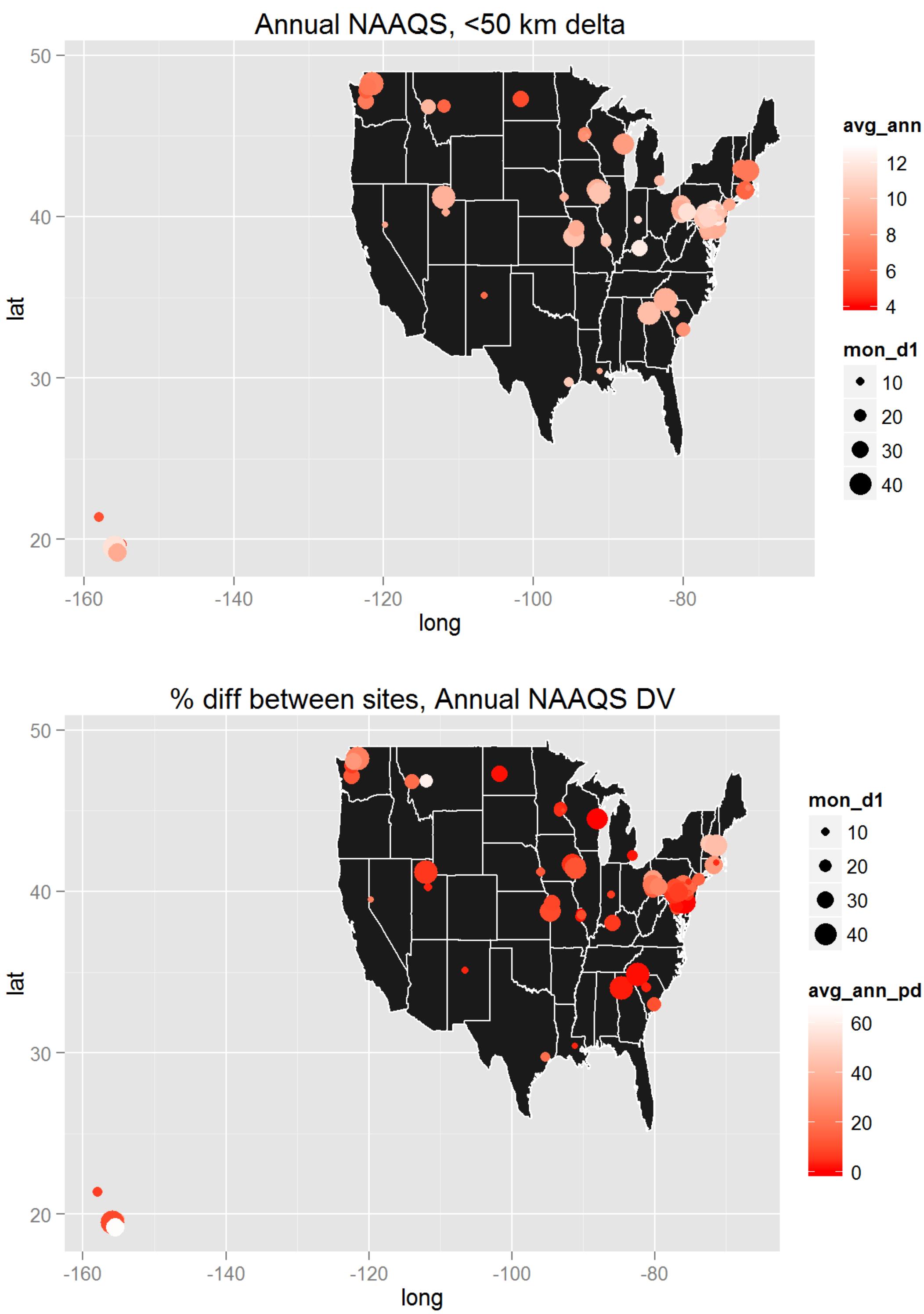
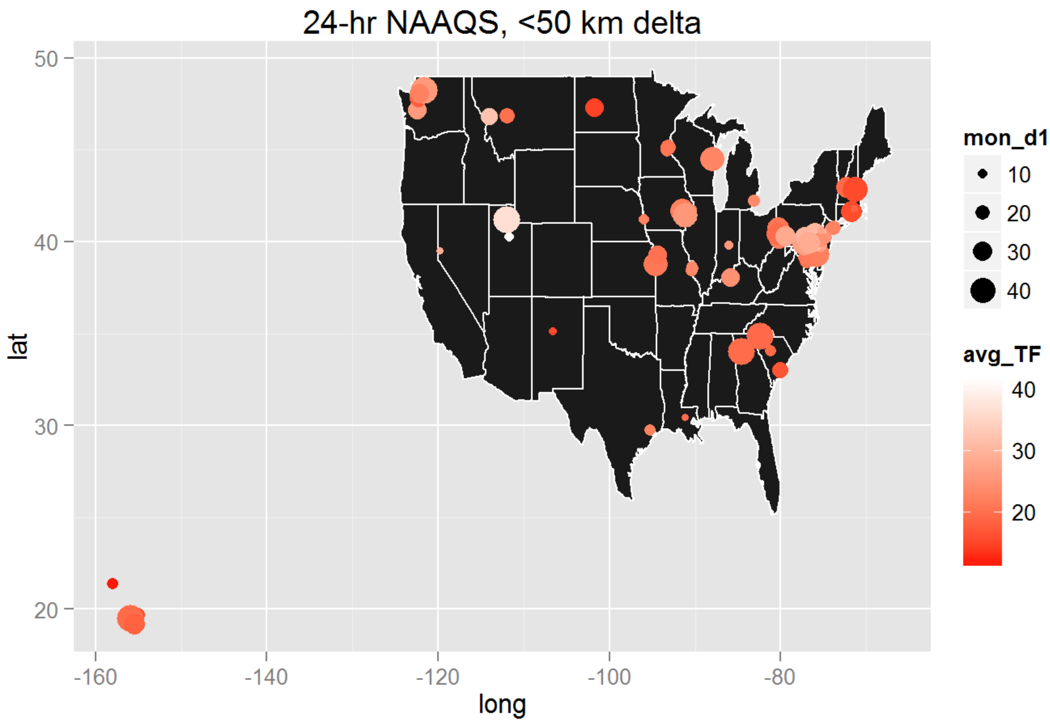


Figure 8 -- Spatial distribution of the difference between the DVs from spatial analysis of the 2012-2014 $PM_{2.5}$ annual DVs. Top panel shows the absolute value of the difference between

the two monitors while the bottom panel shows the difference divided by the mean between the DVs from the two monitors.



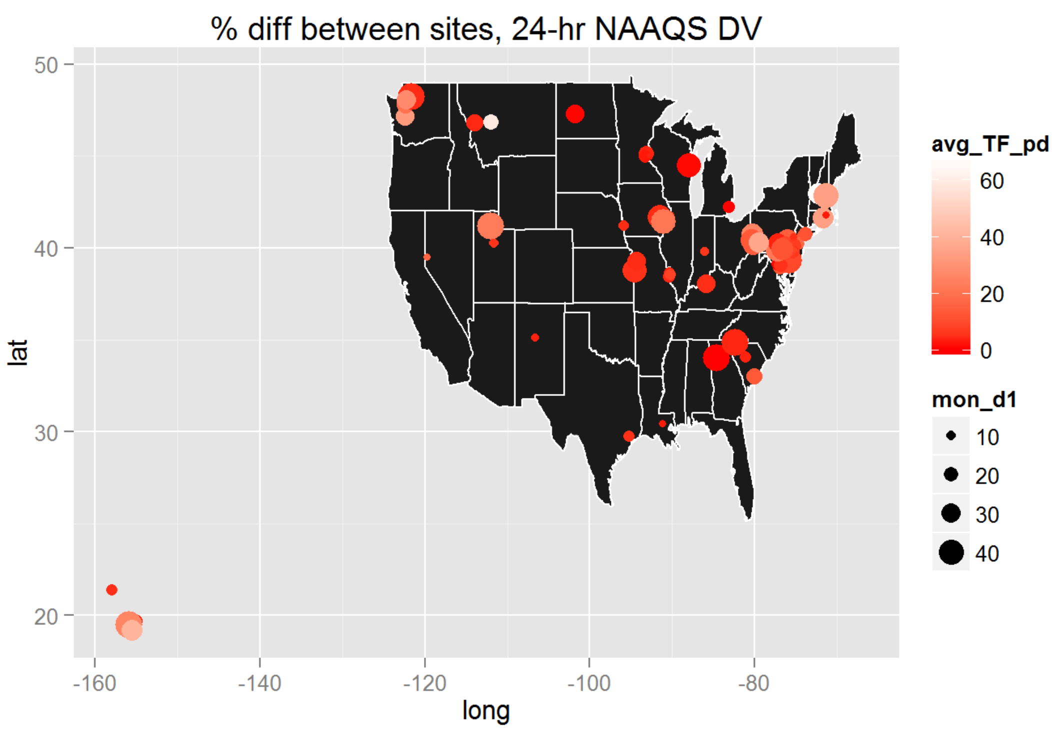


Figure 9 -- Spatial distribution of the difference between the DVs from spatial analysis of the 2012-2014 $PM_{2.5}$ 24-hr DVs. Top panel shows the absolute value of the difference between the

two monitors while the bottom panel shows the difference divided by the mean between the DVs from the two monitors.

3.2.2 Analysis of the Influence of $PM_{2.5}$ Monitor Sampling Frequency

The PM monitoring network is unique in that it has not been designed to operate continuously. When initially designed and deployed, the monitoring requirements for PM indicated that many sites only needed to sample on every third or sixth day, with a smaller number required to sample every day. This is partly due to the technology available at the time, which required a person to collect the filter sample and reload the filter cartridge for each sample taken. The filters were then transported to a laboratory for weighting analysis. While much of the $PM_{2.5}$ network still relies on filter-based sampling, systems that can load multiple filters and automatically swap out filters after each 24-hr monitoring period have reduced the labor requirements. Non-filter based measurement techniques have also been developed that allow for continuous operation (as well as 1-hr sampling) so that concentration values are provided for every 24-hr period. Additionally, the requirements for sampling frequency have tightened, requiring more frequent sampling, particularly in areas with design values close to the NAAQS. The result of the technological and regulatory changes is a sampling network with varied sampling frequency, with notable changes in the sampling frequency over time (see Figure 10). The total number of sites in the network has decreased, however the number of 1:1 sites has increased. Many 1:6 and 1:3 sites have been replaced by 1:1 sites, a trend most obvious starting around 2008. (The site classification was based solely on the number of daily samples during the course of the year, *i.e.*, sites with 60 or less samples were 1:6, sites with 121 samples or less but more than 60 were classified as 1:3, and sites with 122 or more samples were classified as 1:1.)

Due to the nature of temporal variability, it would generally be expected that statistics from datasets from sites with less frequent sampling would in general have a higher variability. Sensitivity tests conducted with the 2010-2013 DVs indeed showed that statistics from the subset of sites with daily monitoring (1:1) have less variability than the subset of sites with 1:3 monitoring and all data (which includes 1:6 monitors) (see Table 2). However, since the 1:1 monitors are not sampling the same air as the 1:3 monitors, it is difficult to directly compare the results from these subsets as a definitive indicator of the inherent increase in variability due to less frequent sample. However, the results do support what is generally expected from reduced sampling frequency, *i.e.*, increased variability in statistics from the air quality measurement data.

Since the monitor sampling frequency can have a notable impact on the calculated air quality variability, an important question arises regarding which monitors should be used. Using only the 1:1 monitors would produce smaller estimates of the variability. However, the 1:3 and 1:6 monitors are part of the monitoring network and will continue to be present for the foreseeable future. Additionally, despite an increase in the number of 1:1 monitors, the overall air quality variability indicated by the network has been fairly stable for the annual and 24-hr $PM_{2.5}$ NAAQS (see Section 4.3.1). This suggests that the inherent variability in the air quality is more influential than the increased variability induced by the 1:3 and relatively small number of 1:6 monitors. In addition, the much greater number of monitoring sites available when sites with all schedules are considered (see Table 2) provides more confidence that the results are representative of the U.S. as a whole.

Table 2 - Summary of comparison of the air quality variability determined by the bootstrap analysis for three design periods for monitors with different sampling frequencies.

Monitor class	all	1 in 1	1 in 3	all	1 in 1	1 in 3	all	1 in 1	1 in 3
Year/NAAQS	**2014 annual**			**2013 annual**			**2012 annual**		
Difference, median bootstrap vs actual	0.03%	0.02%	0.03%	0.03%	0.01%	0.03%	0.03%	0.02%	0.03%
Avg. 25% CI span	0.78%	0.54%	0.87%	0.80%	0.54%	0.88%	0.82%	0.55%	0.90%
Avg. 50% CI span	1.65%	1.14%	1.83%	1.70%	1.15%	1.85%	1.74%	1.17%	1.91%
Avg. 75% CI span	2.83%	1.95%	3.14%	2.90%	1.96%	3.17%	2.96%	1.99%	3.25%
Avg. 95% CI span	4.81%	3.31%	5.32%	4.95%	3.33%	5.39%	5.05%	3.40%	5.57%
Year/NAAQS	**2014 24-hr**			**2013 24-hr**			**2012 24-hr**		
Difference, median bootstrap vs actual	0.79%	0.55%	1.04%	0.89%	0.65%	1.01%	0.84%	0.56%	0.92%
Avg. 25% CI span	1.78%	1.43%	2.01%	1.79%	1.41%	1.96%	1.79%	1.31%	1.93%
Avg. 50% CI span	3.76%	2.97%	4.18%	3.66%	2.90%	3.96%	3.63%	2.68%	3.90%
Avg. 75% CI span	6.50%	5.14%	7.44%	6.37%	4.90%	6.96%	6.43%	4.77%	6.85%
Avg. 95% CI span	11.41%	8.76%	12.73%	11.10%	8.61%	12.01%	11.30%	8.18%	12.16%
Number of sites	720	242	379	724	227	398	714	196	411

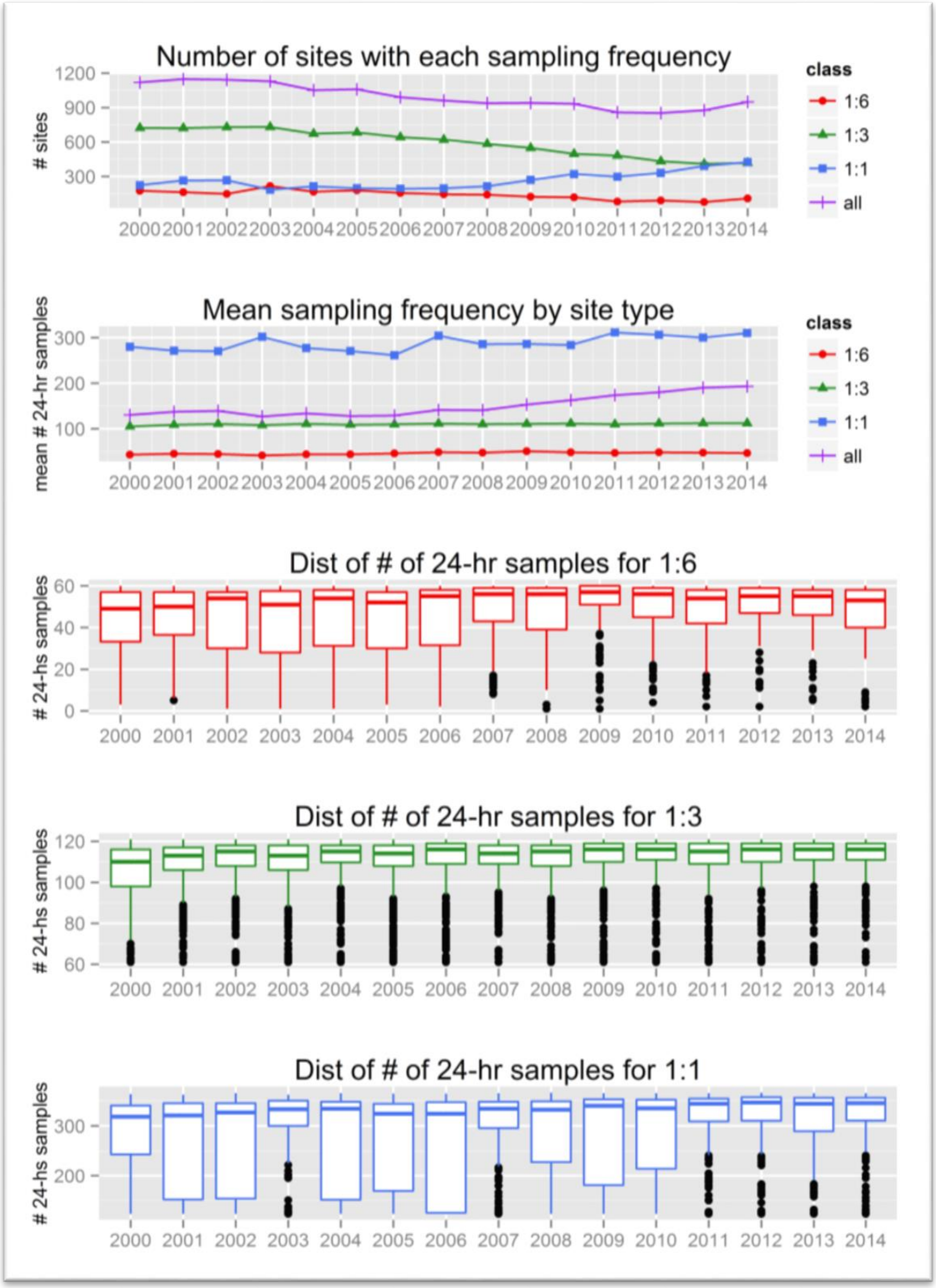


Figure 10 - $PM_{2.5}$ monitor network statistics. Top row shows the number of sites with each sampling frequency by year. Second row shows the average number of samples at each site

type. Third, fourth and fifth rows show the distribution of the number of samples for each site type.

4.0 Application of Air Quality Variability to Determine SILs for the PSD Program

Section 3 presented the results from the bootstrap analysis, which produced variability estimates at the 25%, 50%, 75%, and the 95% CIs for all the AQS data across the U.S. from 2000-2014. In order to identify an appropriate value that represents an "insignificant impact" on change in air quality concentration to apply as a SIL for each NAAQS, these variability estimates need to be narrowed down and summarized into a representative value for each NAAQS considered here. This section presents the choice of a single CI level to reflect an "insignificant impact" for each NAAQS,[84] an analysis of changes to these values over time and the resulting value determined to represent an "insignificant impact" for each NAAQS.

4.1 PSD Air Quality Analyses and Statistical Significance

Since the SIL is part of the EPA's PSD program, there are programmatic and policy considerations that also must be considered when determining a SIL. These policy needs, along with the results from the technical analysis can assist in making decisions in order to narrow down the results from the larger dataset of variability estimates to produce a single SIL value for each NAAQS. In particular, four selections have been made to produce one SIL value for each NAAQS from the collective results described above. These four selections include the specific CI to represent the inherent variability, the approach used to scale local variability to the level of the NAAQS, the geographic extent of each summary value, and the DV year or years from which to use the variability results. While these selections are ultimately matters of policy, they can be informed by the results of the analysis and the underlying statistical interpretation of the results. This section presents the policy and technical aspects of these decisions.

4.1.1 Selection of the 50% Confidence Interval for the SIL

The bootstrap analysis produced estimates for the 25%, 50%, 75%, and the 95% CIs to characterize the range of the measured of variability and to provide multiple options for selection as an appropriate "insignificant impact" level for each SIL. From these estimates, a specific CI needs to be selected to represent the variability that will be applied to determine each SIL. While there are physical and statistical components of the analysis that can guide this selection, the selection is ultimately based on a

[84] Once an appropriate CI is selected, the SIL is calculated by subtracting the lower bound from the upper bound and dividing by 2 (i.e., half the width of the confidence interval).

consideration of both the policy implications and technical aspects of this analysis in the context of the application of the SIL in the PSD program.

The statistical framework for the bootstrap creates CIs which can be related to an assessment of statistical significance. The traditional application seeks to determine if a deviation from the base value is *significant*. In order to make this determination, larger CIs are typically selected, *e.g.*, 95-99%. In practice, the smallest CI that might be considered for this determination would be the 68% CI, which corresponds to values within one standard deviation of the mean for a normally distributed sample. Thus, any deviation outside the bounds of CIs of a level of confidence greater than 68% could traditionally be identified as a *significant* deviation from the mean. In this PSD application, however, we are seeking for each NAAQS a SIL value below which we can conclude that there is only an "insignificant impact", *i.e.*, that there will not be a notable difference in air quality after the new source begins operation. While these selections are ultimately matters of policy, they can be informed by the results of the analysis and the underlying statistical interpretation of the results. This is done using the underpinnings of *statistical insignificance* at a very conservative (low) level. Therefore, a CI smaller than a 68% CI is chosen to ensure that the physical limit selected would identify values that would very likely not be found to have a *statistically significant* effect.

For the purposes of the PSD program, the intent is to develop a SIL that is large enough to allow some predicted impact from sources and account for inherent variability, but keep those impacts from being a "significant impact". Thus, a SIL of zero or a very small SIL does not fit this purpose. From a policy perspective, a 0% CI (which would give a SIL of zero) or a CI of a very low level of confidence (which would provide a very small range and give a very small SIL) would not fulfill the policy intent of a SIL. From a physical and mathematical perspective, the synoptic patterns of meteorology and emission profiles that correspond to the CIs of very low levels of confidence (*i.e.*, close to 0%) are effectively recreating the years as actually sampled, *i.e.*, there is little or no measure of the variability represented by the bootstrap results within CIs of very low confidence. While there are many other levels of confidence for CIs that may also be useful, there is no scientific reason to select any one CIs over another. Based on the factors above as well as both policy and technical considerations, the 50% CI was chosen as the benchmark statistic from the bootstrap analysis as a range of values reflecting a *statistically insignificant* difference from the average bootstrap DV. Air quality changes below the increase (SIL) would thus represent an "insignificant impact" on air quality in the context of each NAAQS. From a physical perspective, this means that the 50% CI represents the bounds of the variability that captures the central 50% of the inherent air quality values. In terms of a DV that exceeds the upper bound of this variability, there is a 25% chance that a DV of this magnitude or higher would intrinsically occur due to the inherent variability in emissions and meteorology.

4.1.2 Adjustment to the Level of the NAAQS

The SIL analysis conducted as part of the PSD air quality demonstration is focused on determining if a "significant ambient impact will occur" from the emissions from a proposed new or modifying major source.[85,86] This impact is used to determine whether the proposed source will “cause or contribute to an air quality violation”. [87] Due to this second clause, the test only applies when the projected air quality is very close to the level of the NAAQS. When the air quality is well below the NAAQS, there will be no violation, so no “cause or contribute” analysis will be necessary. Conversely, when the air quality is well above the NAAQS, the area would not fall under the PSD program and, as such, no modeling demonstration would be conducted for the permitting process. Thus, it is reasonable for the purposes of this analysis to either evaluate the variation in the air quality only when the DVs are very close to the NAAQS or attempt to relate the variation at any air quality level to the variation that would occur when the DVs are very close to the NAAQS. Sections 4.2 and 4.3 present the 50% CI values on both an absolute scale (ug/m3 and ppb) and a relative scale (percentage), where the relative variability is defined as the percent deviation from the base DV at each site. The figures in these sections indicate that there is a trend in the absolute uncertainty, increasing with increasing DVs, but that there is no particular trend in the relative uncertainty at any of the CIs, *i.e.*, the relative variability is not particularly higher or lower at higher or lower baseline DVs (see Figures 11 and 14). These results suggests that there is an inherent aspect to the variability, regardless of the baseline air quality, and that the relative uncertainty can be used to broadly characterize the variability from the data. Thus, the analysis here focuses on determining a characteristic relative uncertainty from all the results, which can be directly applied to any baseline air quality level and scaled up or down to the level of the NAAQS.

4.1.3 Selection of a Single National Value

A fundamental question raised in using air quality variability to inform the selection of a value for a SIL is whether the variability-based SIL value should be based on an analysis of air quality variability at the particular site of the new source or modification, or whether the SIL value should reflect the central tendency of all monitored sites in the U.S., regardless of the new source’s or modification’s planned location. In other words, if one location has more air quality variability than another location, should the SIL value used in the more variable location be higher that the SIL value used in the less variable area? The implication of implementing area-specific SILs could be that that a larger source with a greater

[85] Revision to the Guideline on Air Quality Models: Adoption of a Preferred General Purpose (Flat and Complex Terrain) Dispersion Model and Other Revisions (70 FR 68218) November 9, 2005.

[86] Air Quality Model Guidelines: Enhancements to the AERMOD Dispersion Modeling System and Incorporation of Approaches to Address Ozone and Fine Particulate Matter; Revisions (80 FR 45340) July 29, 2015.

[87] While the "cause or contribute" use of the SIL only applies when there is a projected violation from cumulative modeling, this property also allows the use of the SIL in a single-source analysis. Both of these analyses are described in *Guidance on Significant Impact Levels for Ozone and Fine Particles in the Prevention of Significant Deterioration Permitting Program*, memorandum from Stephen D. Page to the EPA Regional Air Division Directors, [date].

contribution to air quality could be allowed to be permitted in the more variable area while it might not be permitted in the less variable area. Since the NAAQS are set nationally, the historical practice for NAAQS SILs has been to set a single national SIL value per NAAQS.[88] Thus, from a policy perspective, it would be preferable to continue with a national SIL for each PM2.5 NAAQS and to set a single national ozone SIL.

While the policy aspects of this work are indicative of a single national value, the EPA recognizes that the air quality data and the nature of the emissions and chemical formation of PM2.5 and ozone should be considered as part of these decisions. In particular, it is important to determine if there are regional trends in the variability that suggest large areas, rather than single sites, exhibit a higher or lower amount of variability. The analysis presented in sections 4.2 and 4.3 (Figures 11 and 14) examine the relative uncertainty represented by the 50% CI in order to determine if there are spatial trends in the data. The analysis indicates that there are no large scale, *i.e.,* region-to-region, trends in ambient air variability. While there is a fairly consistent range of variability across the U.S., within a state or region, the magnitude of the variability differs from site-to-site within a state or region. Developing a SIL for each monitor would create a situation where it would be difficult to determine how the impact of proposed new source or modification should be compared to a particular available air monitoring site (*i.e.*, which local or more distant monitor should be used for the local SIL value). A national SIL value provides a SIL value for any location and eliminates the need to determine local or regional approaches for developing a SIL. Additionally, because of the adjustment of each monitor's variability to the level of the NAAQS, even if a site-specific approach were used (see the discussion of adjustment to the level of the NAAQS above), it would not be the site's actual variability that was reflected in the SIL. This fact further supports the usage of a national value over regional or site-specific values. It can also be noted that the approach of having one SIL value across the U.S. for a NAAQS would provide a level playing field among areas in which a new source might consider locating, such that there would be no incentive to "shop" for a higher SIL value.

Based on these considerations, the EPA has determined that aggregating the air quality to a national level is appropriate. Such aggregation is frequently done by averaging across sites in different areas. However, in this case the results also show that the small number of sites with particularly high variability have an effect on the average network-wide variability. Instead of averaging across all sites, the use of the median network-wide variability ensures that the SIL value represents the central tendency of the monitoring sites and is not overly influenced by a few outliers and thus produces a more conservative (more protective of air quality) estimate of the network variability. Therefore, using the median variability from the 50% CIs from the entire U.S. ambient monitoring network satisfies the policy needs for a SIL and is congruent with physical and chemical processes that result in this variability.

[88] The now-removed SILs for PM2.5 were an exception, in that the SIL level applied for the NAAQS depended on the "classification" of the affected area under the visibility protection program. However, this feature was tied to the use of the same set of SIL values for purposes of protecting PSD increments, which vary depending on the same classification.

4.1.4 Selection of the Three Most Recent Design Value Years

Sections 4.2.1 and 4.3.1 present trends in the median nation-wide variability at the 50% CI from 2000-2014 (equivalent to DV years of 2002-2014). For all three NAAQS considered here, there are general downward trends across these years. Since the SILs should reflect the most representative state of the atmosphere, the value selected here for each NAAQS should reflect the lower variability observed in the more recent periods, rather than all the data since 2000. However, it may be advantageous to avoid relying on a single 3-year period that may have been influenced by unusual circumstances, particularly in light of the slightly different trends in the last several years across pollutants (*i.e.,* most recently the 24-hr PM2.5 NAAQS has increased, while the annual PM2.5 and ozone NAAQS has continued to decrease). Faced with a similar selection of DV periods for use in attainment demonstrations for nonattainment areas,[89] the EPA also recommended using the average of three DV periods to be used along with a modeling analyses. Thus, as a matter of policy, we have adopted this approach here and determined to use the average variability from the three most recent DV periods (*i.e.*, 2010-2012, 2011-2013, 2012-2014) for determining SILs for PM2.5 and ozone.

4.2 SIL Values for Ozone

Figure 11 shows, for each monitoring site, the half-width of the 50% CI divided by the actual design value, from the 2012-2014 data for the ozone NAAQS.[90] The scatter plot for the relative variability values shows that the data are fairly well concentrated around 1-2%, with a small number of sites exceeding 3% and a maximum around 4.5%. While there are only a few outliers, they occur across the range of baseline air quality levels, indicating that there is no particular trend with actual design value in the occurrence of sites with especially high variability. When assessed as a whole, despite their relatively infrequent occurrence, these outliers do tend to increase the average of the variability estimates from all the monitoring sites. The median variability, however, is less influenced by these outliers and appears to be more representative of the central tendency of the distribution of relative variability values than the average. Since the median is smaller than the average, it is also a more conservative measure (more protective of air quality) of the air quality variability. The median relative variability can be applied to the level of the NAAQS to determine an ozone concentration for use in air quality modeling demonstrations.

The spatial distribution of the relative variability from the 50% CI is also shown in Figure 11, with 2012-2014 DV period site data colored according to their relative uncertainties and sites with insufficient data during this period in gray. There appears to be no notable large-scale spatial trends in highest relative

---

[89] Draft Modeling Guidance for Demonstrating Attainment of Air Quality Goals for Ozone, PM2s, and Regional Haze. R. Wayland, AQAD, Dec. 3, 2014.

[90] The plots for ozone show a distinct banding in the results. This is a feature of the truncation conventions that were applied to the AQS data prior to the air quality variability analysis.

variability. The lack of any large-scale spatial trend indicates that there is indeed a fundamental characteristic to the relative ambient air quality variability (see section 4.1.3).

4.2.1 Ozone Temporal Trends

The median air quality variability from the 13 DV periods for ozone is shown in Figure 12. This analysis shows how the combination of changes in the network design (*e.g.*, the change in the monitoring season) and the changes in emissions and meteorology over this period have impacted the variability in the DVs from the network. There has been a small decrease in the variability for ozone (0.03 percentage points per year), though most of that decrease occurred in the form of a large drop in the variability between the 2003-2005 and 2004-2006 DV periods. There were increases in the variability for the 2008 and 2012 DV periods, indicating that there is some variability between years. The median air quality variability values at the 50% CI for the three most recent DV periods (*i.e.*, 2010-2012, 2011-2013, 2012-2014) as shown in Table 3, when averaged result in a SIL value for the ozone 8-hour NAAQS of 1.42%. This corresponds to 1.0 ppb at the level of the NAAQS (70 ppb).

Table 3 - Summary of ozone bootstrap results for three design periods, 2010-2012, 2011-2013, and 2012-2014

Year/NAAQS	2014 annual	2013 annual	2012 annual
Difference, median bootstrap vs actual	0.47%	0.46%	0.45%
Avg. 25% CI span	0.70%	0.71%	0.72%
Avg. 50% CI span	1.39%	1.43%	1.45%
Avg. 75% CI span	2.44%	2.52%	2.56%
Avg. 95% CI span	4.28%	4.44%	4.52%
Number of sites	1136	1124	1107

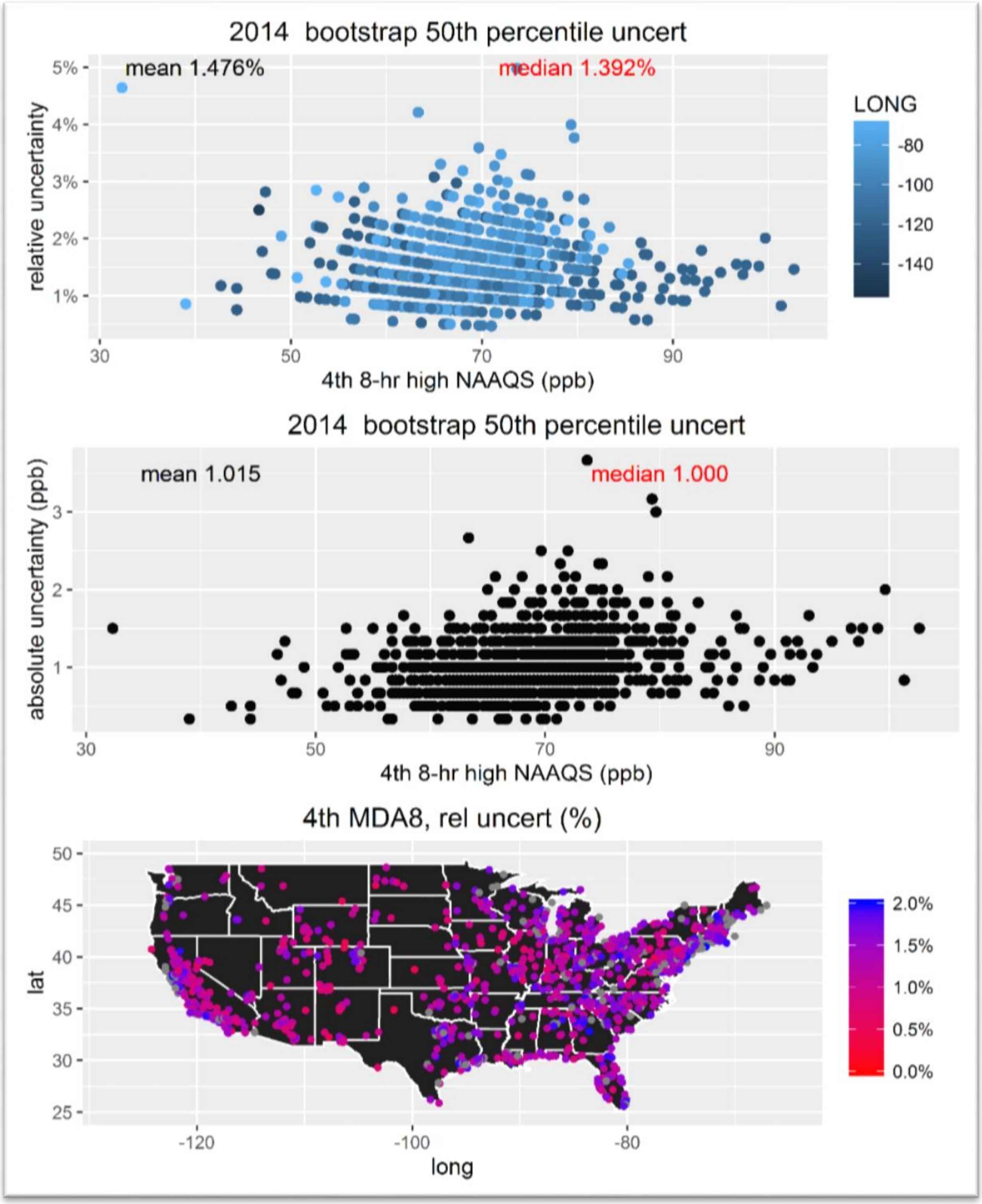


Figure 11 - Bootstrap results from the 50% CIs for the 2014 $PM_{2.5}$ DVs. The top panel shows the relative difference between the CI and the actual DV across the range of actual DVs, the middle panel shows the absolute difference between the values across the same range, and the

bottom panel shows the spatial distribution of the relative difference between the values at each site.

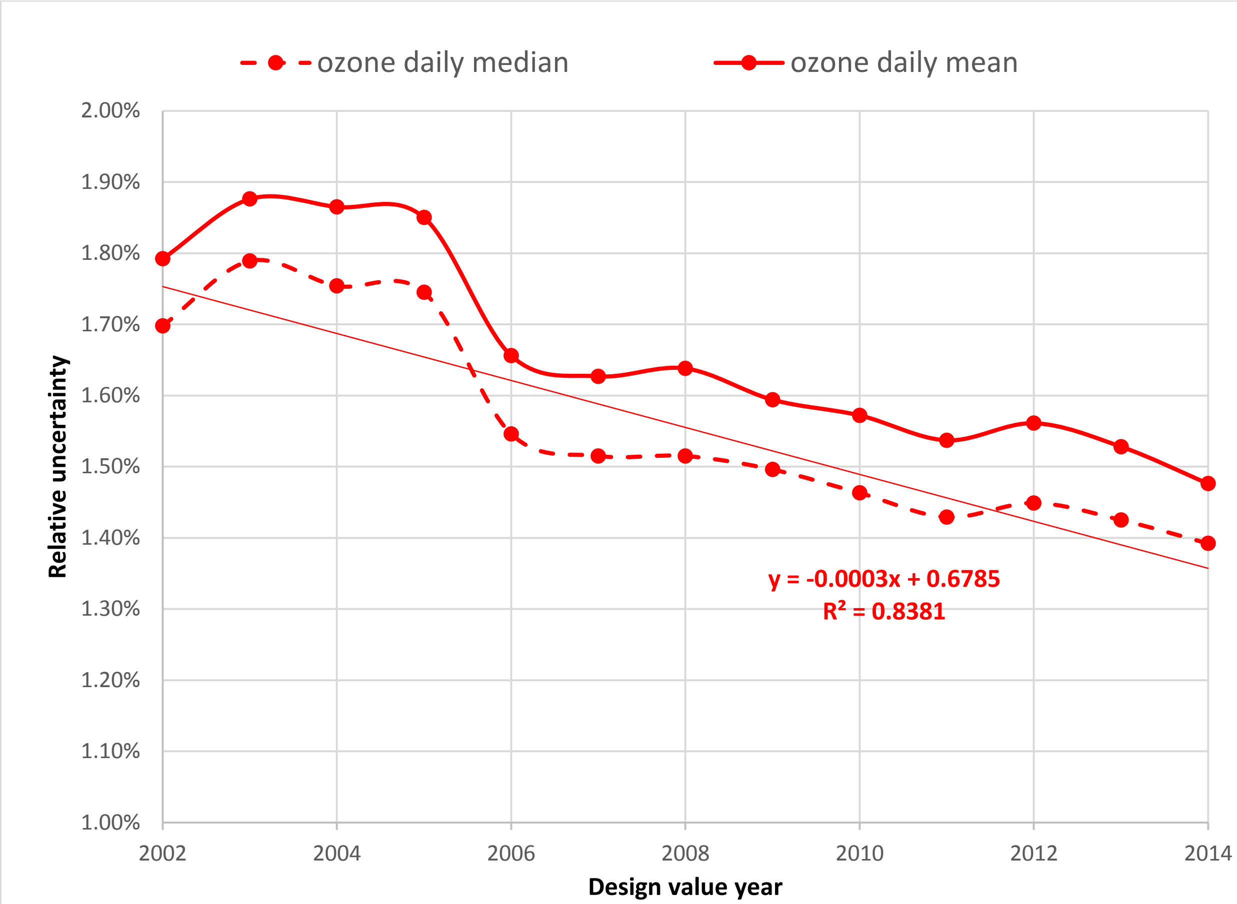


Figure 12 – Median and mean variability in the network determined from the bootstrap analysis for the 13 DV periods from 2002-2014 for ozone.

4.3 SIL Values for $PM_{2.5}$

Figure 13 shows, for each monitoring site, the half-width of the 50% CI divided by the actual design value, for both the annual and 24-hr $PM_{2.5}$ NAAQS. This figure shows even more clearly than Figure 5 that the relative variability is indeed stable across the range of baseline air quality levels, while the absolute variability increases as the baseline air quality levels increase. The scatter plot of the relative variability shows that the values for relative variability are fairly well concentrated around 1-2% for the annual NAAQS, with a small number of sites exceeding 3% and a maximum slightly less than 6%. For the 24-hr NAAQS, the data are concentrated around 4-5%, with a small number of sites exceeding 7% and a maximum around 33%. The outliers occur across the range of baseline air quality levels, indicating that there is no particular trend with actual design value in the occurrence of sites with especially high variability. When assessed as a whole, despite their relatively infrequent occurrence, these outliers do tend to increase the average variability. As with ozone, the median variability is less influenced by these outliers and appears to be more representative of the central tendency of the distribution of relative variability values than the average. As with ozone, the median is smaller than the average, it is also a more conservative measure (more protective of air quality) of the air quality variability and will be used as the initial benchmark to determine the relative variability across the AQS network for the value of the SILs.

The spatial distribution of the relative variability from Figure 13 is shown in Figure 14, with sites having data during the 2012-2014 DV period colored according to their relative variability and sites with insufficient data during the 2012-2014 DV period colored gray. Based solely a visual inspection, there appears to be no notable large-scale spatial trends in geographic locations highest relative variability in either the annual or 24-hr $PM_{2.5}$ NAAQS. The sites with larger variability tend to occur in the western half of the U.S., though the sites are isolated and not grouped into any specific geographic region. This result may be related to the nature of high PM events in the western half of the U.S. (*e.g.*, the typical $PM_{2.5}$ levels may be lower in the western states, but the events that do occur produce much higher concentrations than the typical background, which would result in greater skew and thus greater uncertainty in DVs computed from these data, particularly in the 24-hr $PM_{2.5}$ DVs). There are also trends in missing data. In particular, for the period 2008 through 2013, the data were invalidated for the states of TN, IL, and FL. Late in 2014, a problem was found with the $PM_{2.5}$ data from these states and, as a result, the data were invalidated for the period of 2008 through 2013 (the analysis summarized here uses the most recently validated dataset).[91] The lack of any spatial trend indicates that there is indeed a fundamental characteristic to the relative ambient air quality variability (see section 4.1.2).

[91] The dates and specific monitors affected in each state vary. For FL and IL, data was invalidated from 2011-2013. For KY, data was invalidated from 2009-2012. For TN, data was invalidated from 2008-2012. The invalidation may not have affected every monitor in each state, but these dates cover the time spans for which the data invalidation occurred.

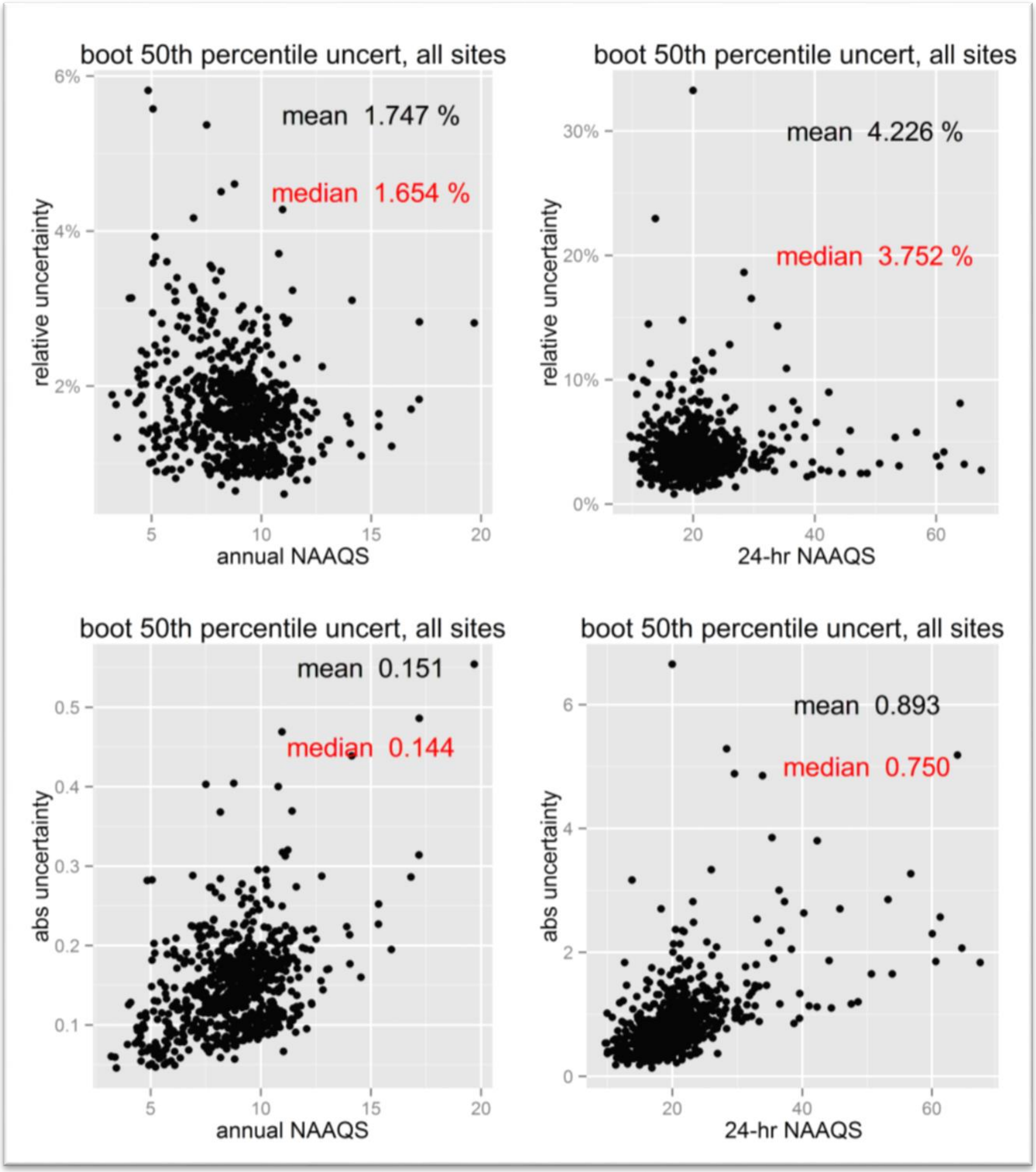


Figure 13 - Bootstrap results from the 50% CIs for the 2014 $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV across the range of actual DV, and the bottom two panels show the absolute difference between the values across the same range.

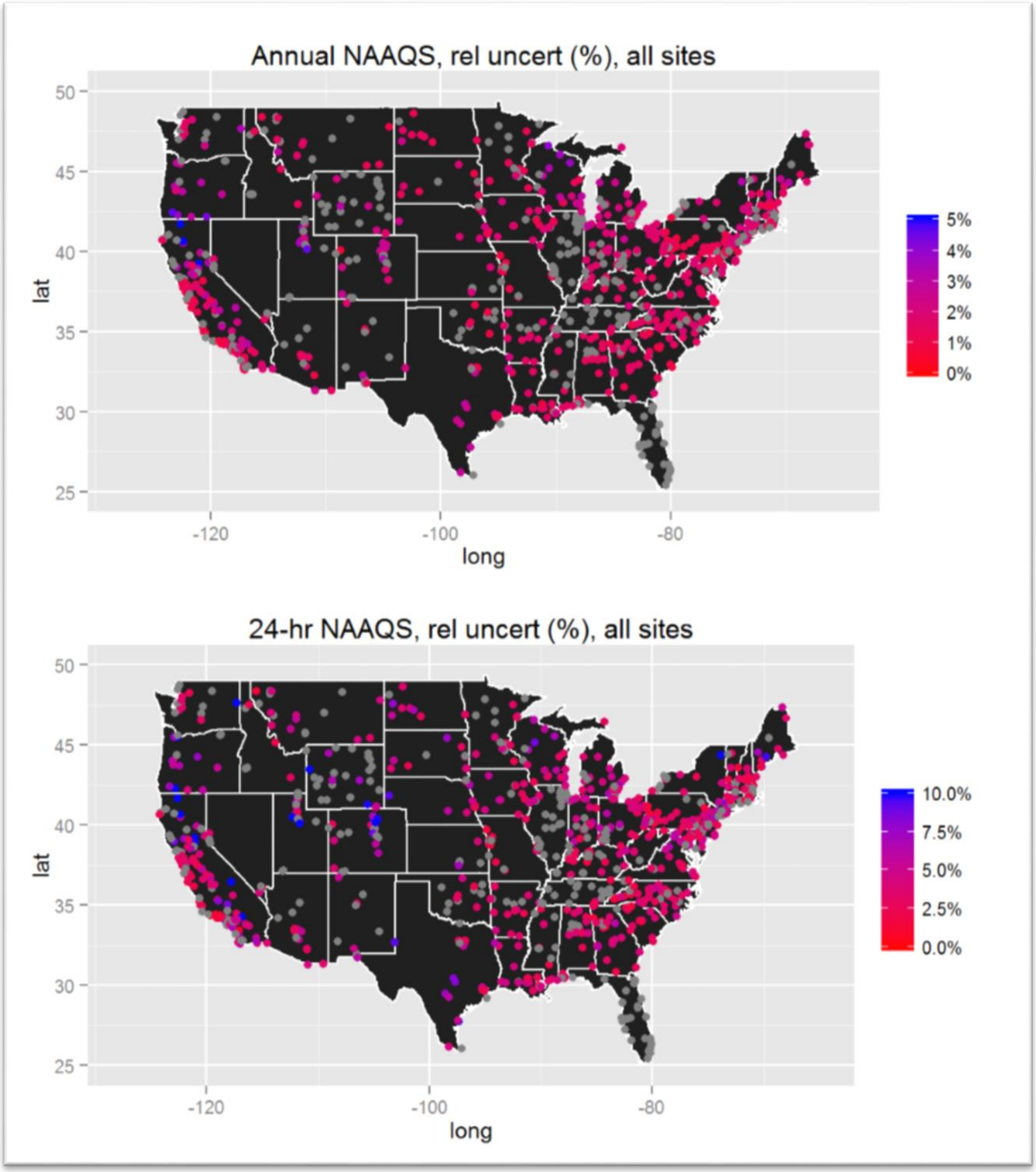


Figure 14 - Spatial distribution of the relative difference between the 50% CI and the actual DV for the 2012-2014 $PM_{2.5}$ DVs.

4.3.1 $PM_{2.5}$ Temporal trends

The median air quality variability from the 13 DV periods for both $PM_{2.5}$ and ozone are shown in Figure 15. This analysis shows how the combination of the changes in the network design (*e.g.*, the change in the monitoring frequency) and the changes in emissions and meteorology have impacted the network variability. There has been a greater decrease in the variability in the 24-hr $PM_{2.5}$ NAAQS than in the variability for the annual $PM_{2.5}$ NAAQS (0.07 percentage points per year versus 0.02 percentage points per year). The analysis in Section 3.2.2 showed that the 24-hr NAAQS is more affected by the monitoring frequency than the annual NAAQS, so it is likely that the change in monitoring frequency played some role in the larger decrease in the variability for the 24-hr $PM_{2.5}$ NAAQS. The median air quality variability at the 50% CI for three most recent DV periods (*i.e.*, 2010-2012, 2011-2013, 2012-2014) as shown in Table 4, when averaged result in a SIL value of 1.70% for the annual $PM_{2.5}$ NAAQS (12 $\mu g/m^3$) and 3.68% for the $PM_{2.5}$ 24-hr NAAQS (35 $\mu g/m^3$). These values correspond to 0.2 $\mu g/m^3$ at the level of 12 $\mu g/m^3$ for the annual NAAQS, and 1.3 $\mu g/m^3$ at the level of 35 $\mu g/m^3$ for the NAAQS.

Table 4 - Summary of comparison of the air quality variability determined by the bootstrap analysis for three design periods.

Year/NAAQS	2014 annual	2013 annual	2012 annual
Difference, median bootstrap vs actual	0.03%	0.03%	0.03%
Avg. 25% CI span	0.78%	0.80%	0.82%
Avg. 50% CI span	1.65%	1.70%	1.74%
Avg. 75% CI span	2.83%	2.90%	2.96%
Avg. 95% CI span	4.81%	4.95%	5.05%
Year/NAAQS	**2014 24-hr**	**2013 24-hr**	**2012 24-hr**
Difference, median bootstrap vs actual	0.79%	0.89%	0.84%
Avg. 25% CI span	1.78%	1.79%	1.79%
Avg. 50% CI span	3.75%	3.66%	3.63%
Avg. 75% CI span	6.50%	6.37%	6.43%
Avg. 95% CI span	11.41%	11.10%	11.30%
Number of sites	720	724	714

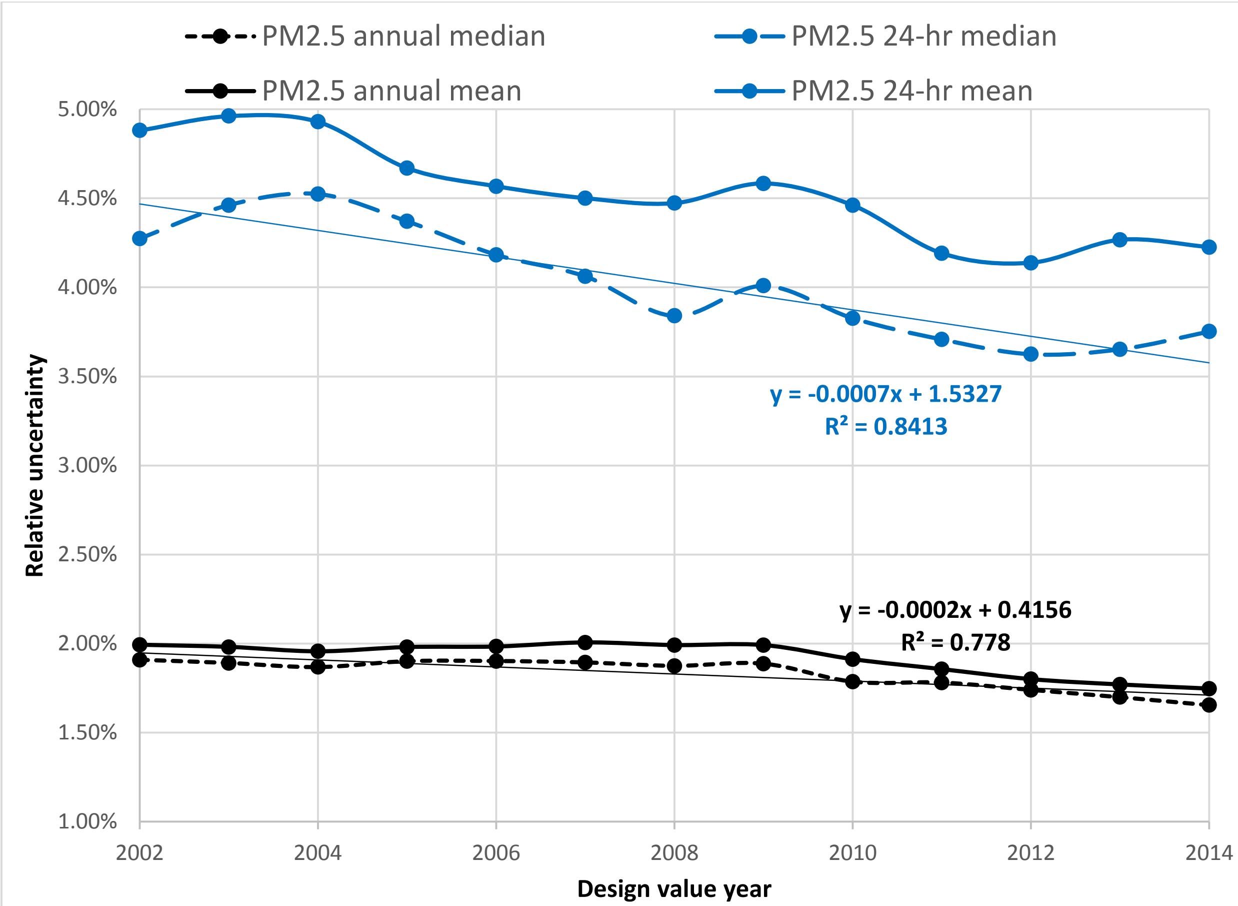


Figure 15 – Median and mean variability in the network determined from the bootstrap analysis (50% CI) for the 13 DV periods from 2002-2014 for $PM_{2.5}$.

5. Additional Information

Data for the analyses presented in this document can be obtained by contacting:

Chris Owen, PhD
Office of Air Quality Planning and Standards, U. S. EPA
109 T.W. Alexander Dr.
RTP, NC 27711
919-541-5312
owen.chris@epa.gov

United States Environmental Protection Agency	Office of Air Quality Planning and Standards Air Quality Analysis Division Research Triangle Park, NC	Publication No. EPA-454/D-16-001a July, 2016