UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC 27711

APR 17 2018

OFFICE OF
AIR QUALITY PLANNING
AND STANDARDS

MEMORANDUM

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

FROM: Peter Tsirigotis
Director

TO: Regional Air Division Directors, Regions 1-10

The purpose of the attached document is to provide guidance on compliance demonstration tools for use with ozone and fine particles (PM2.5) in the Prevention of Significant Deterioration (PSD) permitting program. The Environmental Protection Agency (EPA) 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 for the $PM_{2.5}$ PSD increments. Permitting authorities may use these values to help determine whether a proposed PSD source causes or contributes to a violation of the corresponding NAAQS or PSD increments. Separately, we have developed a technical document that provides a detailed discussion of the technical analysis used in the development of these values and a legal memorandum that provides further detail on the legal basis that permitting authorities may choose to adopt to support using SILs to show that requirements for obtaining a PSD permit are satisfied.[1] This guidance provides a summary of 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 the review of PSD permit applications. This guidance, and the technical and legal documents, are not final agency actions and do not create any binding requirements on permitting authorities, permit applicants or the public.

Please share this guidance with permitting authorities in your Region. If you have questions regarding the guidance, please contact Raj Rao at *rao.raj@epa.gov* or (919) 541-5344. For questions regarding the technical document, please contact Tyler Fox at *fox.tyler@epa.gov* or (919) 541-5562. For questions regarding the legal document, please contact Brian Doster at *doster.brian@epa.gov* or (202) 564-1932.

Attachment

[1] "Technical Basis for the EPA's Development of Significant Impact Thresholds for $PM_{2.5}$ and Ozone," EPA-454/R-18-001, April 2018; "Legal Memorandum: Application of Significant Impact Levels in the Air Quality Demonstration for Prevention of Significant Deterioration Permitting under the Clean Air Act," April 2018.

Attachment

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

I. INTRODUCTION

When a Prevention of Significant Deterioration (PSD) permit applicant has shown through air quality modeling that the projected air quality impact from a proposed source for a particular pollutant is not significant or meaningful, the EPA believes there is a valid analytical and legal basis in most cases for the permitting authority to conclude that the proposed source will not cause or contribute to a violation of a National Ambient Air Quality Standard (NAAQS) or PSD increment for that pollutant. To show that the proposed source will not have a significant or meaningful impact on air quality, permit applicants and permitting authorities may elect to use these Significant Impact Level (SIL) values (air quality concentration values) as a compliance demonstration tool. In this guidance and accompanying documents, the EPA has provided policy, technical and legal analyses that permitting authorities may choose to adopt in supporting the use of the SILs to make the required demonstration in particular PSD permitting actions. The use of SILs can help satisfy PSD requirements while expediting the permitting process and conserving resources for permit applicants and permitting authorities.

The EPA has previously issued guidance describing particular uses of SILs.[1,2,3,4] The EPA has also recognized that permitting authorities have the discretion to apply SILs on a case-by-case basis in the review of individual permit applications, provided such use is justified in the permitting record.[5] In an effort to reduce the need for case-by-case justification by permitting authorities, the EPA finalized a rule in 2010 to codify, among other things, particular $PM_{2.5}$ SIL values and specific

---

[1] 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.
[2] 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.
[3] 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.
[4] 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.
[5] 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 (Environmental Appeals Board (EAB) 2011).

applications of those values ("2010 rulemaking").[6] However, in the course of subsequent litigation over this rule, the EPA conceded the regulation was flawed because it did not preserve the discretion of permitting authorities to require additional analysis in certain circumstances, and the court granted the EPA's request to vacate and remand the rule so that the EPA could address the flaw.[7]

Following the litigation, the EPA began developing a new rule to address the flaw identified in the 2010 rulemaking.[8] 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.[9] Thus, 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. During this second step, to assess whether it is appropriate to codify particular SIL

---

[6] 75 FR 64864 (October 20, 2010).

[7] *Sierra Club v. EPA*, 705 F.3d 458, 463-66 (D.C. Cir. 2013). In its litigation brief at n. 10, the EPA stated an intent to issue guidance in the near future concerning $PM_{2.5}$ values remaining in 40 CFR 51.165(b)(2). 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.

[8] 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.*

[9] *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).

values for ozone and $PM_{2.5}$, the EPA will consider whether permitting experience has confirmed that the recommended SIL values are suitable in all circumstances to show that an increase in air quality concentration below the value does not cause or contribute to a violation of the NAAQS or PSD increments.

Permitting authorities retain discretion to use or not to use these EPA-derived SILs in particular PSD permitting actions. If a permitting authority chooses to use these SIL values to support a case-by-case permitting decision, it must justify the values and their use in the administrative record for the permitting action.[10] Permitting authorities also have discretion to develop their own SIL values, provided that such values are properly supported in the record for permitting actions or decisions in which the values are used to make the required showing. Detailed technical guidance on the development of alternative SIL values is beyond the scope of this document; however, we provide a limited discussion later in this document (*see*, e.g., page 12). This guidance (including the legal and technical documents) supporting the EPA's recommended SIL values may be viewed as a model for permitting authorities that seek to develop alternative SIL values. Permitting authorities may elect to utilize alternative "confidence intervals" as well as regional or local factors in developing their own SIL values.[11]

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 a stronger analytical foundation for the EPA recommended SIL values. This guidance 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 to demonstrate that a proposed source does not cause or contribute to a violation of NAAQS or PSD increments. 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 increments. However, this guidance is not a final agency action and does not reflect a final determination by the EPA that any particular proposed source with a projected impact below the recommended SIL value does not cause or contribute to a violation. 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

[10] *Rocky Mountain Steel Order* at 16-18, supra footnote 5. Such a justification may incorporate the information compiled by the EPA to support the SILs recommended in this memorandum.

[11] A description of the "confidence interval" is provided at page 12 of this document and in the technical document at section 2.2 (Statistical Methods and Assessing Significance Using Confidence Intervals).

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 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.[12] The EPA has reflected this requirement in its PSD regulations.[13] The Clean Air Act (Act) 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 identify such models and the conditions under which they may be used in the PSD program to make the demonstration required under the Act.[14]

Using the models identified in the EPA's 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 a 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.[15] 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.[16]

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 Act 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.[17] Thus, the EPA and other permitting authorities have concluded that a

---

[12] 42 U.S.C. 7475(a)(3) (section 165(a)(3) of the Act). The EPA interprets the phrase "in excess of" to mean a violation, not the exceedance described in 40 CFR 50.1(l).

[13] 40 CFR 51.166(k); 40 CFR 52.21(k).

[14] The PSD regulations at 40 CFR 51.166(l) and 52.21(l) require the use of "applicable models, data bases, and other requirements" specified in 40 CFR part 51, Appendix W, also known as the *Guideline on Air Quality Models (Guideline*).

[15] 1990 Draft New Source Review (NSR) Workshop Manual at C.51.

[16] 40 CFR part 51, App. W, § 9.2.3; 1990 Draft NSR Workshop Manual at C.52.

[17] *In re: Prairie State Generating Co*., 13 E.A.D. 1, 105 (EAB 2006). This EAB opinion includes a long discussion of the EPA's prior guidance with other examples.

proposed source may meet the requirements in 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.[18]

Historic Use of SILs

In the context of section 165(a)(3), 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.[19] Consistent with the 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.[20] 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, Volume 10 (Revised) and Procedures for Evaluating Air Quality Impact of New Stationary Sources."[21]

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

Since the inception of the PSD program, the EPA has faced technical challenges with providing compliance demonstration tools for those pollutants that are not directly emitted by sources (ozone and secondarily-formed $PM_{2.5}$) and which form through chemical reactions of precursor pollutants. In July 2010, the Sierra Club petitioned the EPA to initiate rulemaking regarding the establishment of air quality models for ozone and $PM_{2.5}$ for use by PSD permit applicants. In January 2012, the EPA granted the petition and committed to engage in rulemaking to evaluate whether updates to the *Guideline* 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

[18] 1990 Draft NSR Workshop Manual at C.52.
[19] 61 FR 38250, 38293 (July 23, 1996); 72 FR 54112, 54139 (September 21, 2007).
[20] 1990 Draft NSR Workshop Manual at C.51-C.52.
[21] October 1977, U.S. EPA, Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711. The 1977 document did not discuss SILs, but did identify procedures for air quality analyses pursuant to the PSD program.

the impacts of individual stationary sources on the formation of these air pollutants"[22] 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 finalizing revisions to the EPA's *Guideline*.[23] As discussed in the preamble to the *Guideline*, 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 and $PM_{2.5}$ PSD increments. The revisions to the *Guideline* include criteria and process steps for choosing single-source analytical techniques or models to estimate ozone impacts from precursor nitrogen oxide (NOx) and volatile organic compound (VOC) emissions and to assess concentrations of direct and secondarily-formed $PM_{2.5}$. The ozone and $PM_{2.5}$ SIL values recommended in this guidance are intended to complement the *Guideline* updates by providing thresholds that may be used to determine whether an increase in air pollutant concentration (impact) predicted by the chosen technique or model causes or contributes to a violation.

In the 2010 rulemaking, the EPA established SIL values for $PM_{2.5}$ in paragraph (k)(2) of the PSD regulations at 40 CFR 51.166 and 52.21. 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.[24] 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 was 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 between (1) that rule text, which left no discretion for the permitting authority, and (2) our statements in the preamble to the 2010 rulemaking, which identified certain circumstances where it may not be

[22] Letter from Gina McCarthy, Assistant Administrator, EPA Office of Air and Radiation, to Robert Ukeiley, Sierra Club, January 4, 2012.
[23] 82 FR 5182 (January 17, 2017).
[24] *Sierra Club v. EPA*, 705 F.3d 458, 466 (D.C. Cir. 2013).

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.[25]

The court left intact the $PM_{2.5}$ NAAQS significance levels separately promulgated at 40 CFR 51.165(b)(2), because the regulatory text in that section did not say that a proposed source that has an impact less than the significance level is always deemed to not cause or contribute to a violation. The regulatory text at 40 CFR 51.165(b)(2) says that a major source or major modification with a projected impact greater than the listed significance level at any location that does not or would not meet the applicable NAAQS will be considered to cause or contribute to a violation, but this provision does not compel the opposite conclusion for projected impacts equal to or below that level.[26]

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 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" in section 165(a)(3) to be inapplicable to an air quality impact that is insignificant. This interpretation is more fully explained in the legal memorandum. In the context of this section of the Act, the EPA believes an insignificant impact is an impact on air quality concentrations that is small and not meaningful (e.g., the EPA has often described such an impact as "trivial" or "*de minimis*").

As discussed in more detail in the legal memorandum, a permitting authority may conclude that a PSD permit applicant will "cause" a modeled violation of a NAAQS when the increased emissions from construction or modification of the proposed source are the reason for, responsible for, or the "but for" cause of the violation. However, a permitting authority must also consider whether emissions "contribute" to a violation in circumstances where a violation of the NAAQS is present before considering the proposed increase in emissions from a PSD construction project, or when

---

[25] 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, 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, 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, 64894.

[26] 40 CFR 51.165(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 former 40 CFR 51.166(k)(2) and 52.21(k)(2), and, in this guidance, that an impact equal to the SIL is treated the same as impacts above the SIL.

emissions from multiple sources may impact a particular area. In the absence of specific language in section 165(a)(3) regarding the degree of contribution that is required (such as the term "significantly"), a permitting authority 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. A permitting authority may also identify criteria or factors that may be used to determine whether something contributes, including qualitative or quantitative criteria that are appropriate to the particular context.[27]

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 cause or contribute to air quality concentrations in violation of applicable standards. One of the goals of the development of SILs as a compliance demonstration tool is to ensure an appropriate balance between maintenance of air quality and PSD permit process streamlining. 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 a significant impact on air pollutant concentrations. Using a quantitative 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 cause or contribute to a violation in a particular context.

Historical Approach for Developing SILs

To determine what is (and is not) a significant impact in the context of section 165(a)(3) of the Act, the EPA has previously supported using the levels in 40 CFR 51.165(b)(2).[28] The EPA has

[27] *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 Act, 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 Act.

[28] The Emison Memo, *supra* footnote 5, references 40 CFR 51.165(b)(2) for the purpose of defining "significant" in this context. 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."

described these levels as "significance levels."[29] 40 CFR 51.165(b)(2) was originally promulgated by the EPA in 1987 as part of an offset provision 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.[30] 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 regulation. For example, 40 CFR 51.165(b)(2) states that a proposed source impact that is greater than 5 micrograms per cubic meter ($\mu g/m^3$) for the 24-hour sulfur dioxide (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, particularly ozone (and not for $PM_{2.5}$ until the 2010 rulemaking), but for those NAAQS covered in this regulation, the application is the same. Over time, these air quality concentration significance levels in 40 CFR 51.165(b)(2) have become known as "significant *impact* levels"[31] [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.[32] The EPA has also issued guidance memoranda that have provided recommended SIL values for the 1-hour nitrogen dioxide (NO_2) and SO_2 NAAQS, to be used for the purpose of determining what are (and are not) significant impacts for these pollutants in the context of the 1-hour standards.[33]

As referenced above, the EPA's values contained in 40 CFR 51.165(b)(2) originally were related to the level of protection afforded by the PSD increments that Congress established for Class I areas.[34] The EPA generally relied on that approach in 2010 by using the ratio of the $PM_{2.5}$ NAAQS

---

[29] The EPA initially promulgated these same concentration values in 1978 and described them as the "minimum amount of ambient impact that is significant." 43 FR 26380, 26398 (June 19, 1978). In the 1979 Emissions Offset Interpretative Ruling (Appendix S to 40 CFR part 51), the EPA used these values 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. 44 FR 3274, 3283 (January 16, 1979). Under Appendix S, as revised in 1980, the EPA considered a source to "cause or contribute to" a violation if the impact of the source or modification would exceed these significance levels at any locality that does not meet the NAAQS. 45 FR 31307, 31311 (May 13, 1980).

[30] 52 FR 24672, 24713 (July 1, 1987).

[31] 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.

[32] 40 CFR 52.21(b)(23) defines the term "significant" and applies discrete values for determining if the emissions increase from a proposed source will be significant. This regulation states that an increase in emissions of each ozone precursor (VOC and NOx) is significant if it equals or exceeds 40 tons per year (tpy) and, for direct emissions of $PM_{2.5}$ the significance level is 10 tpy. For $PM_{2.5}$ precursor emissions, the significance level is 40 tpy for SO_2 and 40 tpy for NOx.

[33] Page memoranda, *supra* footnotes 1 and 2 of this attachment.

[34] 43 FR 26380, 26398.

to the particulate matter 10 micrometers or less in diameter (PM_{10}) NAAQS as a multiplier to add $PM_{2.5}$ values to 40 CFR 51.165(b)(2) and to establish $PM_{2.5}$ SIL values in 40 CFR 51.166(k)(2) and 52.21(k)(2).[35] However, given limitations in the rationale supporting them, the EPA recognized in the preamble to the 2010 rulemaking 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."[36] 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.[37] This approach was intended to guard against misuse of the SILs in situations where the existing air quality was already close to the NAAQS.

Analytical Foundation for Recommended SILs

Since the May 2014 $PM_{2.5}$ modeling guidance was issued, the EPA has conducted a statistical analysis that provides an improved analytical foundation for the EPA's selection, based on the policy considerations described below, of a degree of change in concentration that permitting authorities may use to represent 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 technical document. 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 relevant NAAQS is greater than the SIL value before applying one of the recommended $PM_{2.5}$ SIL values. The EPA's intention with this new method was 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. However, permitting authorities retain discretion whether to apply SILs as a general matter, or in particular permitting actions, based on information in the permit record.

In order for a specific change in air quality concentrations to be used to show that a proposed source does not cause or contribute to a violation of the NAAQS, the concentration change must

35 75 FR 64890.

36 75 FR 64864, 64892.

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

represent a level of impact on ambient air quality that is not significant or meaningful. The EPA's judgment is that values representing such a level can be selected from a statistical analysis of the variability of air quality, using data from the U.S. ambient monitoring network for ozone and $PM_{2.5}$. Due to fluctuating meteorological conditions and changes in day-to-day operations of all air pollution sources in an area, there is an inherent variability in the air quality in the area surrounding a monitoring site. This variability can be characterized through the application of a well-established statistical framework for quantifying uncertainty.[38,39] The analysis described in the technical document quantifies the inherent variability in pollutant concentrations (as measured by design values) and informs the EPA's choice of a value for a change in concentrations that the EPA does not consider significant or meaningful because changes of this magnitude are well within the inherent variability of observed design values.[40] Once the precautionary choices described below are built into the calculation, this degree of change in concentration is, thus, indistinguishable from the inherent variability in the measured atmosphere and may be observed even in the absence of the increased emissions from a new or modified source. Therefore, a permitting authority can reasonably conclude that emissions of a proposed source that have a projected impact below the SIL values provided in this memorandum are not the reason for, responsible for, or the "but for" cause of a NAAQS violation. Likewise, this indicates that changes in air quality within this range are not meaningful, and, thus, do not contribute to a violation of the NAAQS.

Before delving in detail into the technical and policy considerations that inform the EPA's choice of the SILs recommended in this document, it is important to point out that the discretion of the EPA and other permitting authorities is limited by the 2010 rulemaking. Specifically, since the EPA has established by regulation that a $PM_{2.5}$ impact greater than a certain value will be considered to 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 40 CFR 51.165(b)(2), permitting authorities are not precluded from developing a higher ozone NAAQS SIL value than recommended in this guidance. Likewise, 40 CFR 51.165(b)(2) does not address PSD increments and, thus, does not constrain the discretion of a permitting authority to develop a higher SIL value and use it for PSD increment purposes.

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

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

[40] The EPA conducted an external peer review of the technical document containing the statistical analysis used for developing the SILs for ozone and $PM_{2.5}$. The peer review comments were supportive of the air quality variability method as being appropriate for application for SILs. The comments also suggested several considerations for improvements to the technical document and analyses to better support the application of the analysis to determine specific SIL values. Therefore, the EPA made a number of revisions to the technical document, including conducting new analyses to investigate issues raised by the reviewers, edits to a number of sections for clarity and accuracy, and updating the analysis to include the most recent data. A peer review report that outlines the subsequent changes to the technical analysis is available from the U.S. EPA library, library number EPA 454/S-18-001.

Basis for Development of Recommended SILs for Ozone and $PM_{2.5}$

In developing the recommended SILs for ozone and $PM_{2.5}$, we assessed the variability in pollutant concentrations, as determined by the national monitoring network, from the design value at each monitor (i.e., baseline value). The technical analysis uses traditional statistical techniques based on statistical significance testing to characterize the variability in air quality. The conceptual underpinnings of the analysis are an application of the concept of "statistical significance" to inform a policy decision regarding what represents an insignificant impact and, therefore, may serve as the basis for developing a SIL for use in the air quality impact analyses required for PSD permitting. More specifically, traditional statistics is based on the concept of identifying what constitutes a statistically significant change from a baseline value where the "baseline" is the statistic of interest, such as the mean or, in this case, the design value. Rather than focusing on statistically significant changes, the purpose of the analysis was to calculate changes in the design values that, once precautionary choices are applied, may be considered not significant or meaningful. To identify recommended SILs for the desired application in the PSD program, the EPA determined that the findings of the statistical analysis can be used to identify a change in the design value (i.e., an air quality impact) below which a permitting authority may reasonably conclude that the impact does not cause or contribute to a violation of a NAAQS. The principles of statistical significance testing do not by themselves provide a single, unique threshold for determining the statistical significance of a change in the design value. Statistical significance testing provides a range of concentration values that can be considered to represent a statistically significant change in air quality or, in this application, a change in air quality that is not statistically significant. Therefore, it is necessary to consider the function and application of SIL values in the context of the PSD program and to select a change in air quality that is reasonably representative of the showing that a proposed source will not cause or contribute to a NAAQS violation, as required by the Act and PSD regulations.

In making a recommendation for an appropriate SIL value, the EPA balanced two considerations: 1) the usefulness of the SIL as a compliance demonstration tool in the PSD permitting program, and 2) the likelihood of a SIL value representing an impact that is not significant. In balancing these considerations, the EPA made policy decisions concerning the confidence interval (CI) to represent the inherent variability for purposes of the NAAQS compliance demonstration, the approach used to scale local variability to the level of the NAAQS, the geographic extent of each summary value, and the design value year or years from which to use the variability results. As described below, for each of these factors, the EPA chose options that are precautionary, leading to SILs designed to ensure the protection of air quality.

Through the statistical analysis, we calculated CIs, which represent different assessments of the level of change in air quality based on the inherent variability in the air quality of an area. We then selected the recommended SIL values as a function of the CIs, the baseline value, and policy considerations. The selection of a CI in defining a particular SIL value required an exercise of judgment based on the technical and policy considerations (as described below) such that the selected value represents a level of change in air quality concentration that can be considered not significant or meaningful in the context of evaluating the impact of emissions from a proposed source. These policy considerations work in conjunction with the statistical analysis, to provide a rational basis to select values derived from the statistical analysis that can be applied as a tool for making the PSD compliance demonstration required by the Act and PSD regulations. For more information on the design and results of the technical analysis, please refer to the technical document.

The technical analysis relies upon data from the national ambient monitoring network for ozone and $PM_{2.5}$. Because these data generally are the basis for determining NAAQS attainment, they are an appropriate basis to characterize air quality, with the statistical analysis evaluating the variation in the design value at each monitoring site across the nation. This variability in air quality concentrations is described by the different CIs computed from the statistical analysis. The CIs identify a statistically significant deviation from the baseline value. As described in the technical document (Section 3.0), the EPA has calculated CIs at the 25 percent, 50 percent, 68 percent, 75 percent, and 95 percent intervals for consideration in defining SIL values for ozone and $PM_{2.5}$. The smallest CI that might be used to identify a statistically significant change would be a 68 percent CI, which corresponds to one standard deviation from the baseline value. Thus, any change in the design value larger than the variation represented by the 68 percent CI could be considered to be a statistically significant change. However, for purposes of the PSD program, we are seeking to identify a concentration value that constitutes an insignificant impact, meaning a change in the design value that does not reflect a meaningful difference in air quality based on the introduction of a new source. Thus, from a statistical perspective, the EPA believes that the CIs used in determining an appropriate SIL value should be below 68 percent, corresponding to a change of less than one standard deviation.

Very small SIL values would have limited use to permitting authorities (i.e., would lead to "false positives"), while larger values (closer to the air quality change represented by the 68 percent CI) would lead to "false negatives." In weighing these competing considerations to select an appropriate SIL value, the EPA believes that air quality change represented by a 50 percent CI represents a protective approach for a SIL value because it is sufficiently within the 68 percent CI, while still being sufficiently higher than zero such that it can be a useful compliance demonstration tool for the PSD permitting process. Of the available choices, the 50 percent CI has more utility as a screening tool under the permitting program, while providing a value that adequately reflects a change in air quality concentrations that is not significant or meaningful.

The EPA chose to use the relative variability rather than the absolute variability in calculating the SILs because the technical analysis (Section 4.0) showed that the relative variability is fairly consistent across the range of design values, suggesting a commonality in the relative variability across a wide range of geographic regions, chemical regimes, and baseline air quality levels in the development of the SILs.

In order to promote national consistency, the EPA has historically provided national SIL values rather than regional or local values. The EPA considered whether a SIL value should be informed by the statistical analysis at the particular site of the proposed source or the central tendency across all monitored sites in the U.S., regardless of the proposed source's planned location. The EPA continues to recommend using a national SIL value based on the variability aggregated across the nation rather than developing regional or local values. Findings from the statistical analysis indicate that while there are local spatial correlations, there are few instances of large scale (e.g., region-to-region) trends in ambient air variability. Thus, national numbers are supported by the spatial analysis and suitable for use here. Because NAAQS and PSD increments are set on a national basis, the EPA and permitting authorities have historically used national SILs in the PSD program. National SIL values are designed to be used for any location subject to PSD requirements and eliminate the need to determine local or regional approaches for developing a SIL value, including addressing the status of local air quality monitoring (which would be needed if regional or local SILs were to be determined). However, as noted above, local permitting authorities have the discretion to develop alternate SILs.[41] Having a national SIL value promotes consistency in implementation and prevents possible confusion or arbitrary choices that may arise with highly localized SIL values (i.e., determining which monitors to use for computations and other possible deviations from national protocol). Given these considerations, the EPA recommends continuing the practice of using national SIL values. Furthermore, as shown in the technical analysis (Section 4.0), because the median statistic is less influenced by high variability areas, the median statistic is preferred for use in selecting a SIL. Therefore, using the median statistic of the relative variability from the 50 percent CIs from the entire U.S. ambient monitoring network satisfies the policy needs for a SIL and is congruent with the physical and chemical processes that result in this variability.

Next, the EPA chose to use the most recently available years of ambient monitoring data (2012-2016) in the technical analysis to derive the recommended SILs. The SILs should reflect the most recent and representative state of the nation's atmosphere. In assessing the historical trends in ozone and $PM_{2.5}$ air quality levels across the nation, there are observable downward trends in concentrations that indicate more recent data are most appropriate. To have more confidence that the resulting values would not be unduly influenced by temporary circumstances or episodic events, the EPA's recommended SILs are based on an average of the most recent three design value years as a basis for ozone and $PM_{2.5}$ SIL development (i.e., 2012-2014, 2013-2015, 2014-2016).

[41] In the cases where a permitting authority is considering an alternative SIL(s) due to the characteristics of regional variability (e.g., if, based on the analysis presented in the technical document, a specific area appears to have more localized variability than the national average), it is important to understand the factors driving that apparent variability to fully support the application of alternative SIL(s). For example, the results presented in section 4.3 of the technical document show some areas with regional variability for the 24-hour $PM_{2.5}$ standard, though no regional trends were apparent for the annual $PM_{2.5}$ standard and the ozone standard. Furthermore, these regional trends for the 24-hour $PM_{2.5}$ standard were not apparent in the other data years shown in the appendix of the technical document. Additionally, the discussion in the technical document highlights potential causes for some of the variability in these regions (e.g., lower sampling frequency, that can lead to apparently higher variability than would otherwise be shown with higher sampling frequency). Similar issues are discussed in the technical document and can have important consequences for the results and conclusions drawn from more localized analyses of the ambient data and should be thoroughly vetted when considering alternative SILs.

SILs for NAAQS

Using the method described above, the EPA developed SIL values for the 8-hour ozone NAAQS and the annual and 24-hour $PM_{2.5}$ NAAQS. Table 1 lists these SIL values for the NAAQS. Each of these SIL values is based on the level, averaging period and statistical form of its corresponding NAAQS. For the reasons discussed in this guidance and supporting documents, we recommend that PSD permitting authorities use the following values as SILs on a case-by-case basis in the manner described in the next section.

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 accounts for the significance level for the 24-hour $PM_{2.5}$ NAAQS in 40 CFR 51.165(b)(2). Refer to the guidance discussion for details.

For the 8-hour ozone NAAQS, the SIL value we recommend is 1.0 part per billion (ppb). Consistent with the form of the NAAQS, this value is based on the annual 4th highest daily maximum 8-hour concentration, averaged over 3 years. The recommended SIL value for ozone is the same as the derived value from the air quality variability analysis.

For the 24-hour $PM_{2.5}$ NAAQS, the SIL value we recommend is 1.2 µg/m^3. The derived value from the air quality variability analysis is 1.5 µg/m^3 and is based on an analysis of the 98th percentile 24-hour concentrations averaged over 3 years. However, 40 CFR 51.165(b)(2) still lists 1.2 µg/m^3 as the significance level for the 24-hour $PM_{2.5}$ NAAQS. In the 2010 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 1.5 µg/m^3 is not significant, significance levels for $PM_{2.5}$ remain in the EPA's regulations at 40 CFR 51.165(b)(2) and the EPA 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 1.5 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 can 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.[42]

For the annual $PM_{2.5}$ NAAQS, we recommend 0.2 μg/m^3 as the SIL value, which is the value based on a 3-year average of annual average concentrations. This value is lower than the value of 0.3 μg/m^3 listed in 40 CFR 51.165(b)(2). Since 40 CFR 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, the EPA and other 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. However, based on the ambient air quality variability approach, the EPA's judgment is that an impact below 0.2 μg/m^3 is not significant and should be considered to not cause or contribute to any violation of the annual $PM_{2.5}$ NAAQS that is identified.

We recommend that these SIL values apply to the NAAQS everywhere, regardless of the class of the airshed.[43] 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, as described below in the discussion of SILs for 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.

<u>SILs for PSD Increments</u>

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 PSD increment SILs for the $PM_{2.5}$ PSD increments (*see* Table 2), but in an indirect way. The SIL values

---

[42] 40 CFR 51.165(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 guidance, 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.

[43] When Congress established the PSD program requirements under the 1977 Act Amendments, it included specific numerical PSD 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 PSD 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 PSD increments, but to date no Class III areas have been designated. The EPA subsequently defined Class I, II and III PSD increments for NO_2 and PM_{10}, and $PM_{2.5}$ in multiple rulemakings.

for the $PM_{2.5}$ PSD increments are derived from the recommended 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. This is because the EPA recognizes that Congress intended to establish special protection for Class I areas, as observed by the more stringent statutory Class I PSD increments, as well as provisions for use of air quality related values (including protection against visibility impairment).[44] To help reflect 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.[45] 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 $PM_{2.5}$ PSD Increments

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

IV. APPLICATION OF SILS

The EPA recommends that permitting authorities consider using these SIL values for ozone and $PM_{2.5}$ 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 guidance 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.[46] A modeled 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

[44] Section 165(d)(2) of the Act sets forth procedures affording special protection against adverse air quality impacts in Class I areas. Also, section 169A of the Act declares a national goal of preventing future and remedying any existing impairment of visibility in Class I areas. 42 U.S.C. 7475 and 7491.

[45] To derive the Class I PSD increment SIL values, we started with the corresponding NAAQS SIL value as the base number and adjusted it by the ratio of the associated Class I and II PSD increments. For the annual $PM_{2.5}$ increment, we reduced the NAAQS SIL value by the ratio of 1:4, because the Class I PSD increment is 1 $\mu g/m^3$ and the Class II PSD increment is 4 $\mu g/m^3$. We used the ratio of 2:9 for the 24-hour $PM_{2.5}$ increment. For the 24-hour increment, we used the 40 CFR 51.165(b)(2) value of 1.2 $\mu g/m^3$ as our base number.

[46] 1990 Draft NSR Workshop Manual at C.24-C.25, C.51.

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 shows a significant impact, permitting authorities may choose to use the recommended SIL values in a cumulative impact analysis for a NAAQS, which, in addition to the proposed new major stationary source or major modification, includes the impact of existing sources (onsite with the proposed major modification, as well as other existing sources), and the appropriate background concentration. The EPA has described this application of a SIL as a "culpability analysis."[47] 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 recommended 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. This demonstration would, thus, allow the permit to be issued if all other PSD requirements are satisfied. If the proposed source's modeled impact is higher than or equal to the recommended 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 a not significant 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.[48]

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 once 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 for the relevant pollutant, then the permitting authority may conclude that the source does not cause or contribute to a violation of the PSD increment for that pollutant.

---

[47] *Prairie State*, 13 E.A.D. at 100; Mississippi Lime, 15 E.A.D. at 374.

[48] 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 an offset program based on federal regulations at 40 CFR 51.165(b).

In the past, 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 to complete the required modeling with a reasonable effort. As discussed earlier, the EPA recently updated its *Guideline*. The revisions include providing an appropriate, revised basis for determining the modeling domain for NAAQS and PSD increment assessments. Thus, the revised *Guideline* should be used when considering the extent of the modeling domain.

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

Subject to limitations described in this guidance, permitting authorities may use 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. Since this guidance 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.

The case-by-case use of SIL values should be justified in the record for each permit. To ensure an adequate record, any PSD permitting decision that is based on this guidance (including the technical and legal documents) should incorporate the information contained in them. The permitting authority should also consider any additional information in the record that is relevant to making the required demonstration.

Permitting authorities also retain the discretion to use other values that may be justified separately from this guidance 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 the 2010 rulemaking constrains the discretion of state and local permitting authorities, the EPA is committed to reassessing 40 CFR 51.165(b)(2) through a future rulemaking process that will begin within 18 months.

Because ozone is not addressed in 40 CFR 51.165(b)(2), permitting authorities are not precluded from developing a higher ozone NAAQS SIL value than recommended in this guidance. Likewise, 40 CFR 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 PSD increment purposes. Permitting authorities are also not precluded from developing and using lower SIL values than recommended in this guidance. Permitting authorities may elect to utilize alternative CIs, based on regional or local factors, in developing their own SIL values. 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.

Legal 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 (Act), an applicant for a preconstruction 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 Standard (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 Act does not specify how a PSD permit applicant or permitting authority is to determine whether a proposed new or modified source will (or will not) cause or contribute to a violation of a NAAQS or applicable PSD increment. *Id.*

The language of section 165(a)(3) of the Act supports two basic approaches that a PSD permit applicant may use to demonstrate that the proposed source's emissions will not cause or contribute to a violation of a NAAQS or PSD increment. One approach 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 approach is to demonstrate that the emissions from the proposed source do not cause or contribute to any violation of a NAAQS or PSD increment that has been identified prior to preparation of a permit application or that is identified or projected in the course of preparing and reviewing a permit application.[1] Considering the relevant terms of the Act and other factors discussed below, when applying this second approach, 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

[1] *See* NSR Workshop Manual at C.51-52. The EPA has described both of these approaches as elements of an overall "second approach" that the Agency has recommended applying since 1988. *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), at 2 ("Emison Memo"). The EPA did not favor the "first approach" described in the 1988 memorandum -- to automatically consider a source to cause or contribute to any modeled violation that would occur within its impact area.

based solely on an interpretation of the phrase “cause, or contribute to,” as specifically used in the context of section 165(a)(3) of the Act, without relying on the inherent authority to establish exemptions for *de minimis* circumstances.

Analysis

Two aspects of the Act 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, the phrase “cause, or contribute to” and the included terms “cause” and “contribute” are not specifically defined in the Act itself. Second, 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.

The phrase “cause, or contribute to” and the included terms “cause” and “contribute” are not defined in section 169, section 302, or any other section of the Act. Courts have observed that the absence of a statutory definition does not by itself establish that a term is ambiguous. *NRDC v. EPA*, 489 F.3d 1250, 1258 (D.C. Cir. 2007). In the absence of a definition, the ordinary meaning of a term should govern. *Petit v. Dep’t of Education*, 675 F.3d 769, 781 (D.C. Cir. 2012). But courts have also observed that the meaning of a statutory term depends on the context in which it is used. *Bell Atlantic Telephone Co. v. FCC*, 131 F.3d 1044, 1047 (D.C. Cir. 1997).

To discern the ordinary meaning of the term “cause,” one can look to dictionary definitions. For example, according to the Merriam-Webster dictionaries, when used as a verb (as in section 165(a)(3) of the Clean Air Act), the word “cause” means “to compel by command, authority, or force.” <https://www.merriam-webster.com/dictionary/cause>. The American Heritage Dictionary includes a similar meaning when “cause” is used as a verb, but adds “to be the cause or reason for” and “result in.” <https://ahdictionary.com/word/search.html?q=cause>. The term “cause” may also be used as a noun. The Merriam-Webster definition for this usage of “cause” includes “a reason for an action or condition” and “something that brings about an effect or a result.” The American Heritage definition of “cause” includes “the producer of an effect, result, or consequence” and “a person, event, or condition, that is responsible for an action or result.” Thus, based on these definitions of “cause,” emissions from a proposed PSD source that will be responsible for, be the reason for, or result in a violation of the NAAQS may be considered to cause that violation.

Under principles of common law, behavior is generally not considered to be the cause of an injury unless that injury would not have occurred "but for" the behavior. *See* 57A Am. Jur. 2d Negligence § 415. Applying this classic understanding of the concept of causation, a permitting authority may conclude that a PSD permit applicant will "cause" a modeled violation of a NAAQS if the modeled violation would not be projected to occur "but for" the increased emissions from construction or modification of the proposed source.[2] However, it is clear from the "cause, or contribute to" language in section 165(a)(3) of the Act that Congress did not intend for this provision to apply only when emissions from a proposed source are a "but for" cause of a violation of the NAAQS or PSD increment. This is because the term "cause" is followed by the phrase "or contribute to." Given the addition of this phrase, section 165(a)(3) should be read to apply not only where a proposed source would be a "but for" cause of a new modeled violation but also where a proposed source would "contribute" to a violation that might be modeled even without the impact of the proposed source. This could include circumstances where a NAAQS violation is present before considering the proposed increase in emissions from a PSD construction project, or when emissions from multiple sources may impact a particular area.

While the use of "contribute" conveys this meaning in the context of section 165(a)(3) of the Act, one federal appeals court has recognized, based in part on competing dictionary definitions, that the term "contribute" does not itself have a consistent, ordinary meaning. *See Catawba County, N.C. v. EPA*, 571 F.3d 20, 39 (D.C. Cir. 2009). In two different contexts under the Act, the United States Court of Appeals for the District of Columbia Circuit has observed that the term "contribute" is ambiguous with respect to the degree of air quality effect to which it applies. *Id.* at 38-39; *EDF v. EPA,* 82 F.3d 451, 459, amended by 92 F.3d 1209 (D.C. Cir. 1996). In the absence of an ordinary meaning for the term, 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 "contribute

[2] In the April 2018 memorandum titled "Guidance on Significant Impact Levels for Ozone and Fine Particles in the Prevention of Significant Deterioration Permitting Program," the EPA explains how a permitting authority may conclude that increased emissions from a proposed PSD source that would result in changes in air quality concentration that are less than a statistical level of variability are not responsible for, the reason for, or the "but for" cause of a NAAQS violation.

to" air pollution in excess of air quality standards under section 165(a)(3) of the Act. *See Entergy Corp. v. Riverkeeper, Inc.*, 556 U.S. 208, 222 (2009).

In the *Catawba County* case, the court considered the use of "contribute" in section 107(d) of the Act, 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." In doing so, the court noted the Petitioners' citation of one dictionary definition and the EPA's citation of other dictionary definitions of the term "contribute" and concluded that "[t]his alone suggests an ambiguity." *Catawba County,* 571 F.3d at 39. 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 a NAAQS violation. Likewise, in the *EDF* case, the court reasoned that "contribute to" in section 176(c) of the Act is ambiguous and "leaves wide open the question of how large a reduction in emissions must be to constitute a contribution." 82 F.3d at 459.

Similar to sections 107(d) and 176(c) of the Act, section 165(a)(3) uses the ambiguous term "contribute" without specifying 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 or another PSD permitting authority 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.[3] Furthermore, this opinion supports a permitting authority'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 Act 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 or exacerbate 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, in areas that do not violate the NAAQS. Further, two goals of the PSD program are to "insure that economic growth will occur in a manner consistent with the preservation of existing clean air resources" and to "assure that any decision to permit increased air pollution in any area to which this section applies is made only after careful evaluation of all the consequences of such a decision and after adequate procedural opportunities for informed public participation in the decision-making process." 42 U.S.C. § 7470(3), (5); *see also NRDC v. EPA,* 937 F.2d 641, 645-46 (D.C. Cir. 1991) (quoting section 160(3) and (5) of the Act and inferring that "Congress believed that its PSD provisions should balance the values of clean air, on the one hand, and economic development and productivity, on the other other"). Thus, the PSD program strikes a balance that allows construction and modification of major stationary sources that will result in increased emissions in areas meeting air quality standards, but only after appropriate safeguards are in place to prevent the source from causing or contributing to 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

---

[3] *See also Environmental Defense v. Duke Energy Corp.,* 549 U.S. 561 (2007) (where the term "modification" and its definition appear, by cross-reference, in two places in the CAA, the EPA may interpret the term differently in the two contexts, so long as it does so in a reasonable manner consistent with the statutory definition).

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" when a violation of the NAAQS is predicted or pre-existing. Instead, this provision says the source must show it does not "cause, or contribute to" a NAAQS violation. This choice by Congress militates against reading section 165(a)(3) to mean that any degree of a source's projected impact on an area with a predicted or pre-existing violation of a NAAQS or PSD increment must be considered by the permitting authority to cause or contribute to such a violation (without any consideration of whether that degree of impact is meaningful). Under such a reading, a permitting authority could issue a permit only where the applicant has shown either (a) there would be no violation of the NAAQS or PSD increment in the area affected by the source or (b) increased emissions from the source would have no projected impact whatsoever in any area where the NAAQS or PSD increment is already or projected to be violated. This reading of the Act would not allow a permitting authority to exercise any judgment, and thus would fail to give meaning to the terms "cause, or contribute" that Congress used.

This legislative intent for the reviewing authority to exercise judgment in the PSD program is also supported by a comparison of the PSD provisions to the preconstruction permitting requirements applicable in areas that have been designated as nonattainment. Under this program, known as Nonattainment New Source Review (NNSR), sections 173(a)(1) and 173(c) of the Act require increased emissions from a proposed major source or major modification located in a designated nonattainment area to be offset by an equal or greater reduction in actual emissions from other sources. 42 U.S.C. § 7503(a)(1)(A), (c). There is no requirement in this part of the Act (like section 165(e) in the PSD provisions) to examine air quality in the affected area or the level or degree of air quality impact from the proposed emissions increase. The Act does not direct permitting authorities to determine whether emissions offsets are necessary to mitigate the air quality impact of the proposed construction. Rather, when a proposed source will be located in a nonattainment area, the Act in effect conclusively presumes that emissions from the source "cause" or "contribute to" the nonattainment condition because the Act requires the source to offset its emissions increase. In contrast, under the PSD program, when the proposed source will be located in an area that is designated attainment or unclassifiable for the NAAQS for that pollutant, the permitting authority must conduct an analysis of the ambient air quality impact of the source and then determine whether the increased emissions from that source "cause, or contribute to" a violation that may be projected to occur in the attainment area or occurring in an adjacent nonattainment or unclassifiable area. 42 U.S.C. § 7475(a)(3), (e). Thus, in the NNSR program, the Act's emissions offset provisions afford no discretion to the permitting authority and require every NNSR permit applicant to fully offset its emissions increase – in effect, a conclusive, *per se* presumption that an NNSR source will cause or contribute to a nonattainment problem and therefore must provide mitigation in the form of emissions offsets. By contrast, in the PSD program, the Act provides discretion to the permitting authority to determine, through the use of modeling and other analytical tools as identified by EPA, whether the emissions increase from a proposed PSD source will "cause, or contribute to" a violation, before the source would find it necessary to mitigate its ambient impact (to avoid having its permit denied where its emissions are projected to cause or contribute to a violation). This exercise of discretion by permitting authorities in assessing a proposed source's ambient impact is appropriate in light of the context and purpose of the PSD provisions of the Act, including the contrast to the lack of discretion provided to permitting authorities in the NNSR emissions offset provisions.

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. § 7475(e)(3). Section 165(e) of the Act 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. 42 U.S.C. § 7475(e)(3)(D). 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 CFR 51.166(l); 40 CFR 52.21(l); 40 CFR Part 51, Appendix W (Guideline on Air Quality Models). Thus, in section 165(e)(3) of the Act, Congress gave the EPA responsibility 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. This is evidence of legislative intent for the EPA to exercise its judgment to determine the degree of impact that "contributes to" a violation of the NAAQS and thereby fill a gap in the statutory scheme. While section 165(e)(3) addresses the promulgation of EPA rules, this provision of the statute may inform a permitting authority's interpretation of section 165(a)(3) of the Act in the context of a decision on an individual permit, because it underscores Congressional intent that the air quality impact analysis required for the issuance of PSD permits be conducted in a manner informed by EPA expertise with air quality modeling. This expertise may also be communicated by EPA in the form of nonbinding guidance to permitting authorities.

Furthermore, given their mathematical nature, the models used to make the showing required by section 165(a)(3) under the PSD program are capable of predicting increases in air pollutant concentrations that are small in relation to the level of the NAAQS. In order to give meaning to the "cause or contribute" language in section 165(a)(3) as calling for an exercise of judgment by the permitting authority, it is reasonable to conclude that Congress understood there would be a point at which a small projected air quality impact from a proposed new or modified source becomes so inconsequential[4] that PSD permitting authorities may reasonably conclude that such an impact does not cause, or contribute to, an existing or projected 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 under sections 110(a)(2)(C) and (J) of the Act. The focus of the PSD program is on controlling increased emissions from the construction and modification of large stationary sources, while some 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 violate a NAAQS, it is not necessary to construe the PSD provisions to prohibit any increase in air pollutant emissions from a source located in an attainment area or to require that such a source offset its emissions increase as in the nonattainment NSR program. The goals of the PSD program are achieved by demonstrating that

[4] As discussed herein, this conclusion can be grounded on the statutory language and its context, without invoking an agency's inherent authority to establish a *de minimis* exception from a statutory requirement under the doctrine reflected in *Alabama Power v. Costle*, 636 F.2d 323, 361-63 (D.C. Cir. 1980).

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 the existing elevated levels of air pollution in the area.

Recognizing this, the EPA has previously supported the use of concentration values,[5] called "ambient air quality significance levels" or "significant impact levels" (SILs) in the PSD program, 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);[6] NSR Workshop Manual, C.24-C.31 (Oct. 1990). For example, EPA has supported using such values in a preliminary (single-source) analysis that considers only the air quality impact from the construction proposed in a permit application to determine whether a full (or cumulative) impact analysis that also considers background concentrations and the impact of other sources in the

---

[5] 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 Act. *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.

[6] In this rulemaking notice, the EPA proposed to revise 40 CFR 51.166(k) and 52.21(k) to clarify that the emissions from an individual source seeking a PSD permit must make a "significant contribution" to a violation to support denial of a PSD permit, but this rule was not completed. In the EPA's explanation of its proposed action, the EPA used the term "significantly contribute" to mean essentially the same thing as the term "significant impact." However, the term "contribute" is used in various ways in different parts of the Clean Air Act, sometimes before or after the term "significantly." There is also ambiguity in these statutory provisions regarding the degree of impact that "contributes" to a particular air quality condition specified in each provision. Thus, the EPA and other permitting authorities should exercise more care in the future with regard to their usage of these terms in particular contexts under the Clean Air Act. With these considerations in mind, this memorandum intentionally uses the term "significant impact" and does not use the term "significant contribution." The former is used in this memorandum to describe a degree of impact on air quality concentrations that is meaningful (more than "inconsequential" or "negligible") and thus amounts to a "contribution" for purposes of section 165(a)(3) of the Act. The latter phrase ("significant contribution") is not used in this memorandum because that is not the language used in section 165(a)(3) of the Act. In circumstances where Congress has used the term "significant" or "significantly" to modify the term "contribute" or "contribution" elsewhere in the Clean Air Act, EPA should endeavor to read the Act in a way that gives meaning to this modifying language. Depending on the statutory context, one approach may be to construe the use of "significant" or "significantly" in other provisions of the Act to call for a higher degree of contribution than required under section 165(a)(3) of the Act.

area is necessary before reaching a conclusion as to whether the proposed source would (or would not) cause or contribute to a violation. 40 CFR Part 51, App. W, § 9.2.3; NSR Workshop Manual at C.24-C.25, C.51. In reviewing an individual permit decision by the EPA based on this approach, the United States Court of Appeals for the First Circuit rejected an argument that a source with an impact below a significant impact level for sulfur dioxide should have been required to conduct further analysis. *Sur Contra La Contaminacion v. EPA*, 202 F.3d 443, 446-48 (1st Cir. 2000). The court observed that EPA's decision not to require a cumulative analysis to show that emissions from a source did not cause or contribute to a violation of the NAAQS was "within its discretion, under the regulations." *Id.* at 448. EPA has also supported using these values to demonstrate that a source does not cause or contribute to a violation of the NAAQS in the area that is predicted after a cumulative impact analysis is conducted. NSR Workshop Manual at C.52. At the same time, where such a violation is nevertheless identified in the course of the PSD permitting process, the 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) ("Emison Memo"); NSR Workshop Manual at C.52.

This practice in the PSD program has been based, in part, on an interpretation by the EPA that the phrase "cause, or contribute to" in section 165(a)(3) does not apply to an "insignificant" impact. In this context, the 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 Act 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 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 the former section 10.2.3.2(a) of an earlier version of the EPA's Guideline on Air Quality Models (40 CFR Part 51, Appendix W).[7] 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 recently revised and reorganized the Guideline on Air Quality Models, and an examination of whether a proposed source has a "significant ambient impact" is still reflected in the Guideline. 82 Fed. Reg. 5182 (January 17, 2017) (*see, e.g.,* sections 4.2(c) and 8.1.2(a)).

In a 1988 guidance memorandum, the EPA explained that its position has been 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)." Emison Memo at 1. 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." NSR Workshop Manual at C.51. 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 a series of actions between 2006 and 2012, EPA sought to justify the use of SILs as an exemption to the requirement in section 165(a)(3) of the Act 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). The EPA proposed a regulation based on this rationale in 2007 for only the $PM_{2.5}$ pollutant and finalized that rule in 2010. 75 Fed. Reg.

[7] 40 CFR Part 51, App. W, § 10.2.3.2(a) (2006); 70 Fed. Reg. 68218, 68248-49 (Nov. 9, 2005).

64864 (Oct. 20, 2010).[8] 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 SILs … for purposes 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 (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 the 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, considering the interpretation of the phrase "cause, or contribute to" in section 165(a)(3) described above and the intended role and function of SILs, it is not necessary for permitting authorities to cite inherent *de minimis* exemption authority to justify the conclusion that a proposed source with an insignificant impact on air quality does not cause or contribute to

[8] In response to a challenge to the 2010 rulemaking in the District of Columbia Circuit, the EPA requested that the court remand and vacate two of the EPA's SILs regulations for $PM_{2.5}$ so that the 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 noted the EPA's statement in its brief that "the regulatory text it adopted does not allow permitting authorities the discretion to require a cumulative impact analysis, notwithstanding that the source's impact is below the SIL, where there is information that shows the proposed source would lead to a violation of the NAAQS or increments." *Id.* at 464. The court then vacated the two $PM_{2.5}$ SIL provisions "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." Although a rulemaking has not been conducted to date, as discussed below, a permitting authority has discretion to conclude that a proposed source does not cause or contribute to a violation if its predicted impact on air quality concentrations for the relevant pollutant is not significant or meaningful. A permitting authority also has discretion to require other appropriate modeling analyses or information from the permit applicant to make the demonstration required under 42 U.S.C. § 7475(a)(3).

a violation of the NAAQS or PSD increment within the meaning of section 165(a)(3) of the Act.[9] The air quality concentration levels that the EPA has identified as SILs do not function to exempt a source from making the demonstration required by section 165(a)(3) of the Act. Rather, these concentration levels provide a streamlined 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 in the ambient air, the permitting authority may conclude that the applicant has made a demonstration that its increased emissions will not cause or contribute to any air pollutant concentrations that violate the relevant NAAQS or PSD increment. In many circumstances this demonstration can be made by showing through modeling that projected air quality impacts from emissions from the proposed source will fall below the relevant SIL, but permitting authorities have the discretion to require further information or a cumulative impact analysis.

As discussed above, the phrase "cause, or contribute to" in section 165(a)(3) of the Act is reasonably read in context to not apply to impacts on air quality that are not meaningful or significant. In order to show that a particular degree of change in concentration is not meaningful or significant 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

[9] Although the EPA emphasized its inherent authority to establish a *de minimis* exception to a statutory requirement in several actions on the topic of SILs between 2006 and 2012, EPA also continued to recognize in these actions that phrase "cause or contribute" could be construed to exclude insignificant impacts and that a demonstration that the impacts of a source are insignificant can be used to satisfy (rather than avoid) the statutory requirement in section 165(a)(3) of the Act. In its *Prairie State* opinion, the EAB described how the EPA has interpreted the phrase "cause, or contribute to" in section 165(a)(3) to refer to significant emission contributions. *Id.* at 105. In its 2007 proposal of the $PM_{2.5}$ SILs rule, the EPA said that when "a source can show that its emissions alone will not increase ambient concentrations by more than the SILs, 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 expressed similar thoughts in a guidance memorandum. *See* 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, 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."

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 meaningful or significant. In general terms, a trivial or *de minimis* impact on air quality may be considered "meaningless" or "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.

Nevertheless, 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. Given the statutory considerations discussed above, a permitting authority is not required to conclude that any level of ambient impact from a source located in an attainment area automatically "causes or contributes" to a violation. A permitting authority has discretion to conclude that a proposed source does not cause or contribute to a violation if its predicted impact on air quality concentrations for the relevant pollutant is not meaningful or significant. 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 used by the permitting authority is representative of a level below which the projected impact of a proposed new or modified stationary source is not meaningful or significant. *See Rocky Mountain Steel* Order at 18; 2010 NO_2 Guidance at 11. Where SIL values developed by EPA are used to show that a source does not cause or contribute to a violation, this permit-specific record can incorporate the information and technical analysis provided by the EPA to show that a source with a projected impact below the relevant SIL value will not cause or contribute to a violation of the NAAQS or PSD increment. If a permitting authority elects to apply its own SIL value to support a permitting decision, the permitting record should reflect information independently compiled by a permitting authority to make the same showing with respect to that value.

Technical Basis for the EPA's Development of the Significant Impact Thresholds for $PM_{2.5}$ and Ozone

EPA-454/R-18-001
April 2018

Technical Basis for the EPA's Development of the Significant Impact Thresholds for $PM_{2.5}$ and Ozone

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Assessment Division
Research Triangle Park, NC

Contents

1.0 Introduction:

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 calculate levels of change in air quality concentrations that have a "significant impact" or an "insignificant impact" on air quality degradation. *Statistical significance* is a well-established concept with a basis in commonly accepted scientific and mathematical theory. This analysis examines *statistical significance* for a range of values measured by air quality monitors. The statistical methods and data reflected in this analysis may be applicable for multiple regulatory applications where EPA and state agencies seek to quantify a level of impact on air quality that they consider to be either "significant" or "not significant." Note: We have adopted the following convention throughout the document: a "significant impact" (in quotes) refers to a level of air quality change that can be used in the permit analysis of the ambient impacts from a facility to determine if it "causes, or contributes to" a violation of the applicable National Ambient Air Quality Standards (NAAQS) or Prevention of Significant Deterioration (PSD) increment, whereas we use *significant* (italicized) to refer to a mathematical assessment of probabilistic properties.

While this technical analysis may have utility in several contexts, the primary purpose of this document is to quantify the degree of air quality impacts corresponding to different confidence intervals (related to the statistical analysis presented here) that can be used in determining what is an "insignificant impact" when considering an application for a permit under the 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 NAAQS or PSD increment.[1] One way that this criterion can be met is by showing that the increased emissions from a proposed source will not have a significant impact on ambient air quality at any location, including locations where an exceedance of the NAAQS or PSD increment is occurring or may be projected to occur.[2] For the purposes of a PSD permit, the EPA has promulgated analytical methods involving air quality modeling and monitoring for conducting these compliance demonstrations.[3] More generally (*e.g.*, for purposes of designating areas as attainment or nonattainment), 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.[4] A DV is a statistic or summary metric based on the most recent one or three years (depending on the specific standard) of

[1] 40 Code of Federal Regulations (CFR) 51.166 and 52.21.
[2] Memorandum from Peter Tsirigotis, EPA Office of Air Quality Planning and Standards, Guidance on Significant Impact Levels for Ozone and Fine Particles in the Prevention of Significant Deterioration Permitting Program, April 17, 2018.
[3] 40 CFR, part 51, Appendix W, 82 FR 5182 (January 17, 2017), Revisions to the Guideline on Air Quality Models: Enhancements to the AERMOD Dispersion Modeling System and Incorporation of Approaches to Address Ozone and Fine Particulate Matter.
[4] 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*.

monitored data that describes the air quality status of a given location relative to the level of the NAAQS.

The EPA has decided that an "insignificant impact" level of change in ambient air quality can be characterized by the observed variability of ambient air quality levels. Since the cause or contribute test is applied to the NAAQS in the PSD program, 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 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 a change in air quality that is not significant.[5] 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).[6]

Based on these observed ambient data, the EPA has estimated the variability 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.[7,8] The bootstrap approach applied here uses a non-parametric, random resampling with replacement on the sample dataset (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 within the environmental sciences.[9,10,11,12] For example, bootstrapping has been used to evaluate the economic value of

---

[5] This approach is applied here strictly for the purpose of section 165(a)(3) and no other parts of the Clean Air Act.
[6] The 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*.
[7] Efron, B. (1979); Bootstrap methods: Another look at the jackknife. The Annals of Statistics 7 (1): 1–26. doi:10.1214/aos/1176344552.
[8] Efron, B. (2003); Second Thoughts on the Bootstrap. Stat. Sci., 18, 135-140.
[9] Schuenemeyer, J., Drew, L. (2010); Statistics for Earth and Environmental Scientists, John Wiley & Sons, Inc. *http://dx.doi.org/10.1002/9780470650707.ch3*.
[10] Park, Lek, Baehr, Jørgensen, eds. (2015); Advanced Modelling Techniques Studying Global Changes in Environmental Sciences, 1st Edition, Elsevier. ISBN 9780444635365.
[11] 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.
[12] Mudelsee, M. & Alkio, M. (2007); Quantifying effects in two-sample environmental experiments using bootstrap confidence intervals, Env. Mod. & Software, 22, 84-96.

clinical health analyses[13] and environmental policies,[14] in evaluations of environmental monitoring programs,[15] and in determining uncertainty in emissions inventories.[16] 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.[17,18]

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 an "insignificant impact" for PSD applications. This approach 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, *not significant* at the given confidence level. Specifically, the analysis uses 17 years (2000-2016) 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 from which the appropriate design values are calculated. The DVs from the 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. As discussed in Section 4.1.1 of this document and in the Policy Document,[2] the 50% CI was chosen to quantify the bounds of a change in air quality that can be considered an "insignificant impact" for the purposes of meeting requirements under the PSD program.

This technical basis document explains the analysis design and results provide the EPA's rational basis to recommend Significant Impact Levels (SILs) values that can be applied as a tool for making the PSD compliance demonstration required by the Clean Air Act (CAA) and PSD regulations. 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 the implications of the geographical analysis performed in response to peer reviewer

[13] Campbell, M., & Torgerson, D. (1999); Bootstrapping: Estimating Confidence Intervals for Cost-effectiveness Ratios, Q. J. of Med., 92, 177-182.
[14] 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.
[15] Levine, C., et al (2014); Evaluating the efficiency of environmental monitoring programs, Ecol. Ind., 39, 94-101.
[16] Tong, L., et al (2012); Quantifying uncertainty of emission estimates in National Greenhouse Gas Inventories using bootstrap confidence intervals, Atm. Env., 56, 80-87.
[17] Hanna, S. (1989); Confidence limits for air quality model evaluations, as estimated by bootstrap and jackknife resampling methods, Atm. Env., 6, 1385-1398.
[18] Cox, W. & J. Tikvart (1980); A statistical procedure for determining the best performing air quality simulation model, Atm. Env., 9, 2387-2395.

comments. The resulting values chosen by the EPA can serve as SIL 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 for multiple purposes, including to determine compliance with the various NAAQS. In addition, the monitoring network is used 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 part 58, Appendix D (Network Design Criteria for Ambient Air Quality Monitoring). These general requirements and choices made by the state and local air agencies conducting monitoring 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 measured 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.[19] 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.)[20]
- For the primary annual $PM_{2.5}$ NAAQS, the DV is the 3-year average of the $PM_{2.5}$ annual mean mass concentrations.[21] The annual mean is defined as the mean of the data in each of the 4 quarters of the year (*i.e.*, the mean of the quarterly means). 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 $\mu g/m^3$.[22]

[19] Appendix U to Part 50 - Interpretation of the Primary and Secondary National Ambient Air Quality Standards for Ozone.
[20] National Ambient Air Quality Standards for Ozone, 80 FR 65292 – 65468 (Oct. 26, 2015).
[21] Appendix N to Part 50—Interpretation of the National Ambient Air Quality Standards for $PM_{2.5}$.
[22] There is a secondary $PM_{2.5}$ NAAQS, with a level of 15.0 $\mu g/m^3$. The work here focuses only on the primary NAAQS at 12.0 $\mu g/m^3$, since compliance with the primary standard explicitly implies compliance with the secondary standard as well.

- 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.[23] 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 exceed 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 2016, 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.

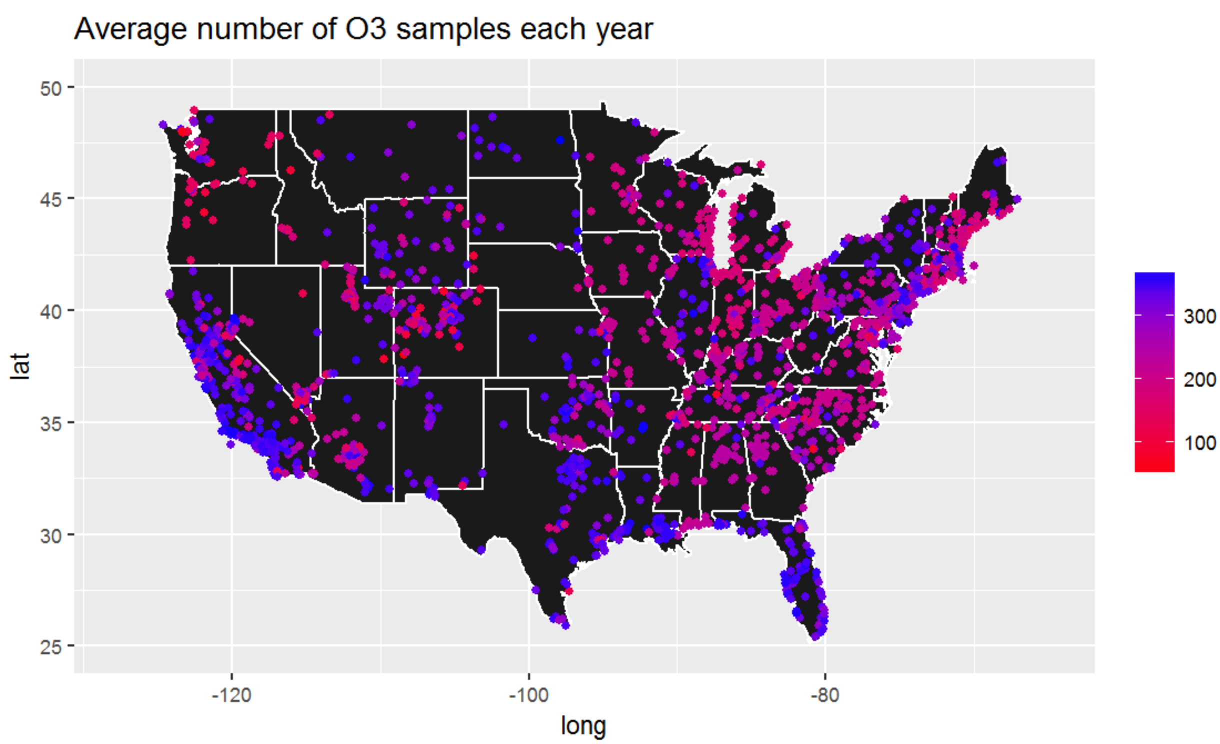


Figure 1 - Location and average number of monitored ozone days each year from the ozone sampling network for the years 2000-2016.

[23] 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*.

2.1.2 $PM_{2.5}$ Monitoring Network

The $PM_{2.5}$ monitoring network consists of two types of monitors: Federal Reference Method (FRM)[24] and FEM[23] 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 the EPA. Newer FEM monitors are "continuous" monitors that can provide hourly (or shorter) $PM_{2.5}$ measurements and have undergone testing to demonstrate conformance (including linear regression, slope/intercept, time series, and mean concentration ratios) with the FRM monitors.[25] 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. Although the FRM and FEM monitors have small differences in their performance, the largest impact to the bootstrap technique of this transition from all FRM monitors to a mix of FRM and FEM monitors is the gradual increase in the frequency of $PM_{2.5}$ measurements over time. During the period of 2000 through 2016, 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.

[24] Appendix B to Part 50—Reference Method for the Determination of Suspended Particulate Matter in the Atmosphere (High-Volume Method).
[25] Noble, C. A. et al (2001); Federal Reference and Equivalent Methods for Measuring Fine Particulate Matter, Aerosol Sci. & Tech, 34:5, 457-464.

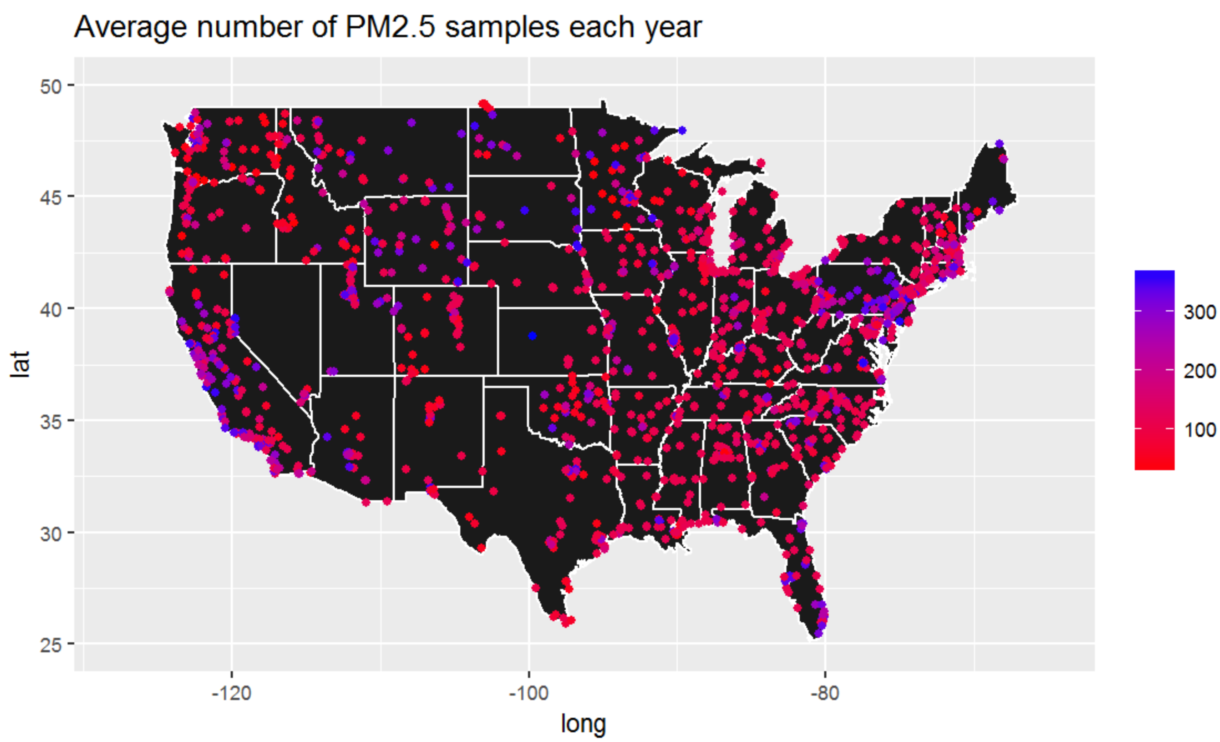


Figure 2 - Location and average number of monitored PM days each year from the $PM_{2.5}$ sampling network for the years 2000-2016.

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 represents. The monitor spatial scales are defined in 40 CFR part 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 $PM_{2.5}$, 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).[6] 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. Assuming 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.[26] *Non-significance* simply shows the data to be compatible with the null hypothesis under the set of assumptions associated with the statistical test.[26] A CI can be used as a mathematically equivalent procedure[26] to a formal hypothesis test for significance. CIs are constructed based on the desired confidence level and characteristics of the *sample*, including the sample variance, to determine error bars for the statistic of interest, such as the mean. Error bars constructed in this fashion are referred to as CI because they convey the confidence in the *sample* estimate of the *population* 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 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 CIs 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

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

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.[27] 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 relate these statistical tests to a practical application, we continue the hypothetical example from above:

> Suppose that the observed 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 Gaussian 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 CI 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. The average of the sample means is 6.98 µg/m^3. In order to determine the 95% CI, the data are 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).

---

[27] 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.

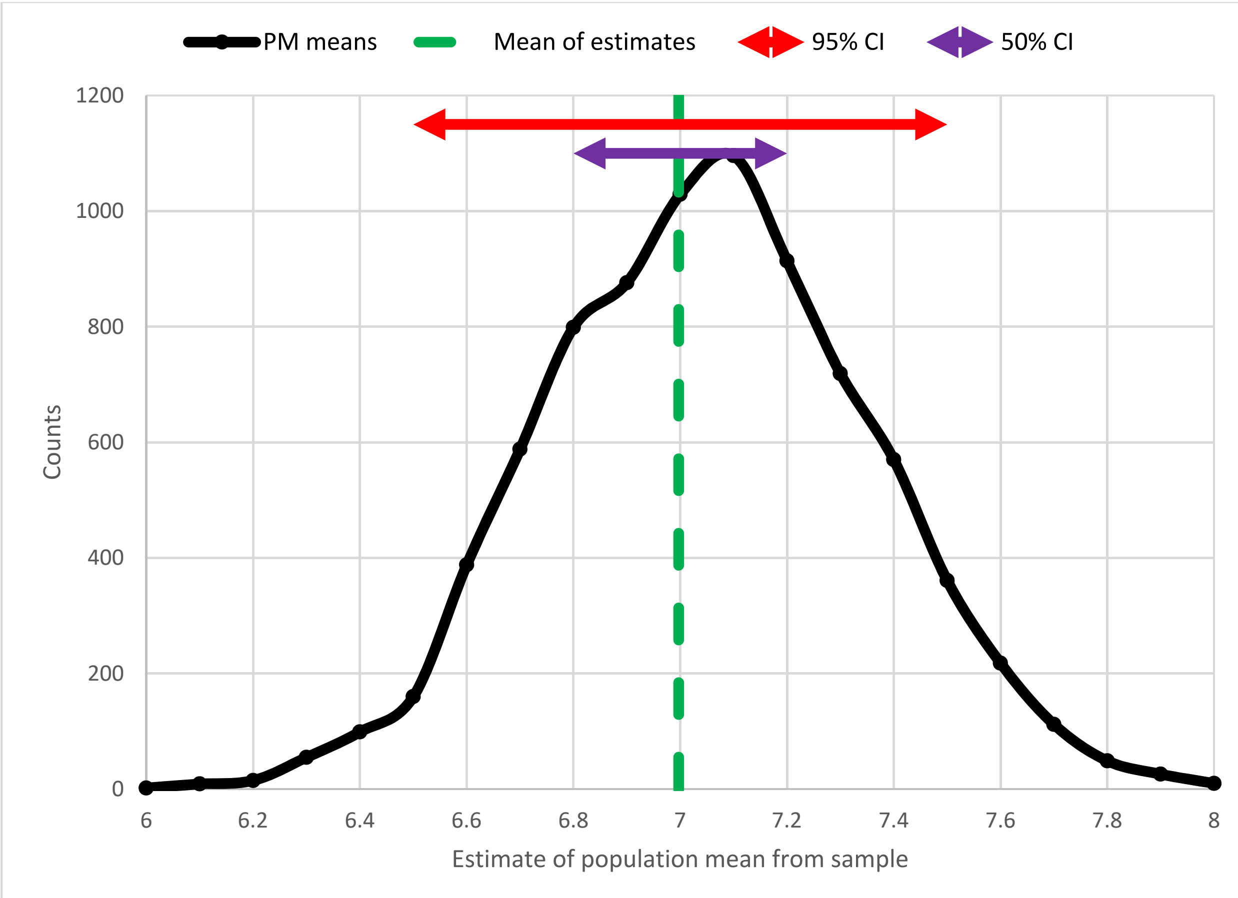


Figure 3- Example of CIs determined from a distribution of sample means.

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 CIs 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% CI. 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 in an area 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 population that sample represents (*e.g.*, a sample from an area-wide monitor theoretically represents the population of air quality across a wide area), and this interpretation has implications for the determination of the uncertainty associated with the sample (*e.g.,* a sample from an area-wide monitor is less likely to accurately represent air quality across the whole area at any moment, thus having greater uncertainty as to its ability to characterize the population of air quality it is intended to represent). 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 (as was the case in the example above).

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 result 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 intended to increase the stability (or reduce the variability) of the DVs.

The importance of temporal variability is perhaps more apparent when the application of the DVs are considered. For area designations purposes, the DVs are historical (updated DVs for a particular year are published in the following calendar year), such that the DV is an estimate of the current state of the air quality in an area. Furthermore, in the permitting process, DVs 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 for the DV contains the value of interest, then the DV 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 not significant* 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.[26]

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 the statistic of interest has a variance that is difficult to estimate with parametric assumptions, such as a rank order statistic, 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 conform to some standard distribution (*e.g.*, log-normal). For small sample sizes, other non-parametric tests such as the Mann-Whitney[28] test or the Wilcoxon signed-rank test[29] 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.[30] 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 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

[28] 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.
[29] Wilcoxon, F. (Dec 1945). Individual comparisons by ranking methods. Biometrics Bulletin 1 (6): 80–83.
[30] 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.

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.[7, 8, 9] 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).[9] 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 and spatial variability in the 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 of air quality variability are used to calculate levels of change in pollutant concentrations that can serve as “significant impact” thresholds in the context of source-specific “cause or contribute” determinations.

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 2016.[31]

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

[31] Raw daily and hourly measurements from FRM and 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 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.

each calendar year. 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.[19,21]

- 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 three annual means. Design values for the annual $PM_{2.5}$ NAAQS were rounded to the tenth $\mu g/m^3$ (*i.e.,* the one decimal), consistent with the computation of DVs for designation purposes.
- 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. Design values for the 24-hr $PM_{2.5}$ NAAQS were rounded to the nearest $\mu g/m^3$, consistent with the computation of design values for designation purposes.
- For the ozone NAAQS, all available data at each site were used. The ozone monitoring regulations require monitoring for the "ozone season," which varies by state. Many states operate a subset of ozone monitors outside of the required monitoring season and when those data are 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. Design values for the ozone NAAQS were truncated to the nearest ppb, consistent with the computation of design values for designation purposes. 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. During initial development of the method, 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 $PM_{2.5}$, 98^{th} percentile for $PM_{2.5}$, and 4^{th} 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 are 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 98^{th} percentile values in each year independently), then averaging the three 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%, 68%, 75% and 95% CIs for the mean,[32] 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 2012-2016).[33] 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[34] 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 98^{th} percentile as the 3^{rd} highest value in the dataset. From Y_0, we calculate the 98^{th} percentile as the 5^{th} highest value in the dataset. From Z_0, we calculate the 98^{th} percentile as the 7^{th} 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.

[32] 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.
[33] 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.
[34] Efron, B.; Tibshirani, R. (1993); An Introduction to the Bootstrap. Boca Raton, FL: Chapman & Hall/CRC. ISBN 0-412-04231-2.

- 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.
- 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 (1-*q*)% CI for the mean is the bounds of the center of the data that contains *q* percentage of the results (*i.e.*, the lower bound is the $(q/2)^{th}$ percentile and the upper bound is the $(1-q/2)^{th}$ 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 six 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 2014-2016 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%, 68%, 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 an apparent 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.

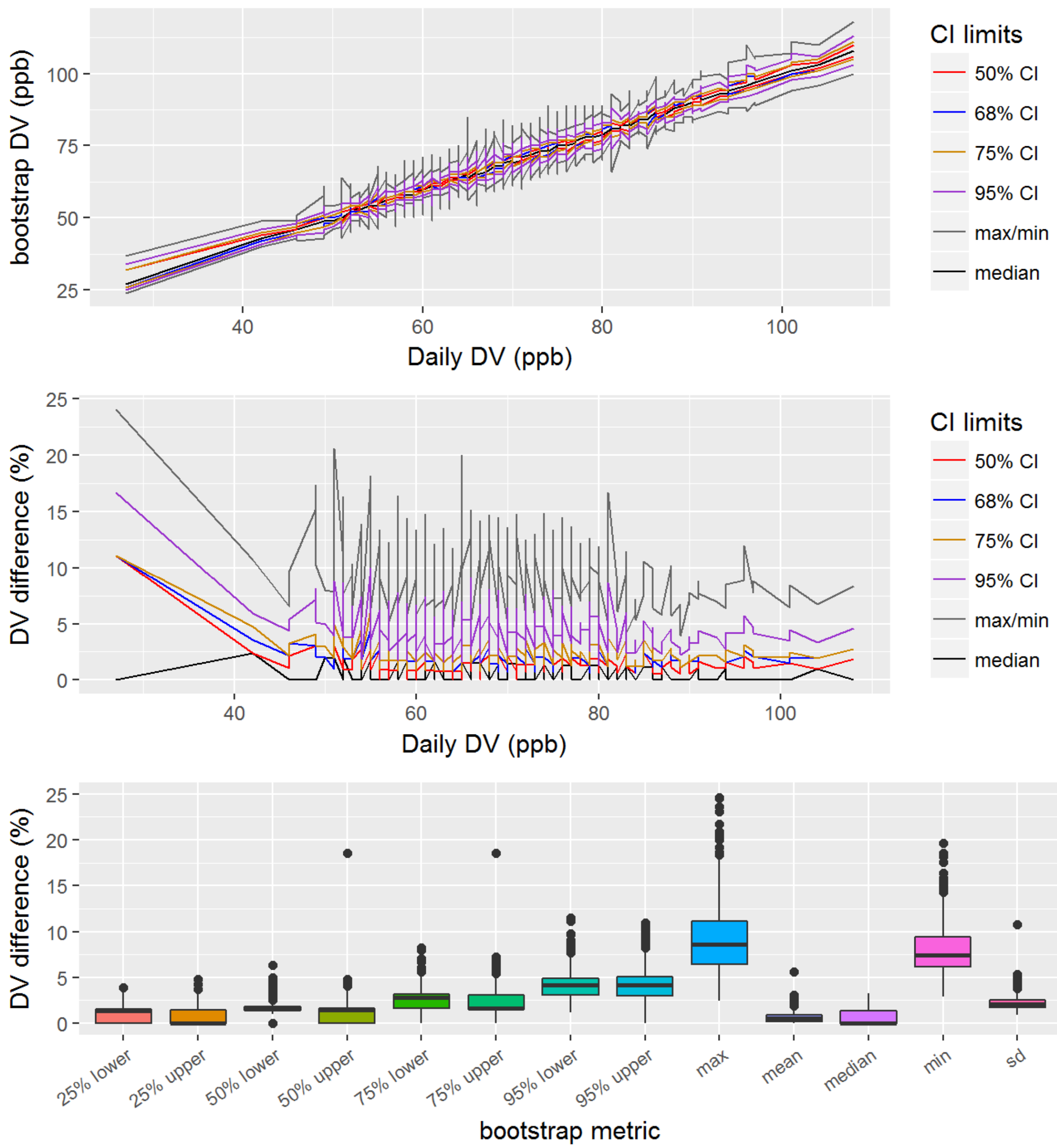


Figure 4 - Bootstrap results for the ozone 2014-2016 DVs (25%, 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs) Top panel shows the values for the DVs at the various CIs, the middle panel shows the average of the relative difference between the upper and lower bounds of the CI 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 2014-2016 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%, 68%, 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 surprising since the mean is 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 bounds of the bootstrapped CI 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%, 68%, and 75% CIs compared to the difference between these and the 95% CI. Similarly, for the 24-hr NAAQS, the values corresponding to the 50%, 68% and 75% CIs are fairly close to each other, with greater differences between these and the 25% CI on the low end and the 95% CI on the high end. 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. That is, there is no notable increase in the relative variability of the bootstrap DV as the actual DV increases. 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 ($\mu g/m^3$ for $PM_{2.5}$) that is appropriately used to characterize variability for sites with air quality that "just complies" with that NAAQS.

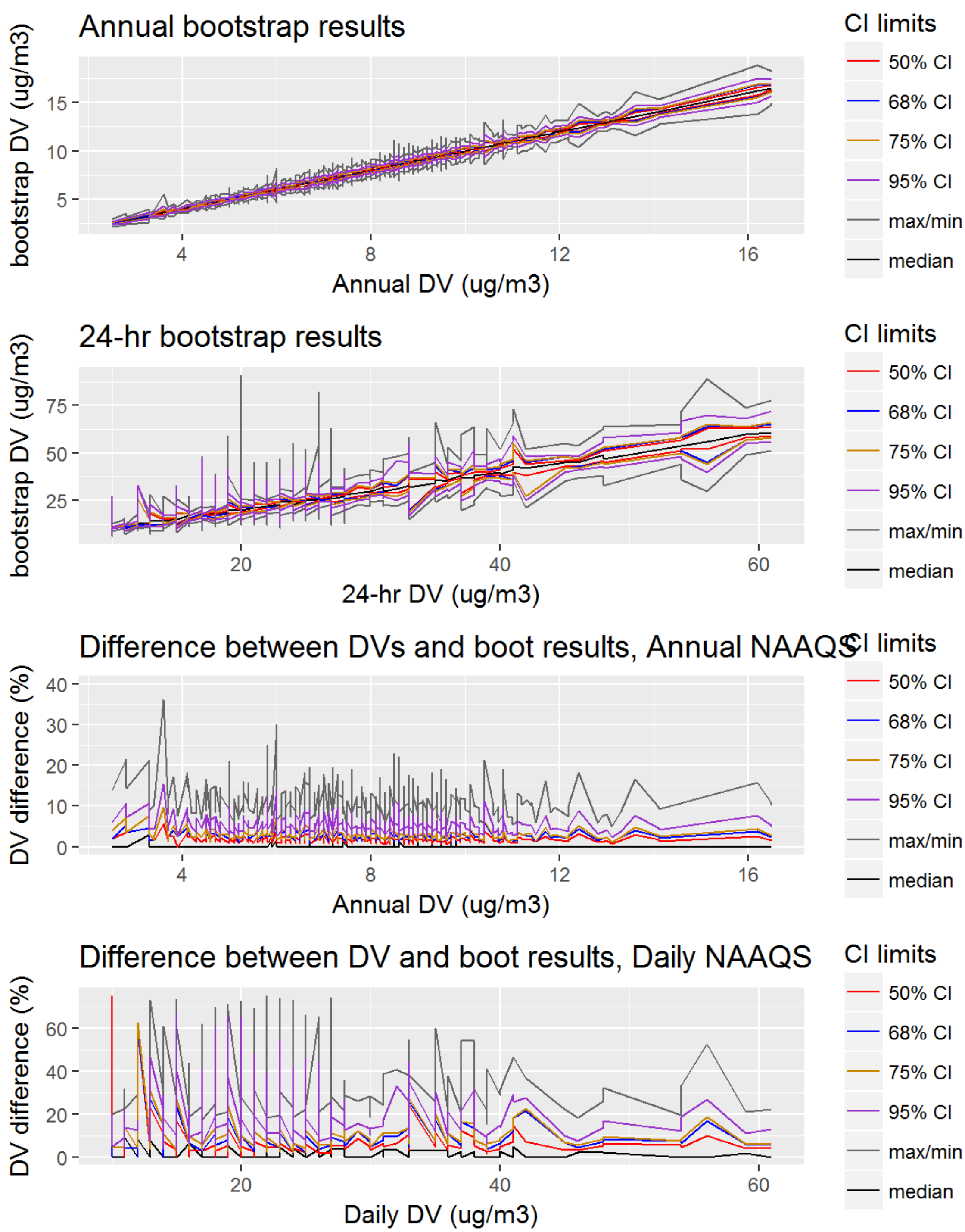


Figure 5 - Bootstrap results for the $PM_{2.5}$ 2014-2016 DVs (25%, 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs). The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the average of the percent difference between the upper and lower bounds of the CI and the actual DV.

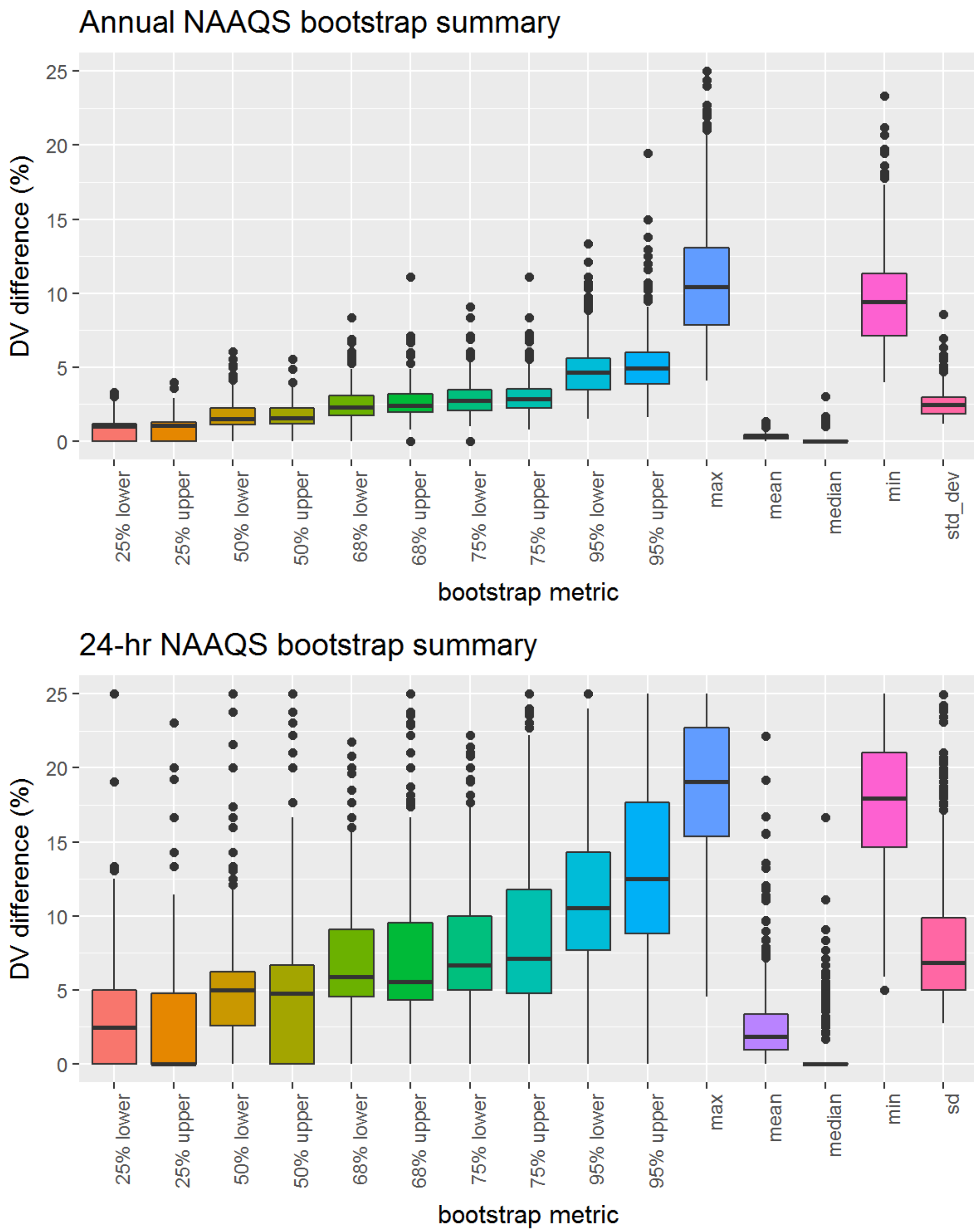


Figure 6 - Bootstrap results for the $PM_{2.5}$ 2014-2016 DVs, showing distribution of the relative differences between the upper and lower bounds of the bootstrap DVs and the actual DV at the 25%, 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

3.2.1 Analysis of $PM_{2.5}$ Spatial Variability

Section 2.1.3 discusses 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 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 2012-2016 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 fairly 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 contrast the results from the bootstrap analysis, which suggest less variability in the annual NAAQS than in the 24-hr NAAQS. 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 would 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 of $PM_{2.5}$ also suggests an important link to temporal variability of $PM_{2.5}$. 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, based on the results from the spatial analysis.

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	Monitor 2 ID	Annual DV 2	Delta (%)[35]
Minnesota	Washington	1.0	271630447	8.1 μg/m³	271630448	8.8 μg/m³	8%
Hawaii	Honolulu	1.7	150031001	4.9 μg/m³	150031004	5.6 μg/m³	14%
Pennsylvania	Philadelphia	2.6	421010047	10.3 μg/m³	421010057	10.9 μg/m³	5%
Pennsylvania	Philadelphia	3.1	421010055	11.6 μg/m³	421010047	10.3 μg/m³	12%
Louisiana	East Baton Rouge	5.4	220330009	9.0 μg/m³	221210001	9.2 μg/m³	3%
Nevada	Washoe	5.5	320310016	7.9 μg/m³	320311005	10.0 μg/m³	23%
Pennsylvania	Northampton	5.7	420950025	10.5 μg/m³	420950027	10.1 μg/m³	4%
Rhode Island	Providence	5.9	440070022	7.1 μg/m³	440071010	7.4 μg/m³	3%
Iowa	Clinton	6.4	190450019	10.6 μg/m³	190450021	9.4 μg/m³	11%
Utah	Salt Lake	7.3	490353006	9.2 μg/m³	490353010	9.7 μg/m³	5%
New Mexico	Bernalillo	7.9	350010023	6.5 μg/m³	350010024	6.3 μg/m³	3%
Indiana	Marion	8.9	180970078	11.1 μg/m³	180970081	11.8 μg/m³	6%
Indiana	Clark	9.3	180190006	11.8 μg/m³	211110067	11.3 μg/m³	4%
State	**City**	**Dist (km)**	**Monitor 1 ID**	**24-hr DV 1**	**Monitor 2 ID**	**24-hr DV 2**	**Delta (%)[35]**
Minnesota	Washington	1.0	271630447	20.6 μg/m³	271630448	21.1 μg/m³	3%
Hawaii	Honolulu	1.7	150031001	10.9 μg/m³	150031004	11.4 μg/m³	5%
Pennsylvania	Philadelphia	2.6	421010047	24.3 μg/m³	421010057	25.2 μg/m³	4%
Pennsylvania	Philadelphia	3.1	421010055	26.4 μg/m³	421010047	24.3 μg/m³	8%
Louisiana	East Baton Rouge	5.4	220330009	19.7 μg/m³	221210001	19.4 μg/m³	2%
Nevada	Washoe	5.5	320310016	26.8 μg/m³	320311005	31.5 μg/m³	16%
Pennsylvania	Northampton	5.7	420950025	27.2 μg/m³	420950027	28.3 μg/m³	4%
Rhode Island	Providence	5.9	440070022	18.3 μg/m³	440071010	18.6 μg/m³	2%
Iowa	Clinton	6.4	190450019	24.7 μg/m³	190450021	22.8 μg/m³	8%
Utah	Salt Lake	7.3	490353006	42.3 μg/m³	490353010	41.0 μg/m³	3%
New Mexico	Bernalillo	7.9	350010023	15.4 μg/m³	350010024	15.1 μg/m³	2%
Indiana	Marion	8.9	180970078	25.0 μg/m³	180970081	26.4 μg/m³	5%
Indiana	Clark	9.3	180190006	24.2 μg/m³	211110067	22.8 μg/m³	6%

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

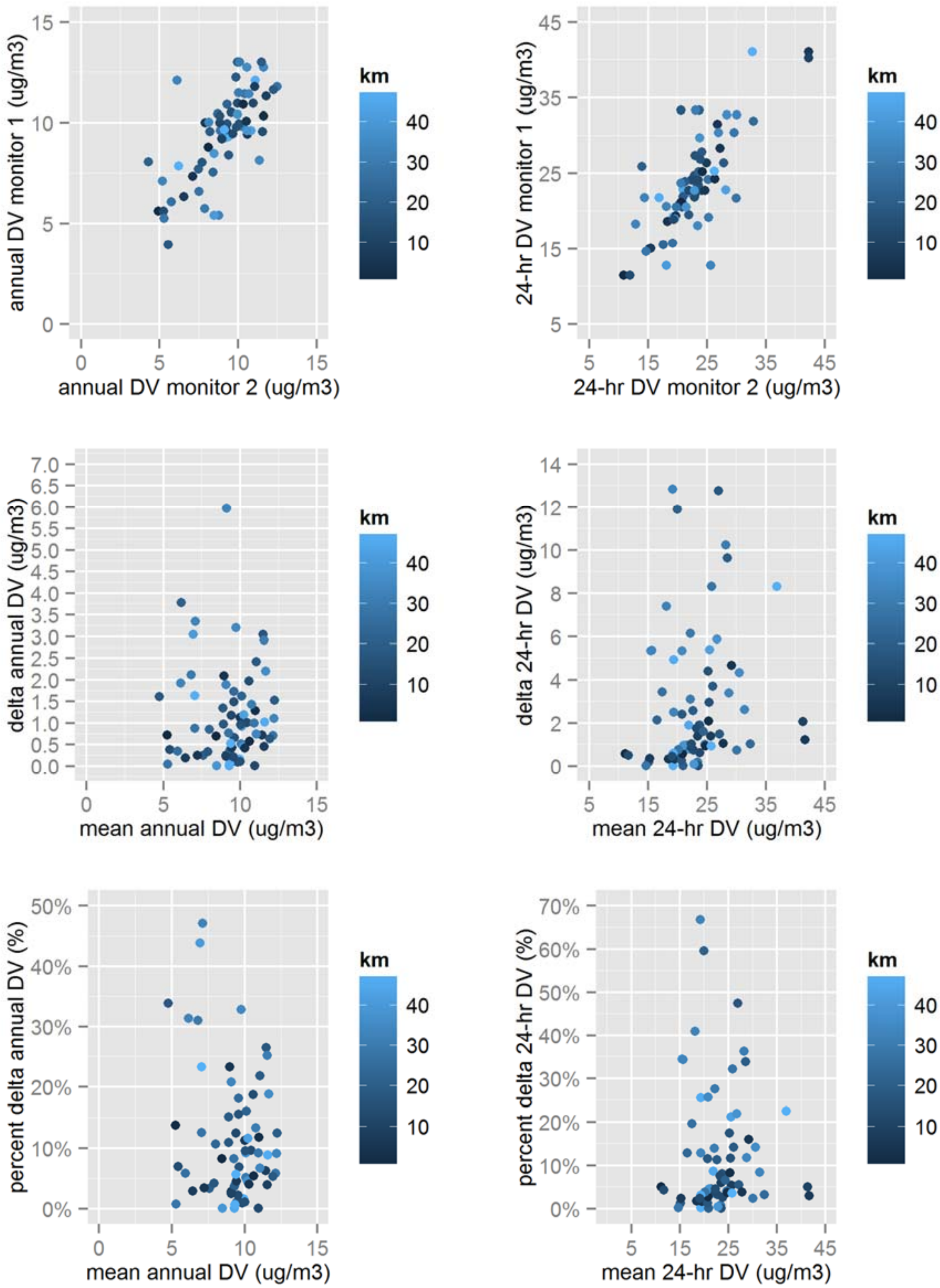


Figure 7 - Results from the analysis of spatial variability. Left column shows results for annual $PM_{2.5}$ NAAQS and the right column shows the results for the 24-hr $PM_{2.5}$ NAAQS.

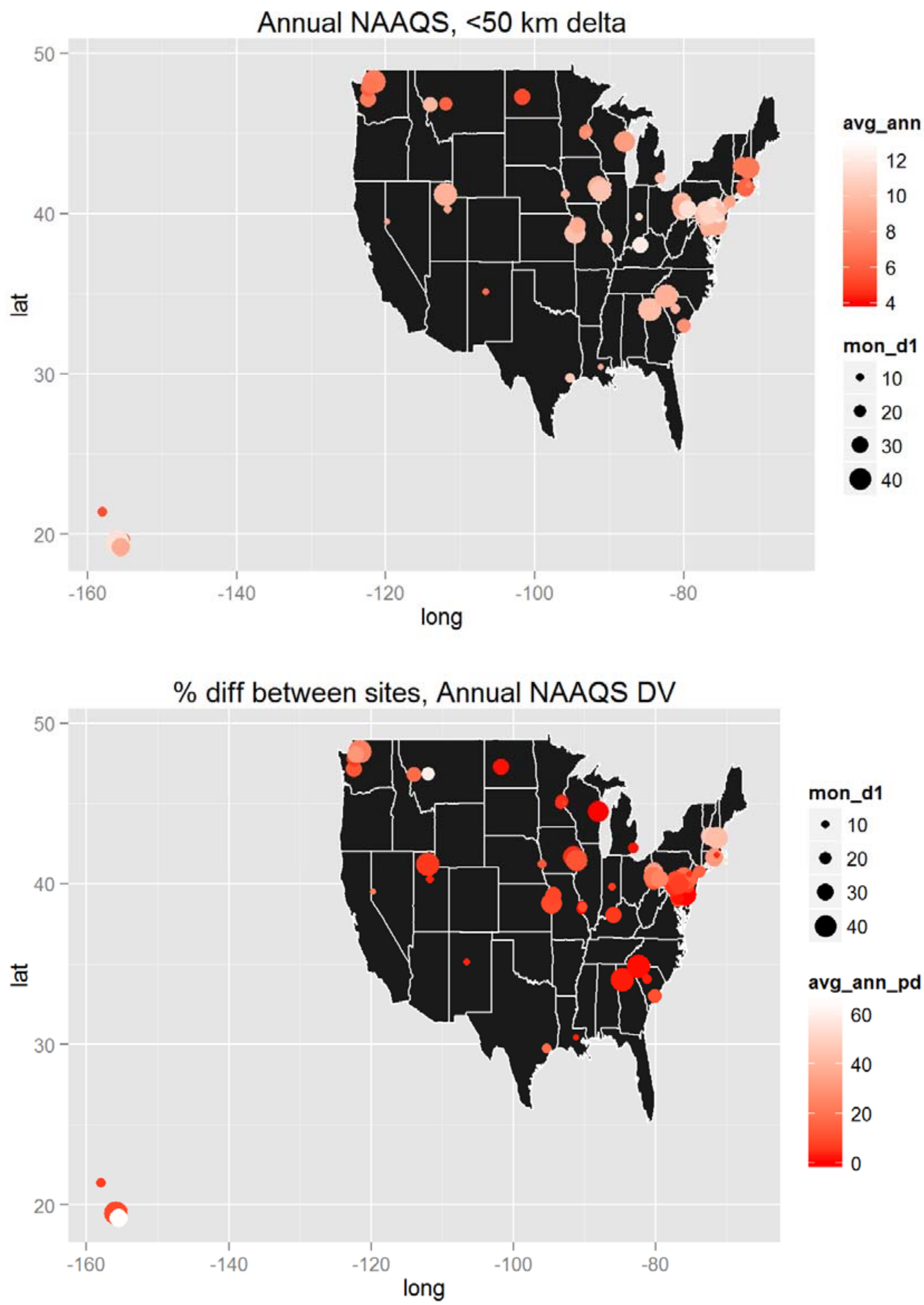


Figure 8 - Spatial distribution of the difference between the DVs from spatial analysis of the 2012-2016 $PM_{2.5}$ annual DVs. Top panel shows the absolute value of the difference between the two monitors while the bottom panel shows the percent difference between monitors.

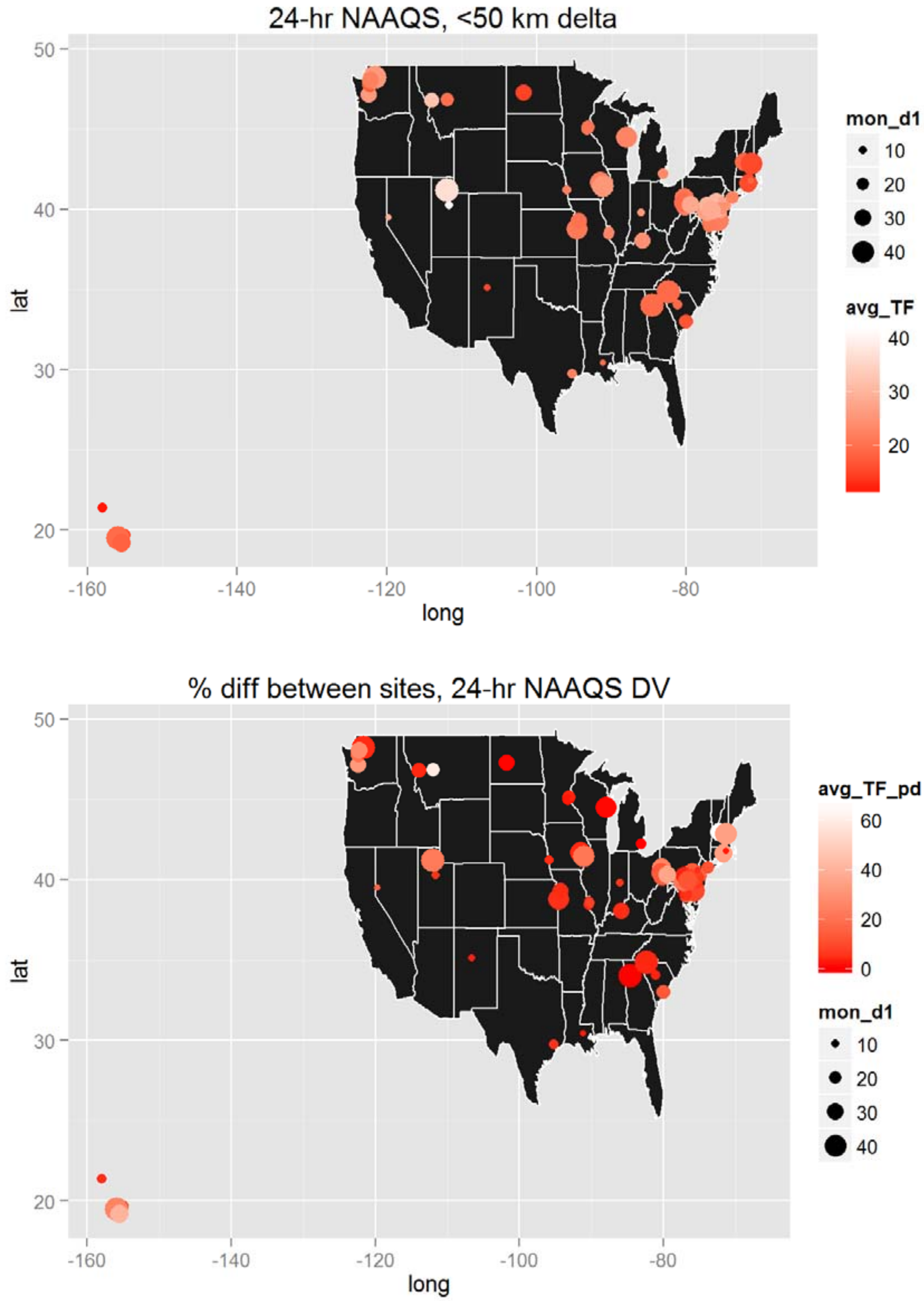


Figure 9 - Spatial distribution of the difference between the DVs from spatial analysis of the 2012-2016 $PM_{2.5}$ 24-hr DVs. Top panel shows the absolute value of the difference between the two monitors while the bottom panel shows the percent difference between the two monitors.

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

The PM monitoring network was 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 was 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 DVs 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, but 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 obviously 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 data from datasets from sites with less frequent sampling would in general have a higher sample variance and therefore wider confidence intervals. Sensitivity tests conducted with the 2010-2013 DVs indeed showed that statistics from the subset of sites with daily monitoring (1:1) have tighter confidence intervals 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 sampling. However, the results do support what is generally expected from reduced sampling frequency (*i.e.*, while 1:1 monitoring might capture a wider range of air quality, less frequent sampling would likely result in increased sample variance and wider confidence intervals for 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 to characterize air quality variability. Using only the 1:1 monitors would likely produce smaller estimates of the sample variance due to the increased sample size while possibly capturing a wider range of air quality across a more widely sampled spectrum. 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 presence of 1:3 and 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 PM 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			2015 annual			2016 annual		
Difference, median bootstrap vs actual	0.04%	0.02%	0.04%	0.03%	0.03%	0.06%	0.04%	0.03%	0.03%
Avg. 25% CI span	0.67%	0.57%	0.94%	0.70%	0.58%	0.91%	0.71%	0.62%	0.88%
Avg. 50% CI span	1.63%	1.14%	1.81%	1.65%	1.24%	1.85%	1.69%	1.22%	1.85%
Avg. 68% CI span	2.44%	1.72%	2.67%	2.46%	1.77%	2.74%	2.45%	1.79%	2.76%
Avg. 75% CI span	2.80%	1.92%	3.11%	2.83%	2.00%	3.09%	2.82%	2.08%	3.18%
Avg. 95% CI span	4.72%	3.33%	5.26%	4.86%	3.43%	5.38%	4.79%	3.47%	5.48%
Year/NAAQS	2014 24-hr			2015 24-hr			2016 24-hr		
Difference, median bootstrap vs actual	1.14%	0.67%	1.54%	1.36%	0.84%	1.78%	1.23%	1.01%	1.40%
Avg. 25% CI span	2.27%	1.89%	2.38%	2.27%	1.92%	2.50%	2.50%	2.17%	2.63%
Avg. 50% CI span	4.29%	2.94%	4.76%	4.17%	3.45%	4.65%	4.35%	3.13%	5.13%
Avg. 68% CI span	6.00%	4.76%	7.02%	6.25%	5.09%	7.14%	6.52%	5.00%	7.89%
Avg. 75% CI span	6.82%	5.36%	8.33%	7.50%	5.56%	8.33%	7.69%	5.77%	8.88%
Avg. 95% CI span	12.50%	9.40%	14.14%	12.50%	10.00%	14.81%	13.16%	9.62%	16.67%
Number of sites	507	182	274	531	210	270	535	237	240

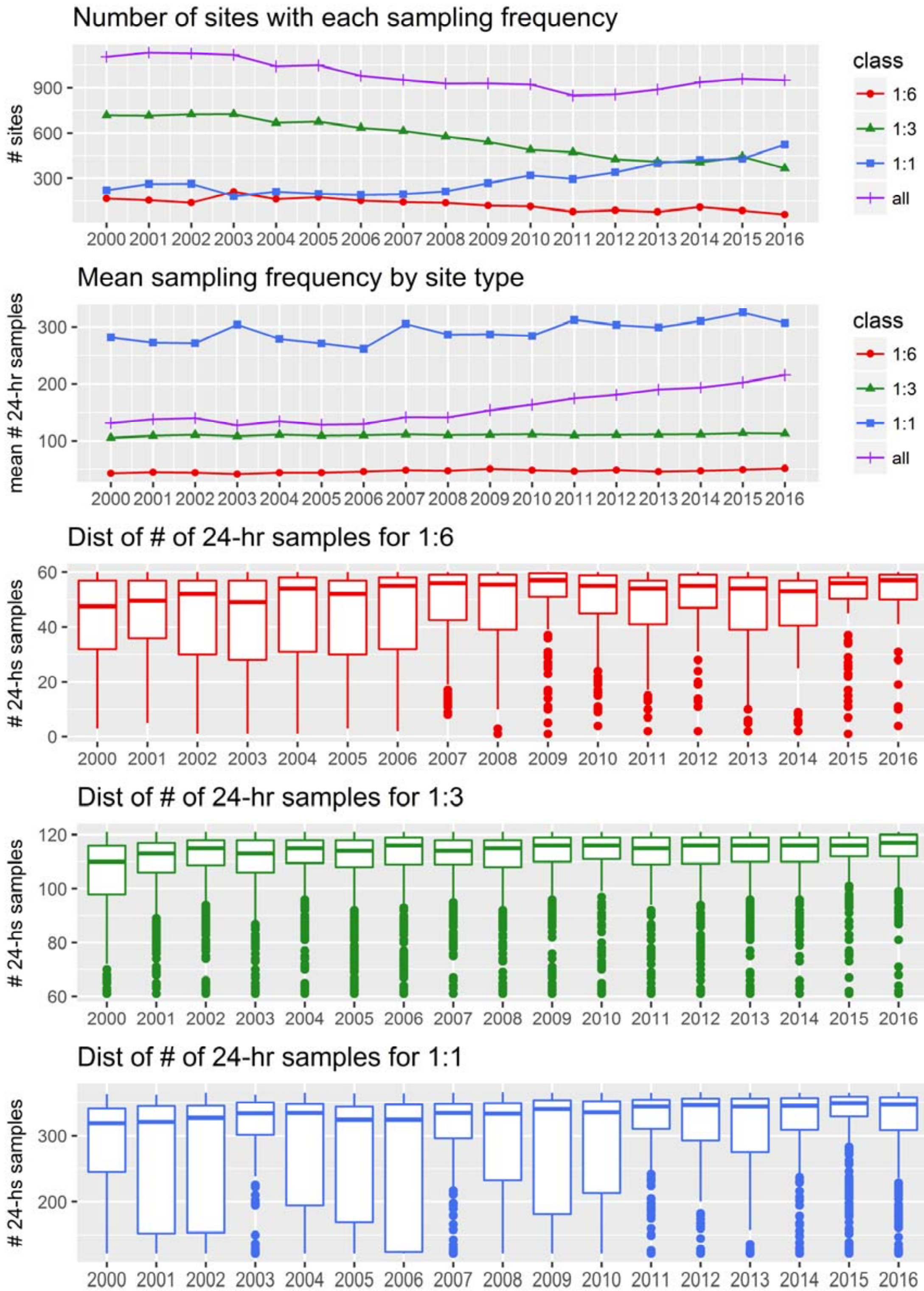


Figure 10 - $PM_{2.5}$ monitor network statistics. Top row shows the number of sites with each sampling frequency by year. Second row shows the average number of samples at each site type. Third, fourth and fifth rows show the distribution of the number of samples for each site type.

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

For a specific change in air quality concentrations to be used to show that a proposed source does not cause or contribute to a violation of the NAAQS, the concentration change must represent a level of impact on ambient air quality that is "insignificant" or not meaningful. The EPA has taken into account the necessary policy considerations in conjunction with the statistical analysis presented here to provide a rational basis to select values derived from the statistical analysis that can be applied to represent "insignificant impacts."

Section 3 presented the results from the bootstrap analysis, which produced variability estimates at the 25%, 50%, 68%, 75%, and the 95% CIs for all the AQS data across the U.S. from 2000-2016. This section presents the technical considerations related to the policy[2] considerations guiding the application of the above results to identify an appropriate SIL for each context, and the final values the EPA has selected from the study results.[36]

4.1 PSD Air Quality Analyses and Statistical Significance

The following four factors are important for EPA's choice of a SIL: determining a CI to represent the inherent variability for purposes of the NAAQS compliance demonstration, an approach for scaling 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. The EPA has balanced the necessary policy considerations in conjunction with technical information discussed here and in the Policy Document[2] to develop SIL values that represents, in the Agency's judgment, an appropriate measure of "insignificant impact" that can be used by PSD permitting authorities to determine if emissions from proposed construction will "cause or contribute" to a violation of the corresponding NAAQS.

4.1.1 Confidence Interval

The bootstrap analysis produced estimates for the 25%, 50%, 68%, 75%, and 95% CIs in order to characterize the range of the inherent variability and to provide options for selecting an appropriate "insignificant impact" level that will be applied to determine each SIL. The statistical framework that forms the basis for the bootstrap CIs can be related to more traditional assessments of statistical significance and statistical significance testing. In contrast to the usage here, the traditional application of statistical significance testing seeks to determine if a deviation from the base value is *significant* (rather than *not significant*, which is the usage here). In order to make this determination, larger CIs are typically selected (*e.g.*, 90-99%, which results in a

[36] The methods, analysis, and application to the PSD program was subject to a peer-review. The results of that peer-review and the subsequent changes to the analysis and the document are detailed in a companion report, U.S. EPA, 2018, Peer review report for the technical basis for the EPA's development of significant impact thresholds for $PM_{2.5}$ and ozone, RTP, NC, EPA 454/S-18-001, available from the U.S. EPA RTP library.

high level of confidence that a deviation from the base value is indeed *significant*). In practice, the smallest CI that might be considered for a similar *significance* determination would be the 68% CI, which corresponds to one standard deviation of the mean for a normally distributed sample. Thus, any deviation larger than the bounds of the 68% CI could traditionally be identified as a *significant* deviation from the mean. In this application for the PSD program, however, we are seeking for each NAAQS a value below which we can conclude that the change in air quality is "not statistically significant" (*i.e.*, that there will not be a notable difference in air quality after the new source begins operation). Thus, a CI that could potentially be considered to represent a *significant* value would not simultaneously be appropriate for identifying a value that is *statistically not significant*. As such, CIs used for identifying a value that is *not statistically significant* value should be below 68%. For the reasons described in the Policy Document, the 50% CI was chosen as the benchmark statistic from the bootstrap analysis to represent the recommended SILs in PSD permitting for ozone and $PM_{2.5}$ NAAQS.

4.1.2 Adjustment to the Level of the NAAQS

Since air quality variability may have different characteristics at different baseline air quality levels (e.g., areas with smaller DVs may have less variability than areas with higher DVs), it is reasonable to characterize the variation in the air quality across a range of air quality levels. Sections 4.2 and 4.3 present the 50% CI value on both an absolute scale (ug/m^3 and ppb) and a relative scale (percentage), where the relative variability is defined as the percent change from the base DV at each site. The figures in these sections indicate that there is less of a trend in the relative variability compared to the absolute variability, and no trend in the relative variability for the ozone DV 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). However, the relative variability was fairly consistent across the range of design values, suggesting a commonality in the relative variability across a wide range of geographic regions, chemical regimes, and baseline air quality levels. These results suggest that there is an inherent aspect to the variability, regardless of the baseline air quality. Thus, for reasons explained in the policy memorandum, the relative variability values are used for the SILs development.

4.1.3 Selection of a Geographic Scale

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.

The EPA recognizes that the air quality data and the nature of the emissions and chemical formation of ozone and $PM_{2.5}$ can impact areas differently and, thus, should be considered as part of this evaluation. The analysis presented in Sections 4.2 and 4.3 (Figures 11 and 14) examine the relative variability represented by the 50% CI to explore any spatial trends in the data. The analysis indicates that while there is evidence of local spatial correlation (*i.e.*, most areas have fairly similar levels of relative variability and that sites with higher variability are isolated), 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., the magnitude of the variability differs from site-to-site within a state or region with few instances of regional patterns and no strong instances of east/west or north/south trends.

The analysis shows that a small number of sites with particularly high variability have an effect on the average network-wide variability. A median network-wide variability is not overly influenced by a few outliers. Thus, for the reasons explained in the Policy Document, the median variability from the 50% CI from the entire U.S. ambient monitoring network is used to calculate SIL values.

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 2000-2016 (equivalent to DV years of 2002-2016). For all three NAAQS considered here, there are general downward trends in the computed variability across these years. Since the SILs should reflect the most representative state of the atmosphere, the analysis uses for each NAAQS 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 $PM_{2.5}$ NAAQS median 50% CI has increased, while the annual $PM_{2.5}$ and ozone NAAQS median 50% CIs have continued to decrease). Faced with a similar selection of DV periods for use in attainment demonstrations for nonattainment areas,[37] the EPA also recommended using the average of three DV periods to be used along with a modeling analyses. Thus, for the reasons explained in the Policy Document, the three most recent DV periods (*i.e.*, 2012-2014, 2013-2015, 2014-2016) were used for determining SILs for $PM_{2.5}$ and ozone.

4.2 Analysis for Ozone

Figure 11 shows, for each monitoring site, the half-width of the 50% CI divided by the actual design value, from the 2014-2016 data for the ozone NAAQS.[38] 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% (with one exception). The variability is fairly consistent across the range of baseline air quality levels, indicating that there is no particular trend with actual design value in the site level variability. The median and mean variability values are fairly similar

The spatial distribution of the relative variability from the 50% CI is also shown in Figure 11, with 2014-2016 DV period site data colored according to their relative variability and sites with insufficient data during this period in gray. There appears to be no notable large-scale spatial

[37] Draft Modeling Guidance for Demonstrating Attainment of Air Quality Goals for Ozone, $PM_{2.5}$, and Regional Haze. R. Wayland, AQAD, Dec. 3, 2014.
[38] 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.

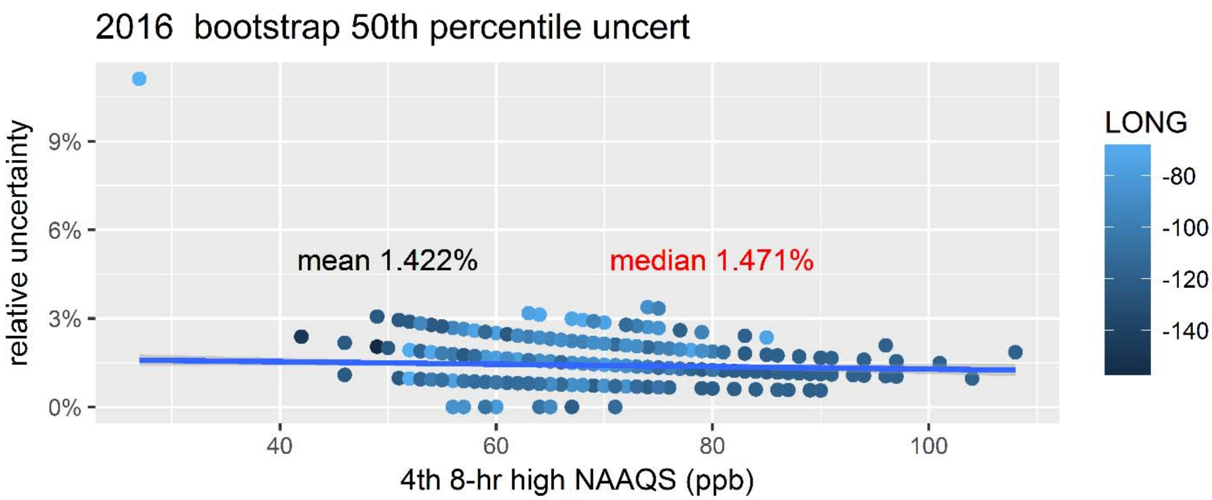
trends in highest relative variability. The lack of any large-scale spatial trend indicates that there is indeed a fundamental characteristic to the relative ambient air quality variability (see Section 4.1.3).

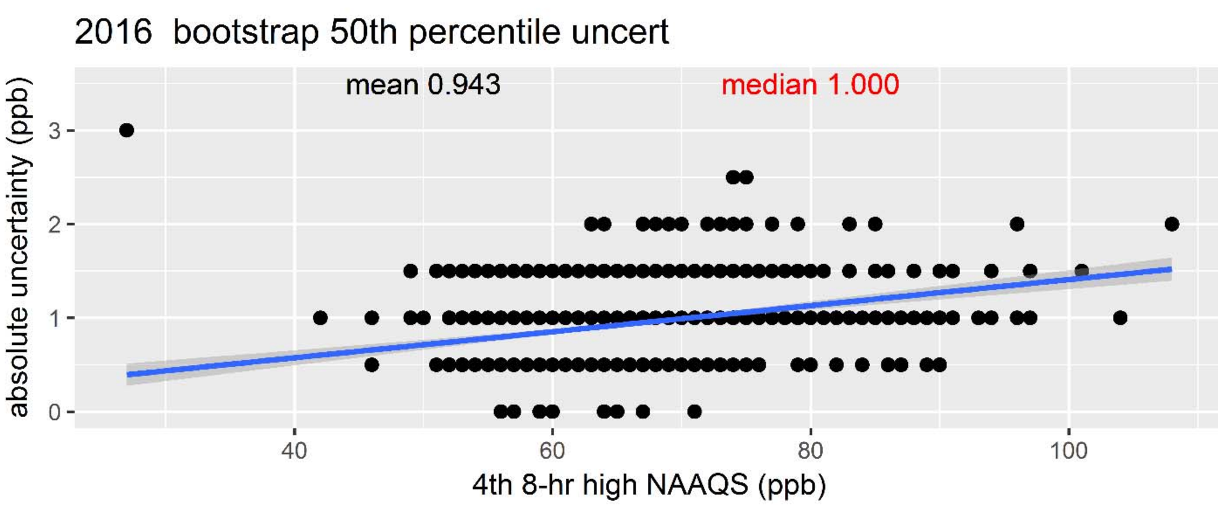
4.2.1 Ozone Temporal Trends

The median air quality variability from the 15 DV periods for ozone is shown in Figure 12 (each period is 3 years). 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.*, 2012-2014, 2013-2015, 2014-2016) as shown in Table 3, when averaged result in a SIL value for the ozone 8-hour NAAQS of 1.47%. This corresponds to 1.0 ppb at the level of the 2015 ozone NAAQS (70 ppb).

Table 3 - Summary of ozone bootstrap results for three design periods, 2012-2014, 2013-2015, and 2014-2016

Year/NAAQS	2014 annual	2015 annual	2016 annual
Difference, mean of median bootstrap vs actual DV	0.44%	0.48%	0.43%
Avg. 25% CI span	0.74%	0.76%	0.75%
Avg. 50% CI span	1.47%	1.47%	1.47%
Avg. 68% CI span	2.14%	2.05%	2.11%
Avg. 75% CI span	2.38%	2.34%	2.31%
Avg. 95% CI span	4.31%	3.97%	3.97%
Number of sites	1148	1131	1131





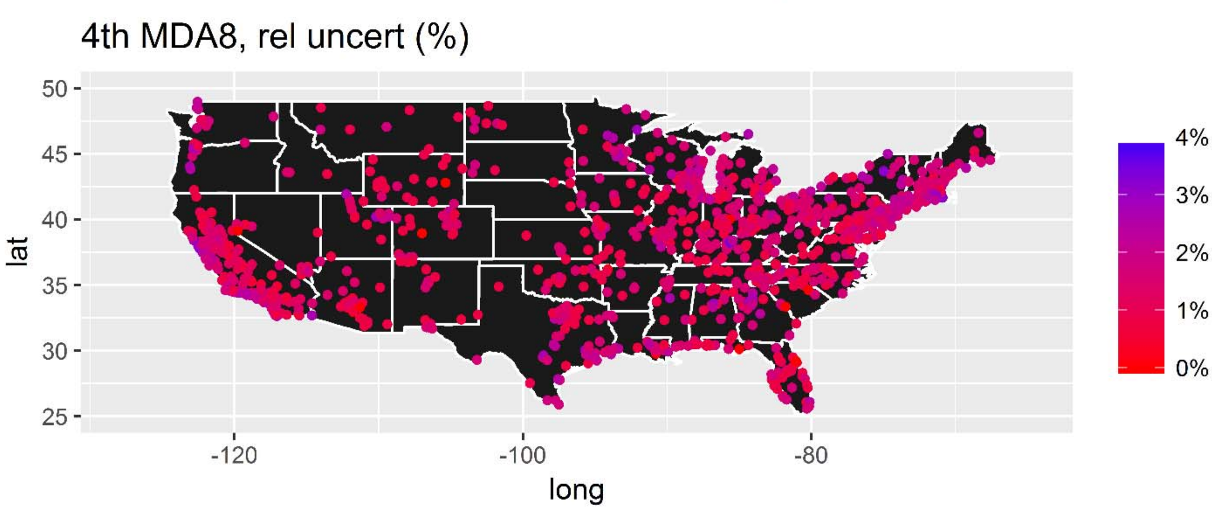


Figure 11 - Bootstrap results from the 50% CIs for the 2016 ozone DVs. The top panel shows the relative difference between the span of the CI and the actual DV across the range of actual DVs, the middle panel shows the absolute difference between the values across the same range, and

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

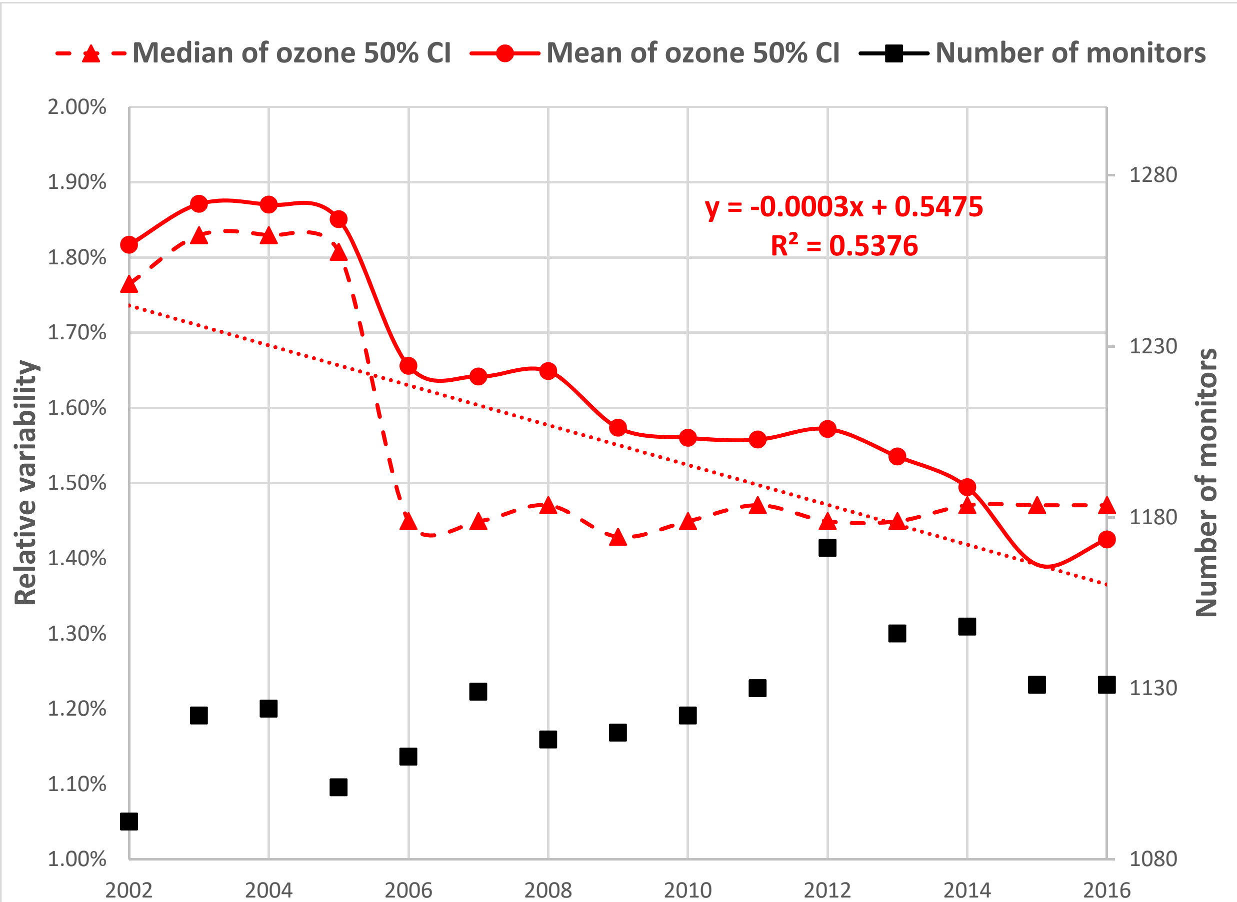


Figure 12 - Median and mean variability in the network determined from the bootstrap analysis for the 15 DV periods from 2002-2016 for ozone (each DV period represents 3 years of data and the data are plotted on the ending year, *i.e.*, the 2016 DV period is from 2014-2016 and plotted at 2016).

4.3 Analysis for $PM_{2.5}$

Figure 13 shows, for each monitoring site, the half-width of the 50% CI divided by the DVs, for both the annual and 24-hr $PM_{2.5}$ NAAQS. This figure shows that the relative variability using these assumptions is indeed stable across the range of baseline air quality levels, while the absolute variability increases as the baseline air quality levels increase.[39] 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 5%. For the 24-hr NAAQS, the data are

[39] The rounding conventions for $PM_{2.5}$ result in striations in the data, which are clearly visible in Figure 13. While these striations appear to represent trends in the data, this is a function of the display and not actual trends in the data. Linear regression lines have been added to each panel, which clearly show an increase in the absolute variability with increasing DVs, while the relative variability is relatively unaffected by changes in the DVs.

concentrated around 4-5%, with a small number of sites exceeding 10%. The outliers occur across the range of baseline air quality levels, indicating that there is no particular trend with actual DV 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. Unlike the ozone results, the median is smaller than the mean

The spatial distribution of the relative variability from Figure 13 is shown in Figure 14, with sites having data during the 2014-2016 DV period colored according to their relative variability (sites with insufficient data during the 2014-2016 DV period are not shown, data from other years are presented in the Appendix). Based solely a visual inspection, there appears to be no notable large-scale spatial trends in 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 generally not grouped into any specific geographic region. The exceptions to this appears in Western U.S. and along the Ohio River Valley, where there are a collection of sites with higher variability (generally above 7.5%) in the 24-hr NAAQS (though the annual NAAQS does not display this apparently higher variability). 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 variability in DVs computed from these data, particularly in the 24-hr $PM_{2.5}$ DVs). These sites also tend to have a lower sampling frequency (see Figure 2), which we have shown to artificially increase the apparent variability. There are also trends in missing data that are important to consider when exploring regional trends in variability. In particular, for the period 2008 through 2013, the data were invalidated for several states. 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 a number of years.[40]

In response to comments received during the peer-review of the initial public draft of this document, several more detailed spatial analyses are presented for the annual and 24-hr $PM_{2.5}$ data in Section 7 of the Appendix to this document. The analysis attempts to identify natural groupings of sites based on location and the level of air quality variability using cluster analysis. The analysis applied both an iterative (K-means) and a hierarchical clustering algorithm using various combination of the site-level variability, latitude, and longitude, resulting in 12 different sets of clusters. The analysis also considered comparing sites by grouping them using the National Oceanic and Atmospheric Administration (NOAA) "climate regions," which are groupings of states known by NOAA to have similar climatic conditions. While some of the analysis did identify some unique clusters, these groups were often not spatially grouped. Many

[40] The dates and specific monitors affected in each state vary. For DC, data were invalidated in Q4 of 2016. For FL, data were invalidated from 2011-2014. For GA, data were invalidated in Q1 of 2011. For ME, data were invalidated from 1998-2015. For ID, data were invalided from 2011-2014. For IL, data were invalidated from 2011-2013 and Q1-Q2 of 2014. For Louisville, KY, data were invalidated from 2009-2013. For the South Coast Air Basin, CA, data were invalidated in 2014. For MS, data were invalidated in 2014. For TN, data were invalidated from 2011-2014. For WA, data were invalidated from 2011-2015. The invalidation may not have affected every monitor in each state, but these dates cover the time spans for which the data invalidation occurred.

of the analyses did not identify any unique clusters. When the results from the special cluster analysis are considered as a whole, they do not indicate any consistent large-scale trends. The lack of any consistent regional trend indicates that there is indeed a fundamental characteristic to the relative ambient air quality variability (see Section 4.1.2).

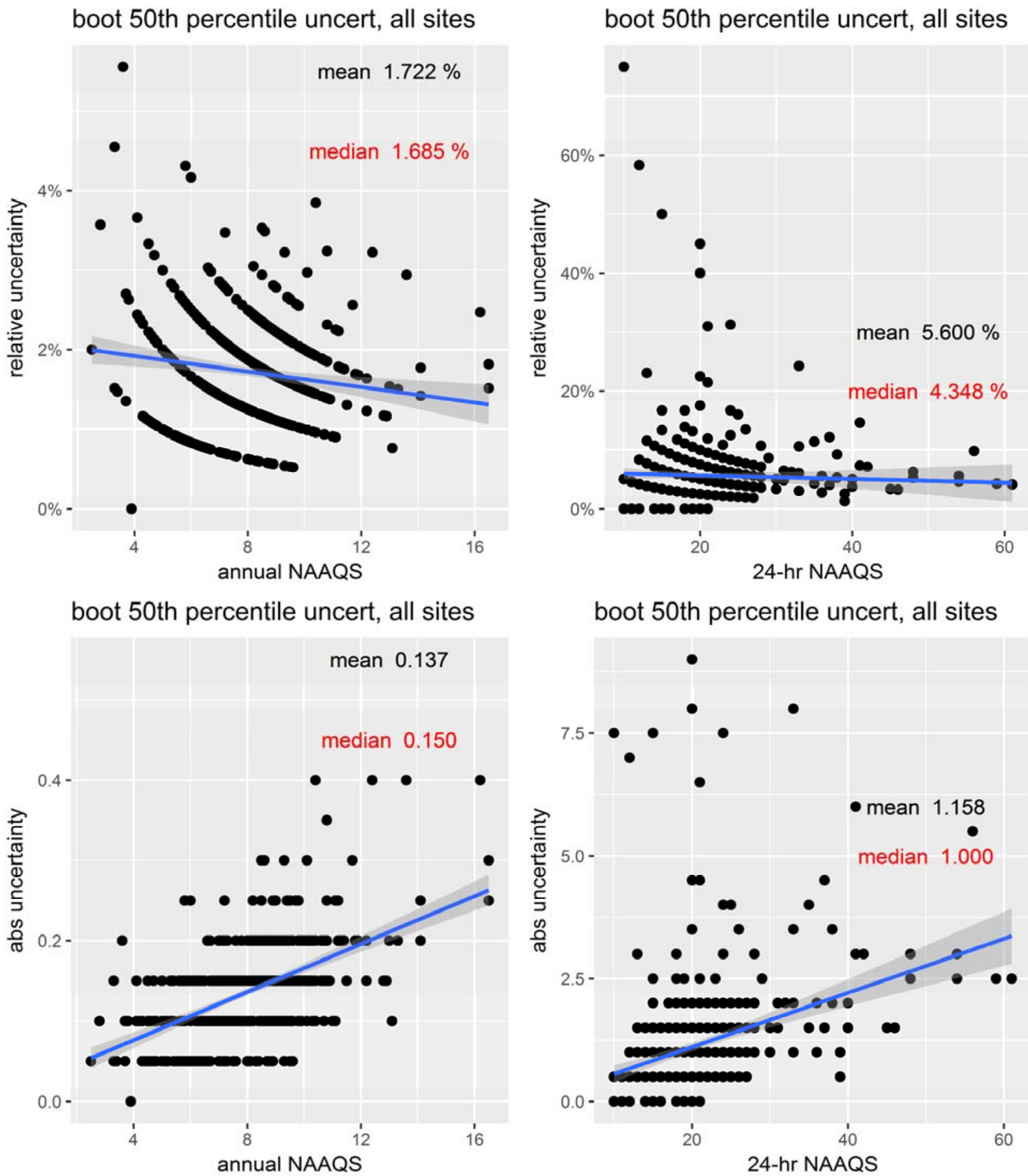


Figure 13 - Bootstrap results from the 50% CIs for the 2016 $PM_{2.5}$ DVs. The top two panels show the relative difference between the span of the CI and the actual DV across the range of actual DV, and the bottom two panels show the absolute difference between the values across the same range.

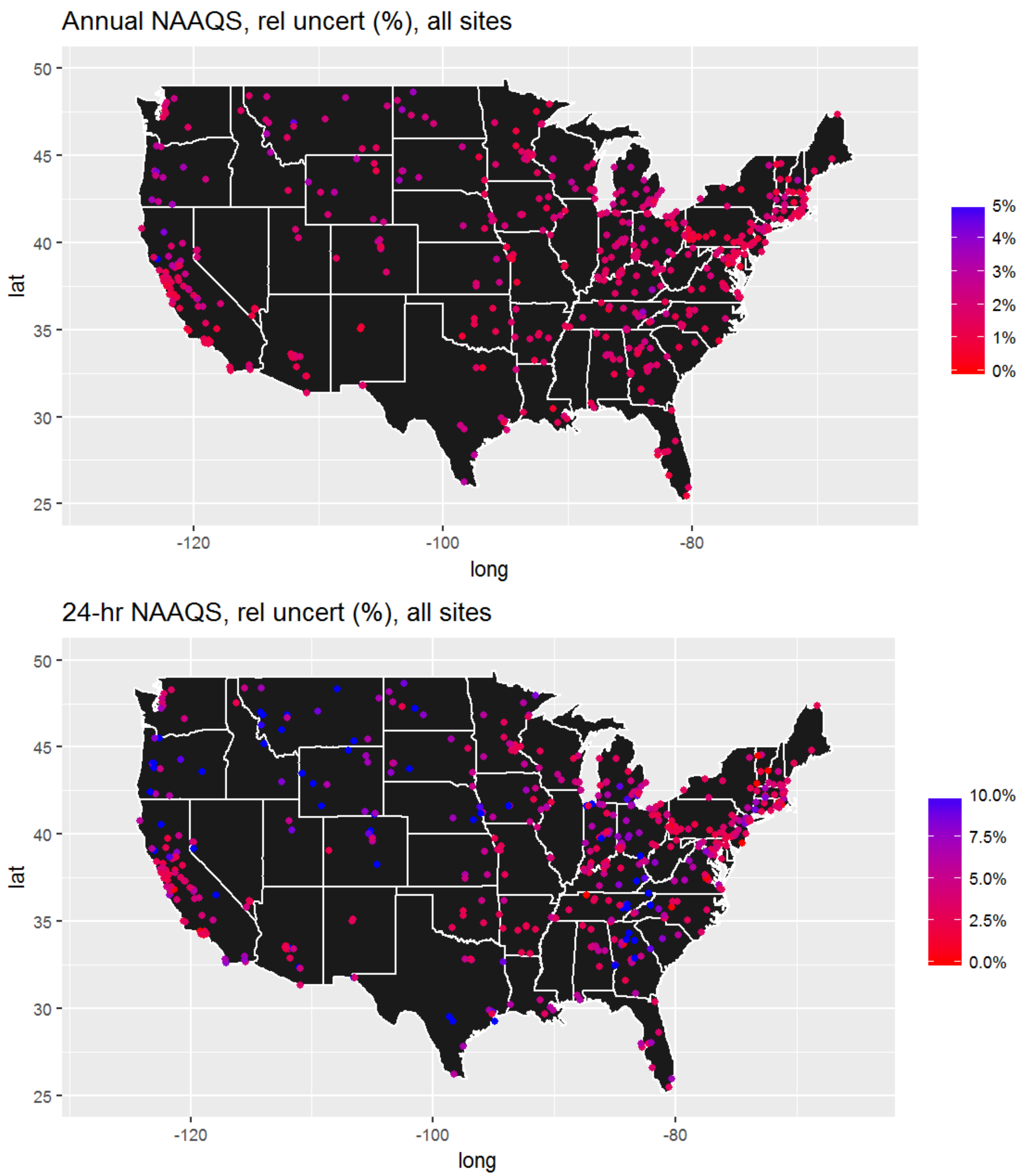


Figure 14 - Spatial distribution of the relative difference between the span of the 50% CI and the actual DV for the 2014-2016 $PM_{2.5}$ DVs.

4.3.1 $PM_{2.5}$ Temporal Trends

The median air quality variability from the 13 DV periods for both the annual and 24-hr $PM_{2.5}$ NAAQS 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 (0.03 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 the three most recent DV periods (*i.e.*, 2012-2014, 2013-2015, 2014-2016) is shown in Table 4, and when averaged result in a SIL value of 1.66% for the annual $PM_{2.5}$ NAAQS (12 $\mu g/m^3$) and 4.27% 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 1.5 $\mu g/m^3$ 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	2015 annual	2016 annual
Difference, median bootstrap vs actual	0.04%	0.03%	0.04%
Avg. 25% CI span	0.67%	0.70%	0.71%
Avg. 50% CI span	1.63%	1.65%	1.69%
Avg. 68% CI span	2.44%	2.46%	2.45%
Avg. 75% CI span	2.80%	2.83%	2.82%
Avg. 95% CI span	4.72%	4.86%	4.79%
Year/NAAQS	**2014 24-hr**	**2015 24-hr**	**2016 24-hr**
Difference, median bootstrap vs actual	1.14%	1.36%	1.23%
Avg. 25% CI span	2.27%	2.27%	2.50%
Avg. 50% CI span	4.29%	4.17%	4.35%
Avg. 68% CI span	6.00%	6.25%	6.52%
Avg. 75% CI span	6.82%	7.50%	7.69%
Avg. 95% CI span	12.50%	12.50%	13.16%
Number of sites	507	531	535

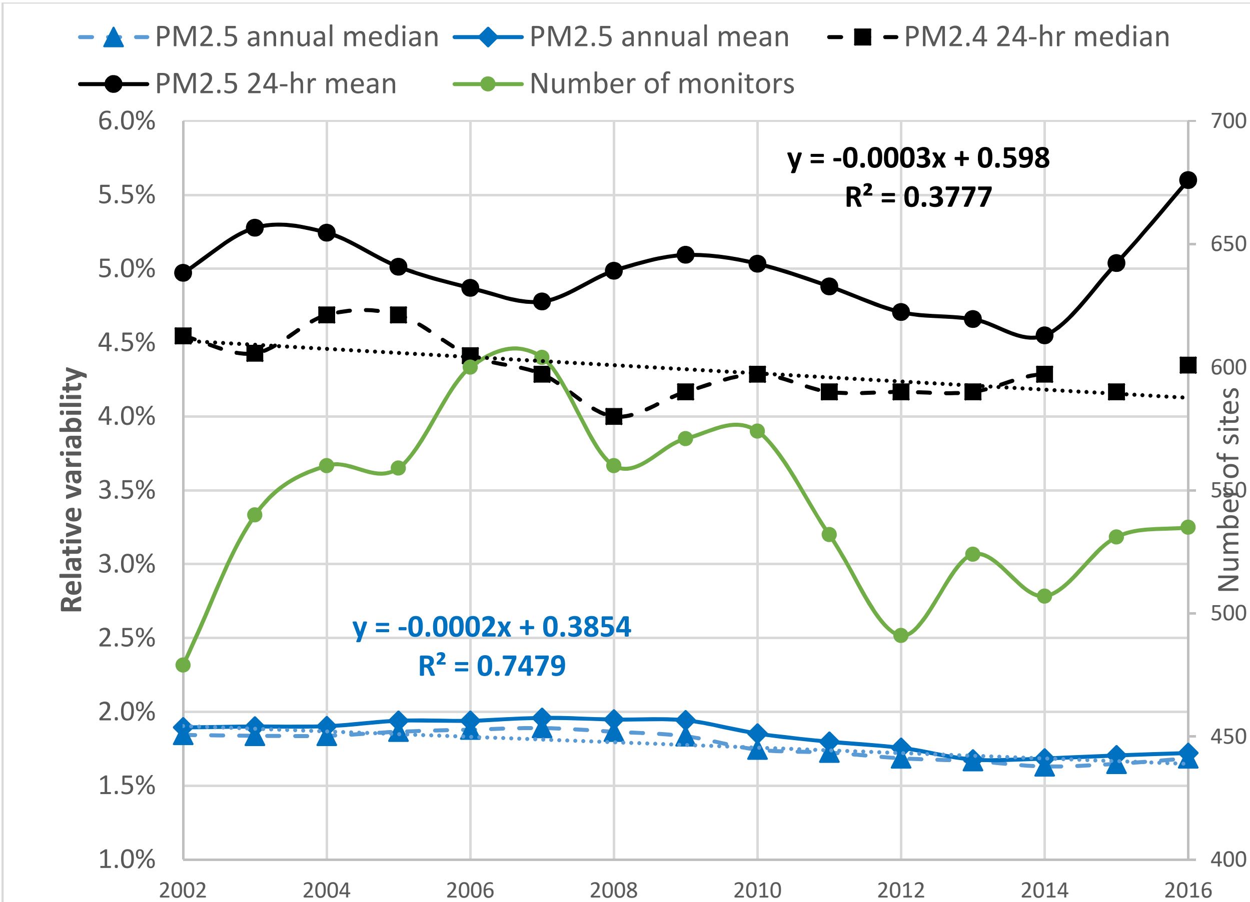


Figure 15 - Median and mean variability in the network determined from the bootstrap analysis (50% CI) for the 15 DV periods from 2002-2016 for $PM_{2.5}$ (each DV period represents 3 years of data and the data is plotted on the ending year: *i.e.*, the 2016 DV period is from 2014-2016 and plotted at 2016).

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/R-18-001 April, 2018

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

EPA-454/R-18-001
April 2018

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

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Assessment Division
Research Triangle Park, NC

Contents

List of Figures

1 Bootstrap examples

Bootstrap examples from selects $PM_{2.5}$ sites for the 2008-2010 DV period. Top left, top right, and middle left plots show the distribution of daily PM concentrations for 2008, 2009, and 2010, respectively. The vertical red line shows the annual mean and the vertical blue line shows the annual 98th percentile. Middle left plots show sample distributions of resampled data from 2008, along with the annual mean and the 98th percentile from each resample. The bottom left plots show the distribution of the annual DVs from the 20,000 resampled DV periods (2008-2010). The bottom right plots show the distribution of the 24-hr DVs from the 20,000 resampled DV periods (2008-2010)

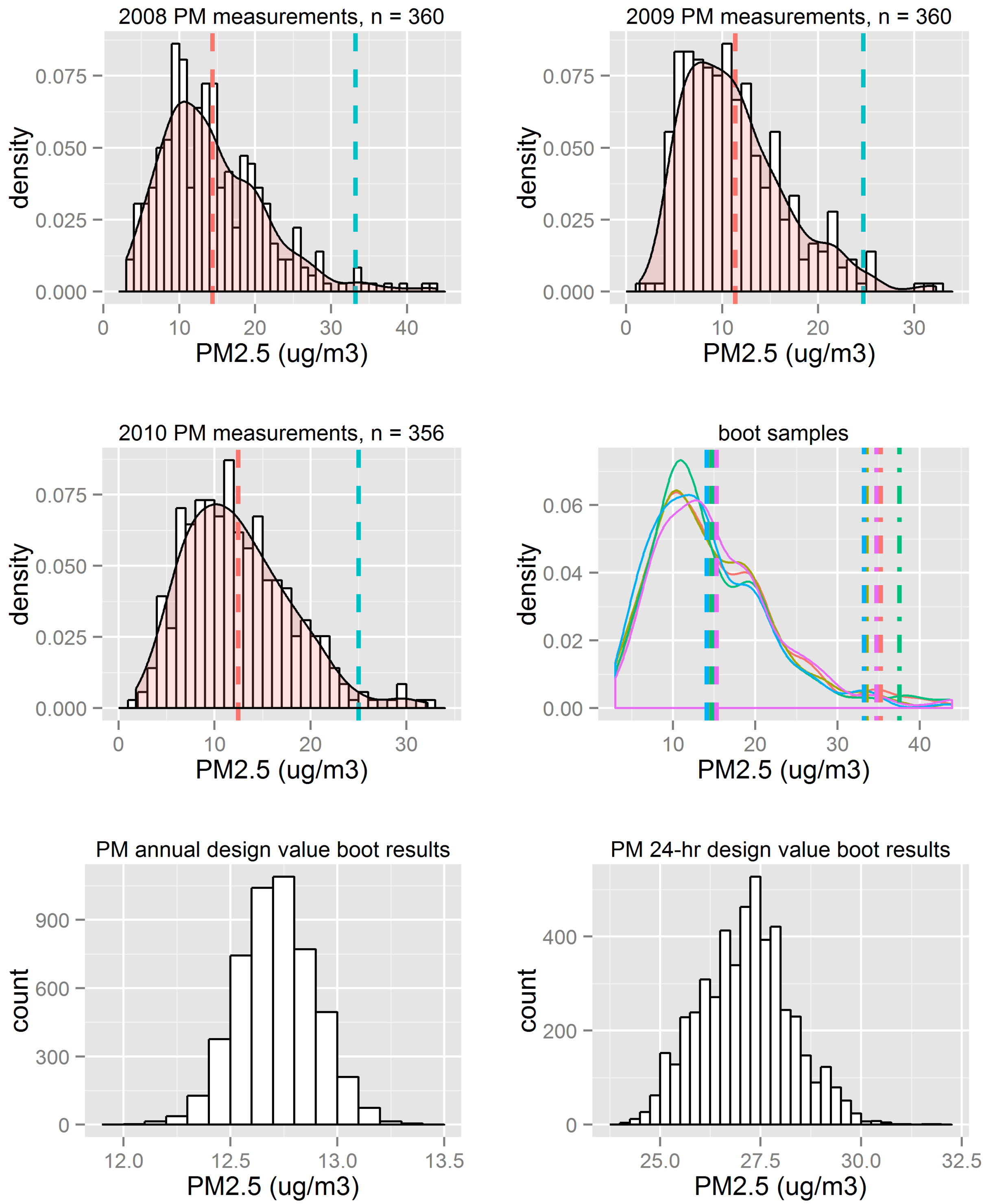


Figure 1: Example from site 10732003.

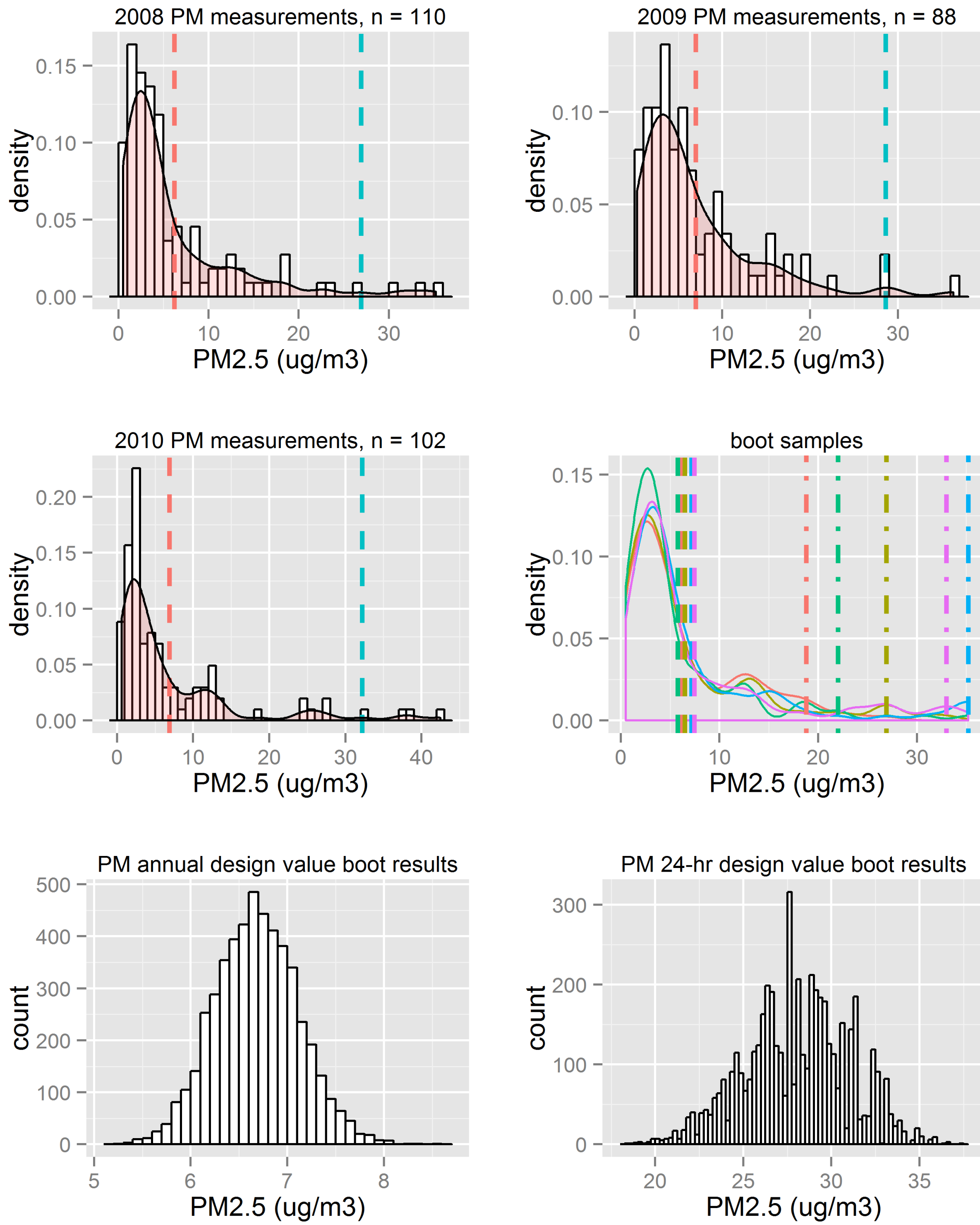


Figure 2: Example from site 21700008.

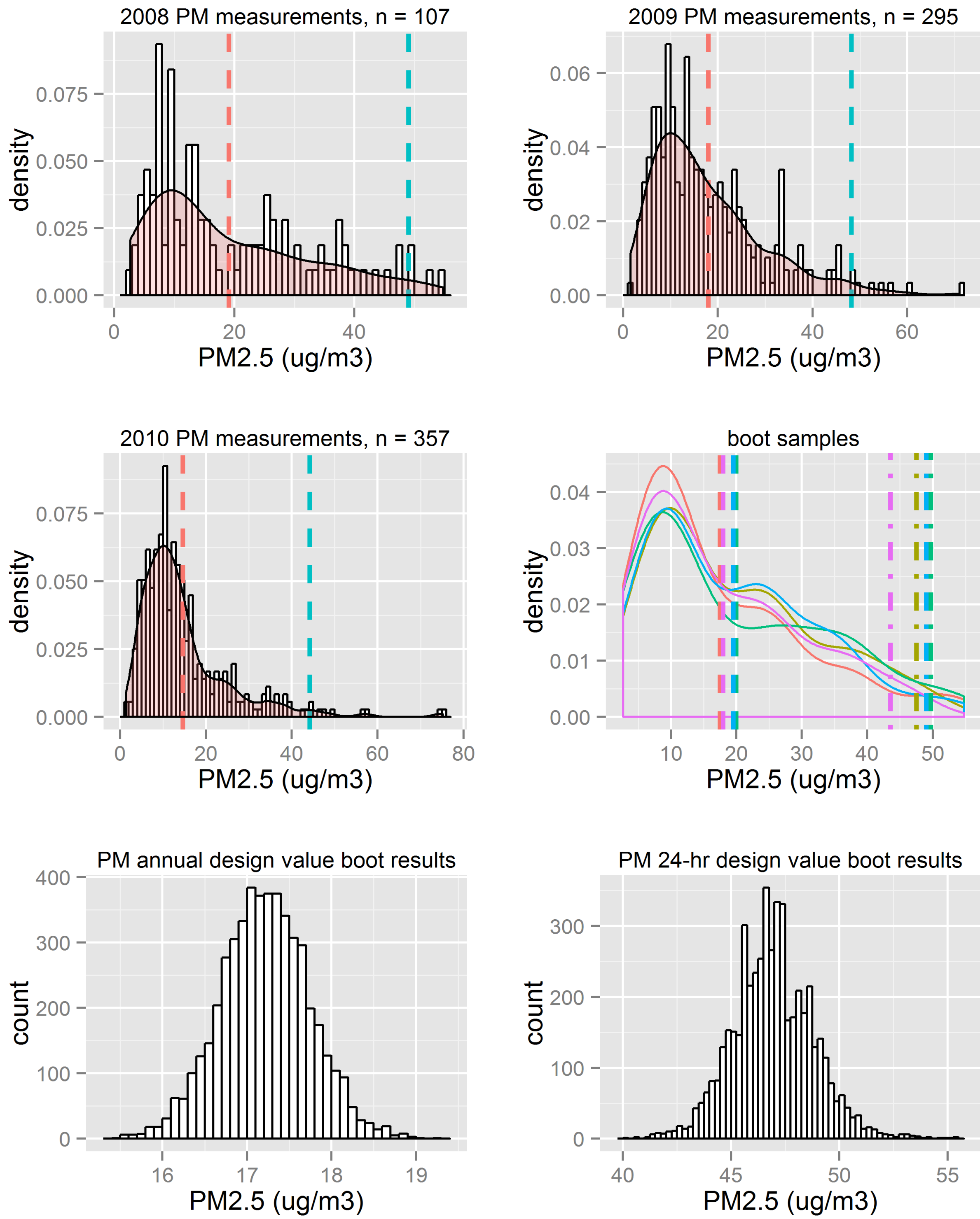


Figure 3: Example from site 60195001.

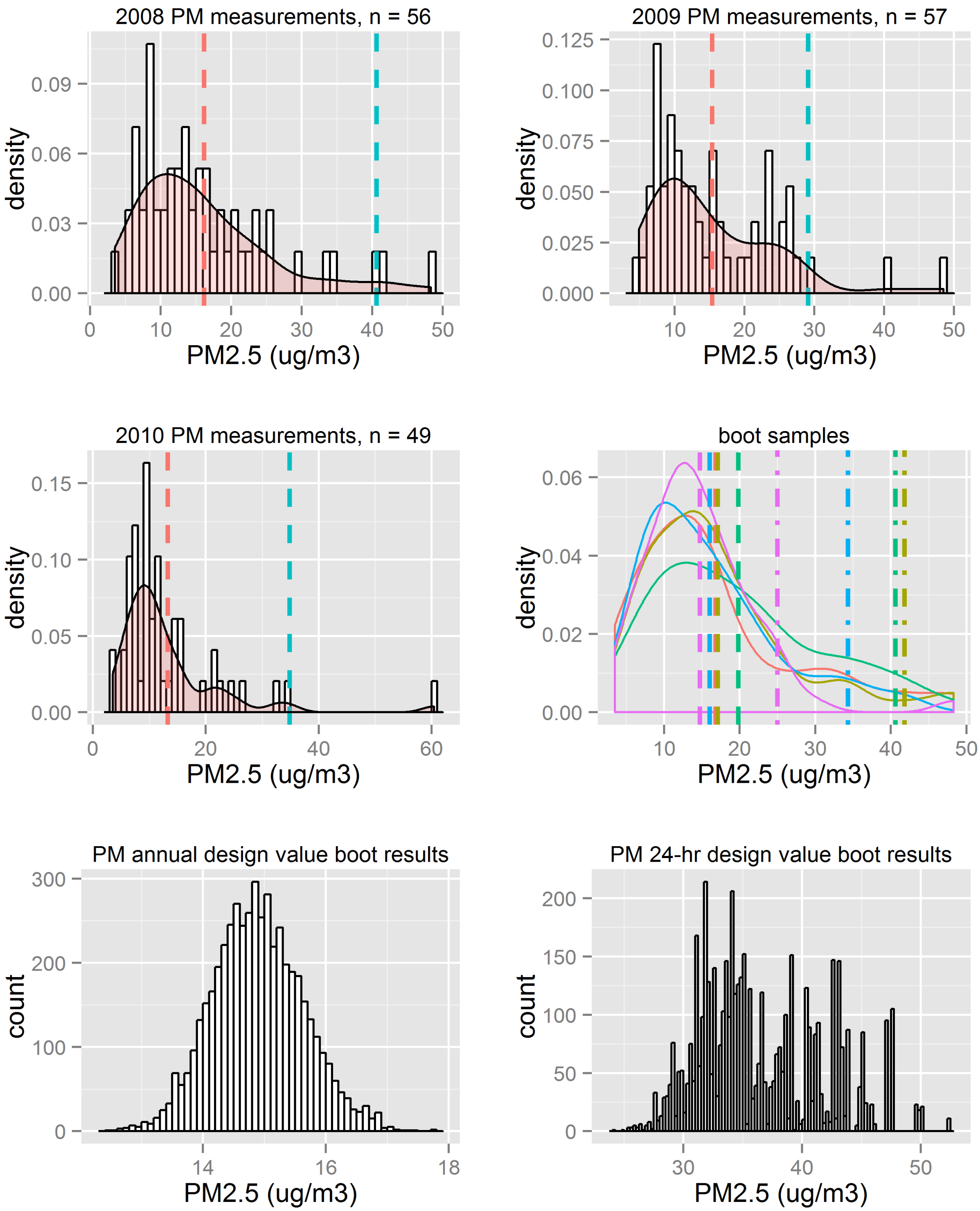


Figure 4: Example from site 481410053.

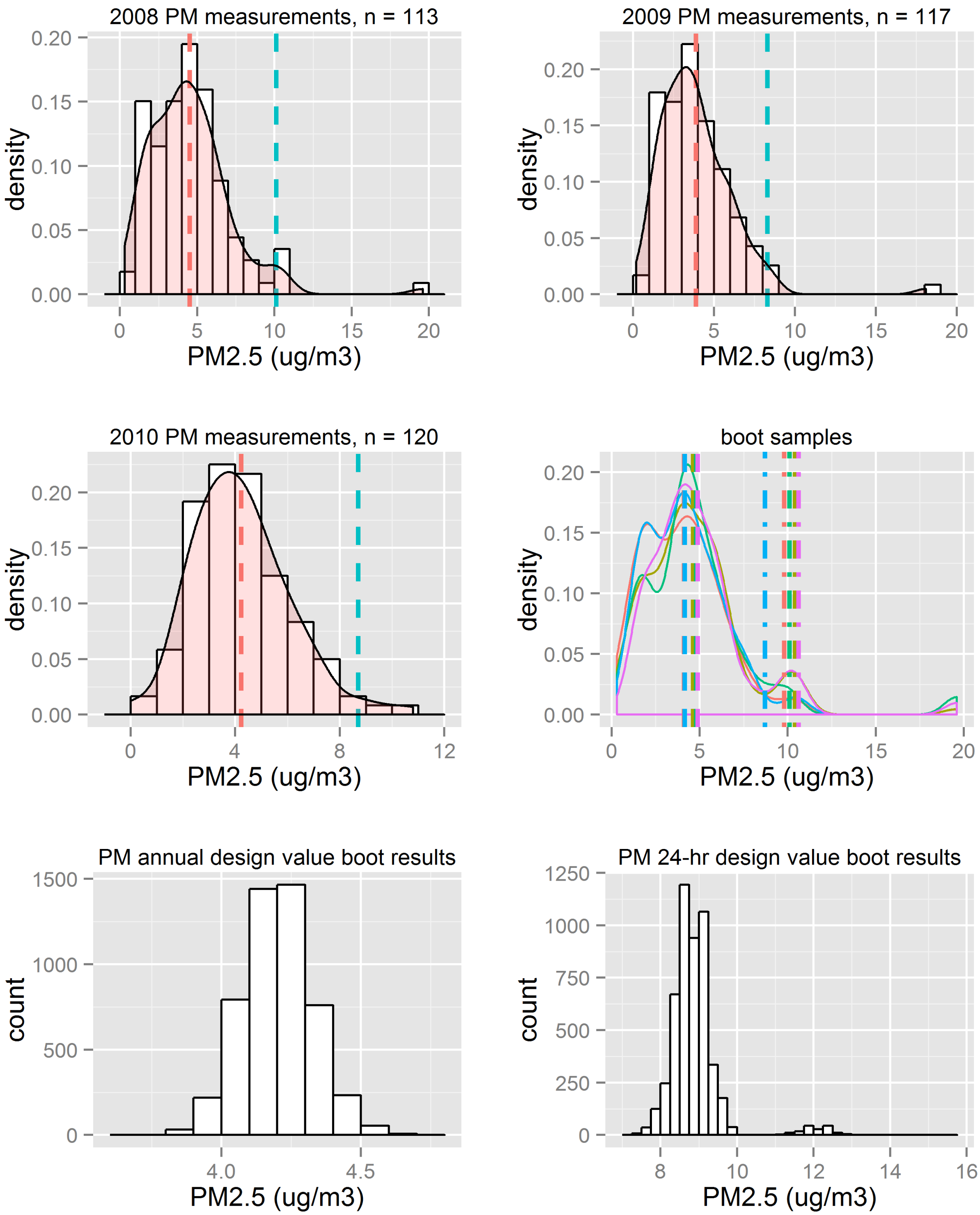


Figure 5: Example from site 560210001.

2 Ozone results

Bootstrap results for ozone data from the years 2000-2013. Each section containts a single DV period,*e.g.*, the results for 2015 include data from 2013-2015.

2.1 2013-2015 ozone bootstrap results

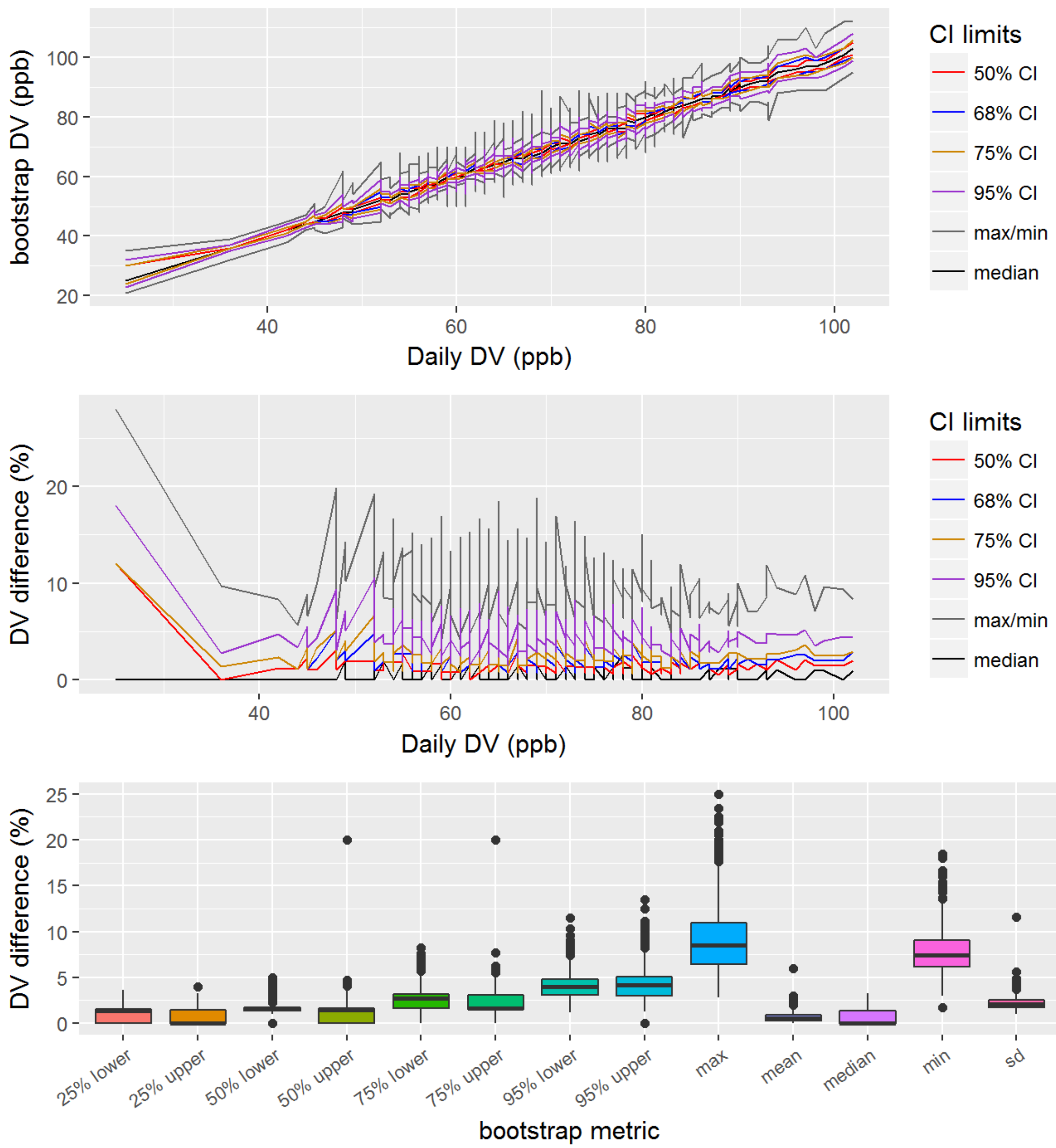


Figure 6: Bootstrap results for the ozone 2015 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

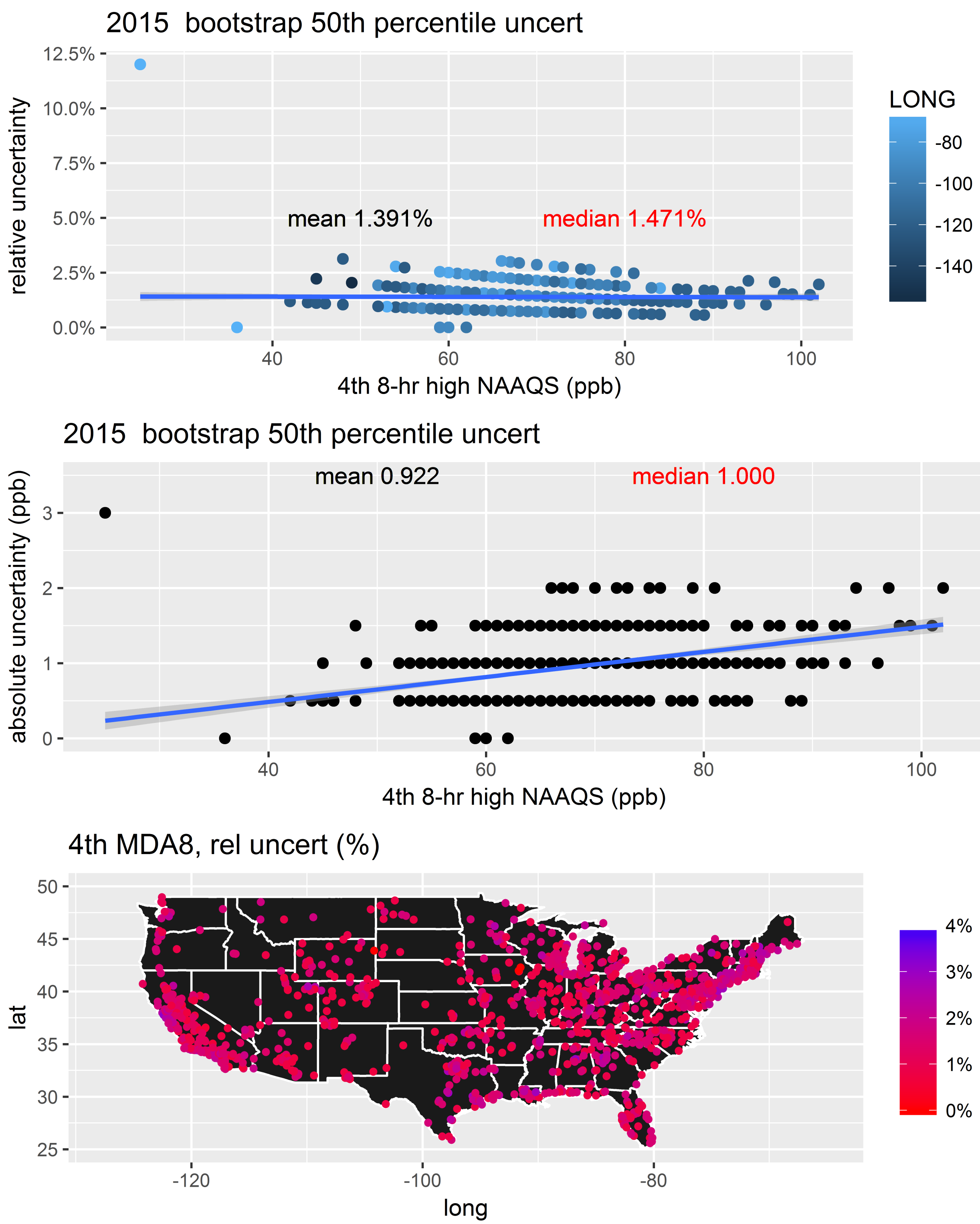


Figure 7: Bootstrap results from the 50% CIs for the 2015 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2015 ozone DV at each site.

2.2 2012-2014 ozone bootstrap results

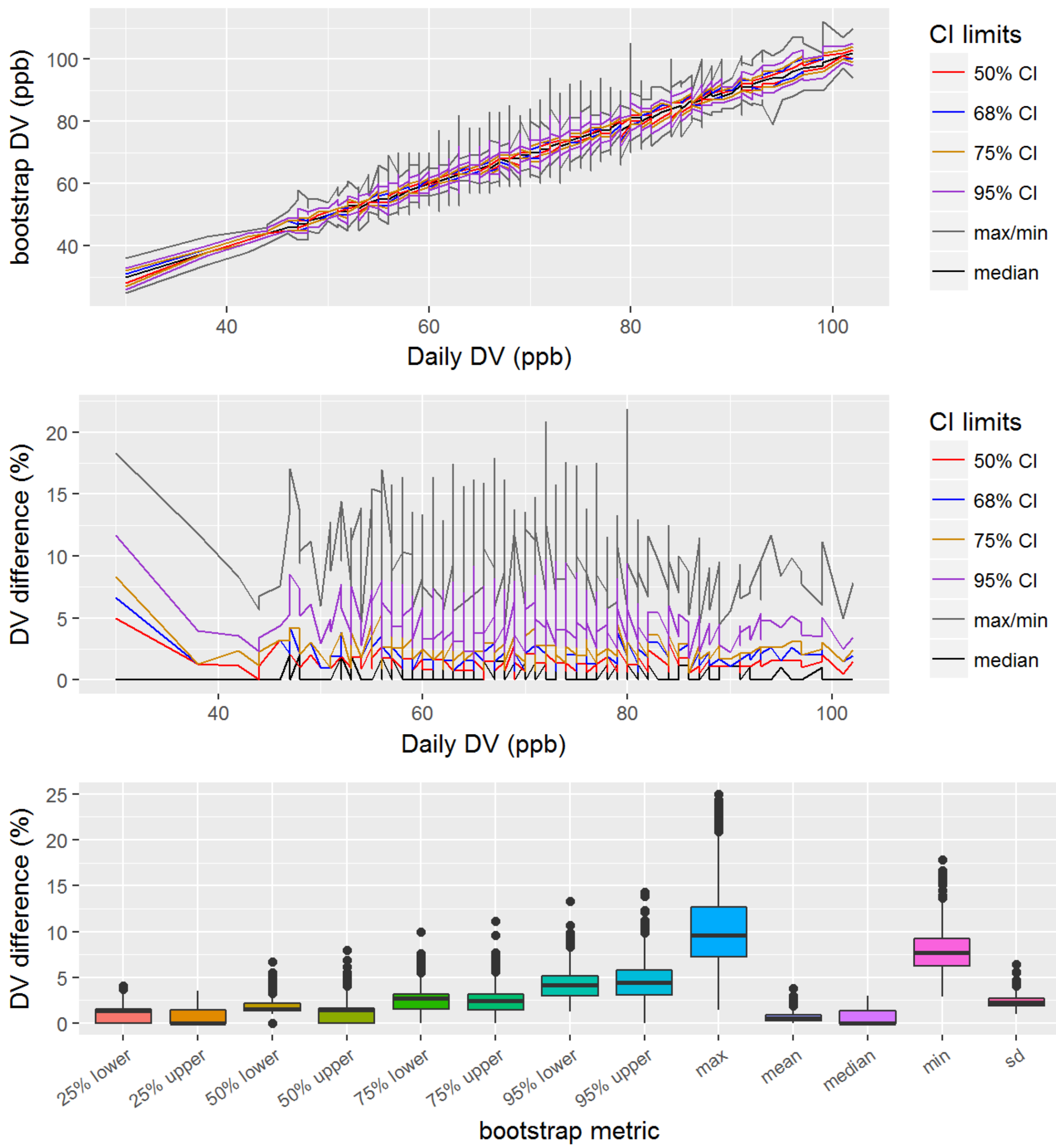


Figure 8: Bootstrap results for the ozone 2014 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

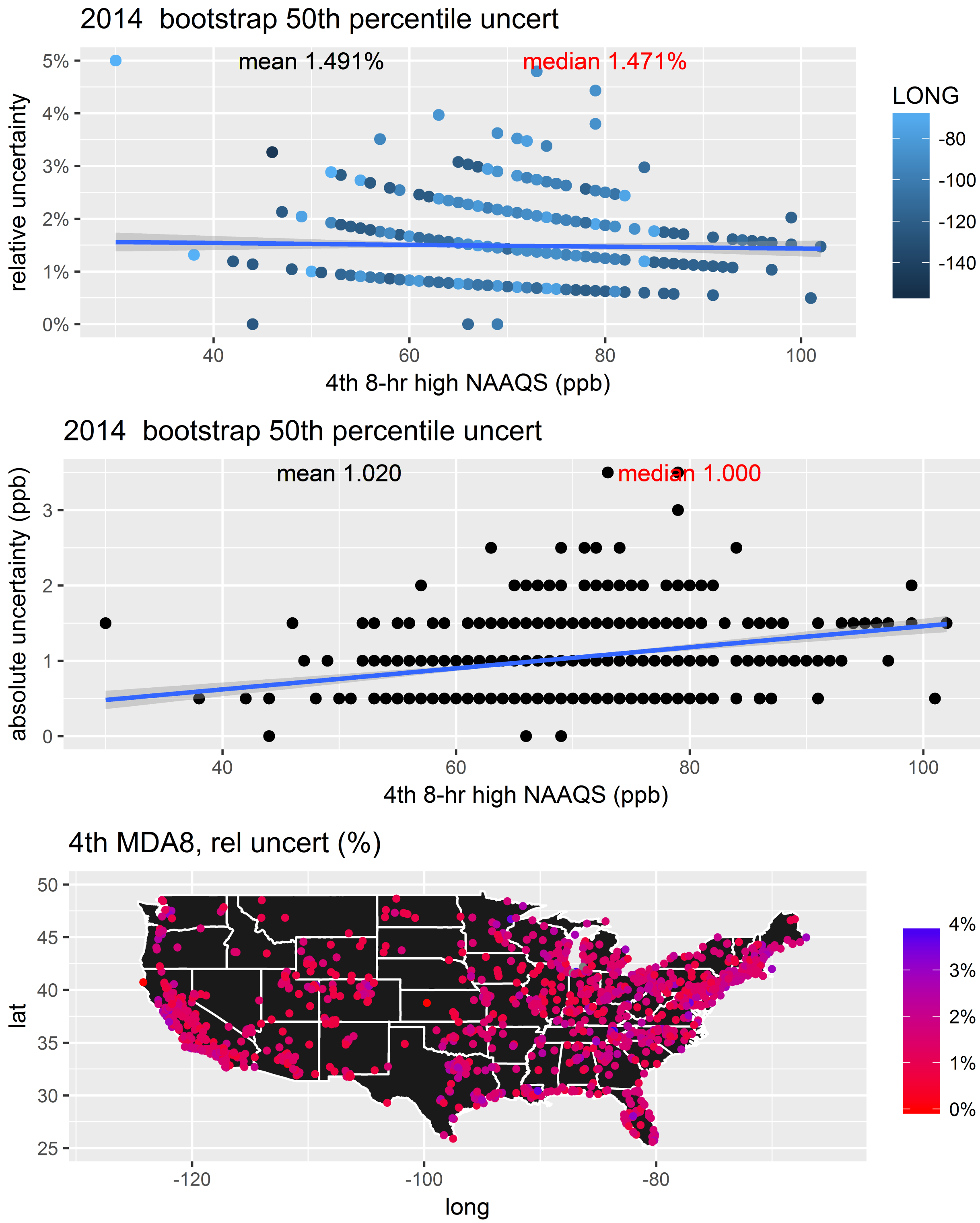


Figure 9: Bootstrap results from the 50% CIs for the 2014 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2014 ozone DV at each site.

2.3 2011-2013 ozone bootstrap results

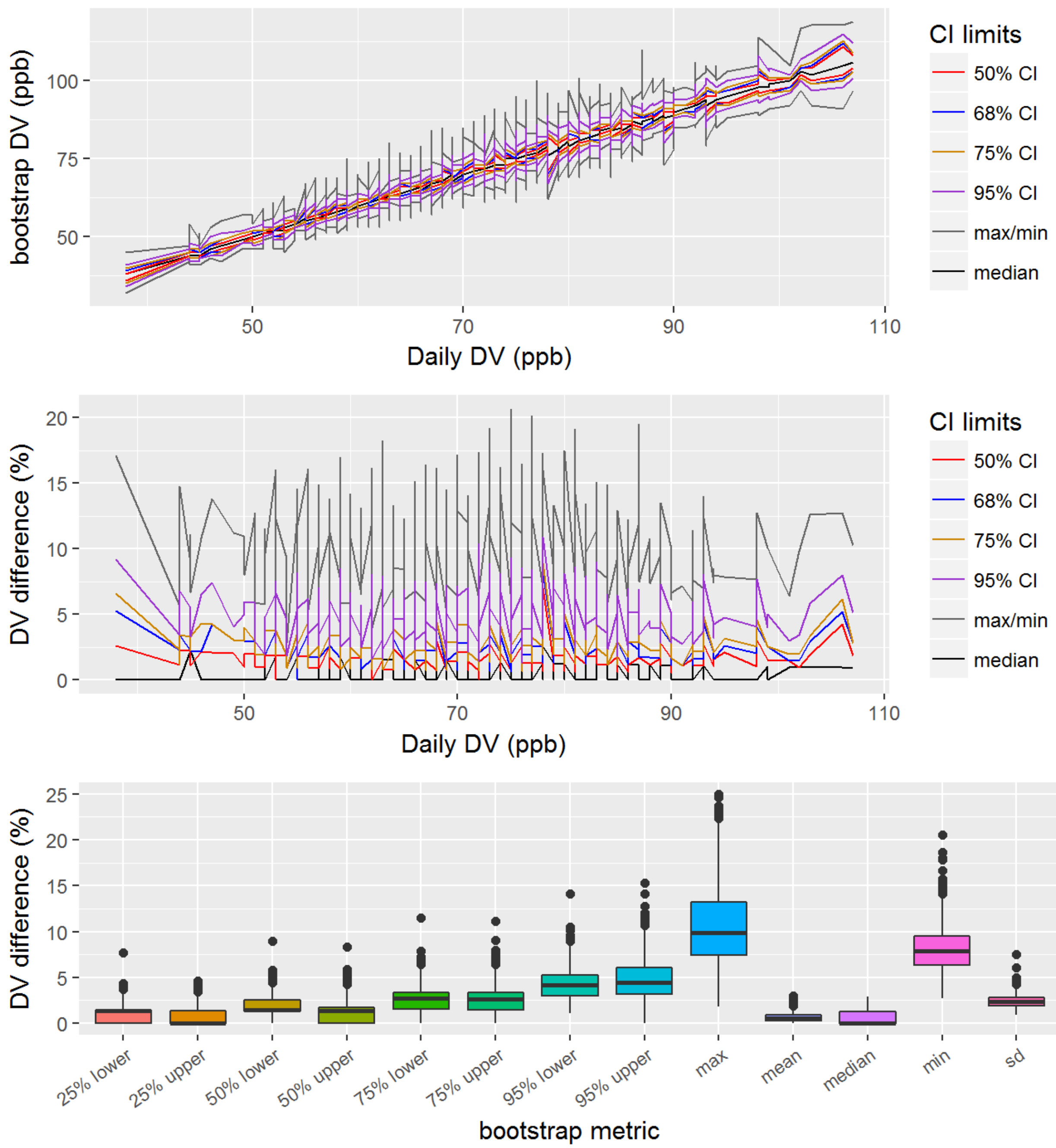


Figure 10: Bootstrap results for the ozone 2013 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

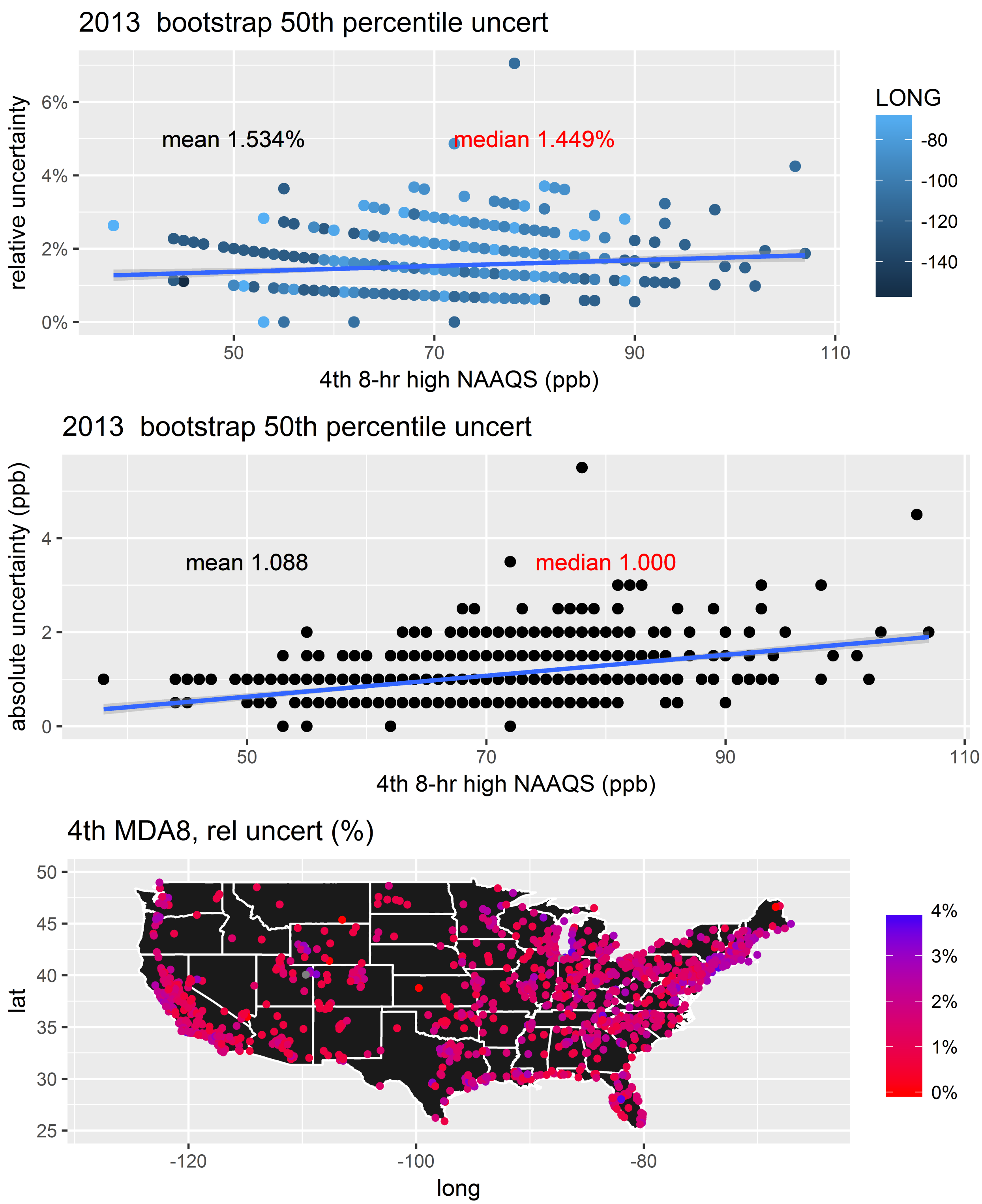


Figure 11: Bootstrap results from the 50% CIs for the 2013 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2013 ozone DV at each site.

2.4 2010-2012 ozone bootstrap results

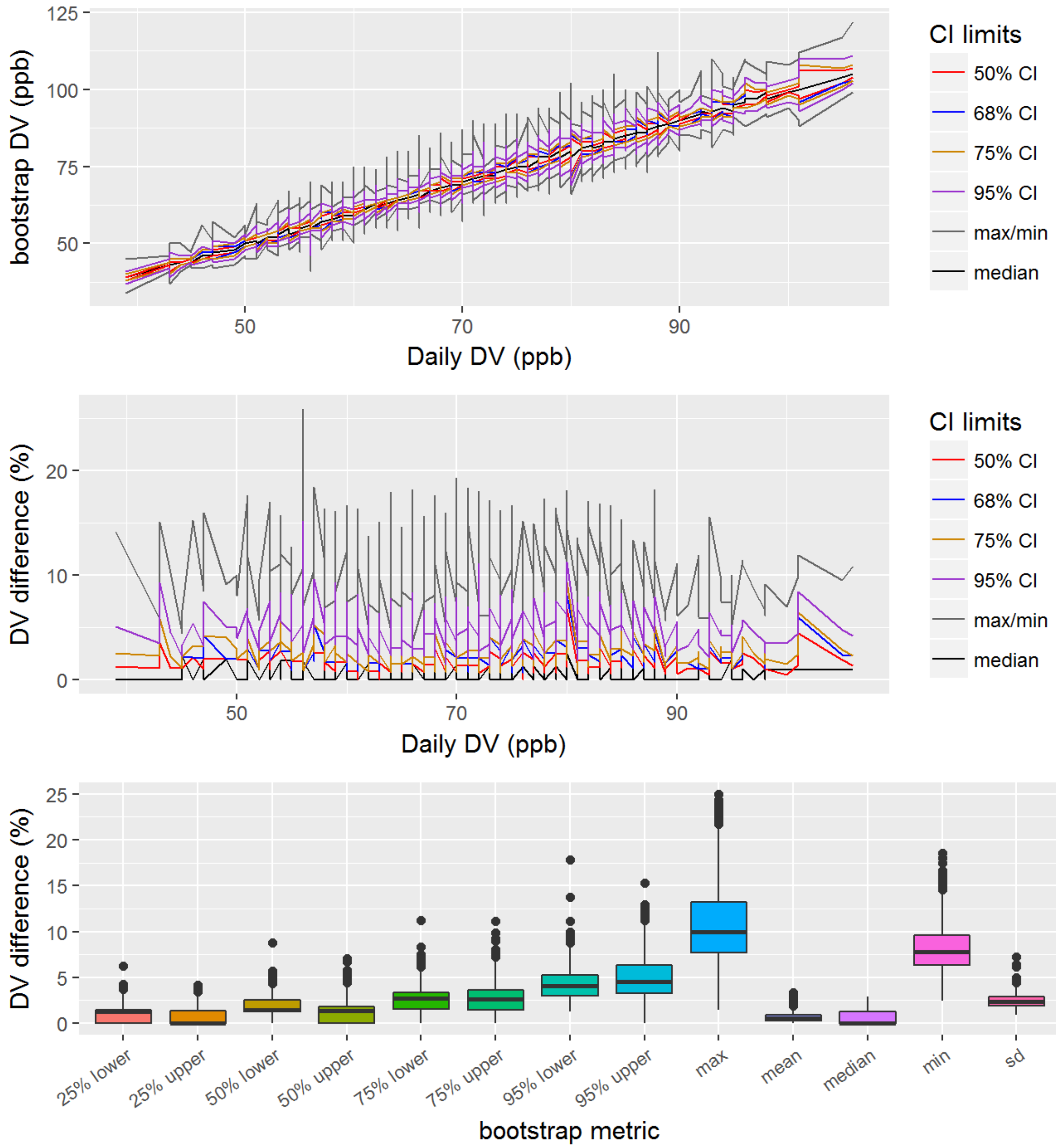


Figure 12: Bootstrap results for the ozone 2012 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

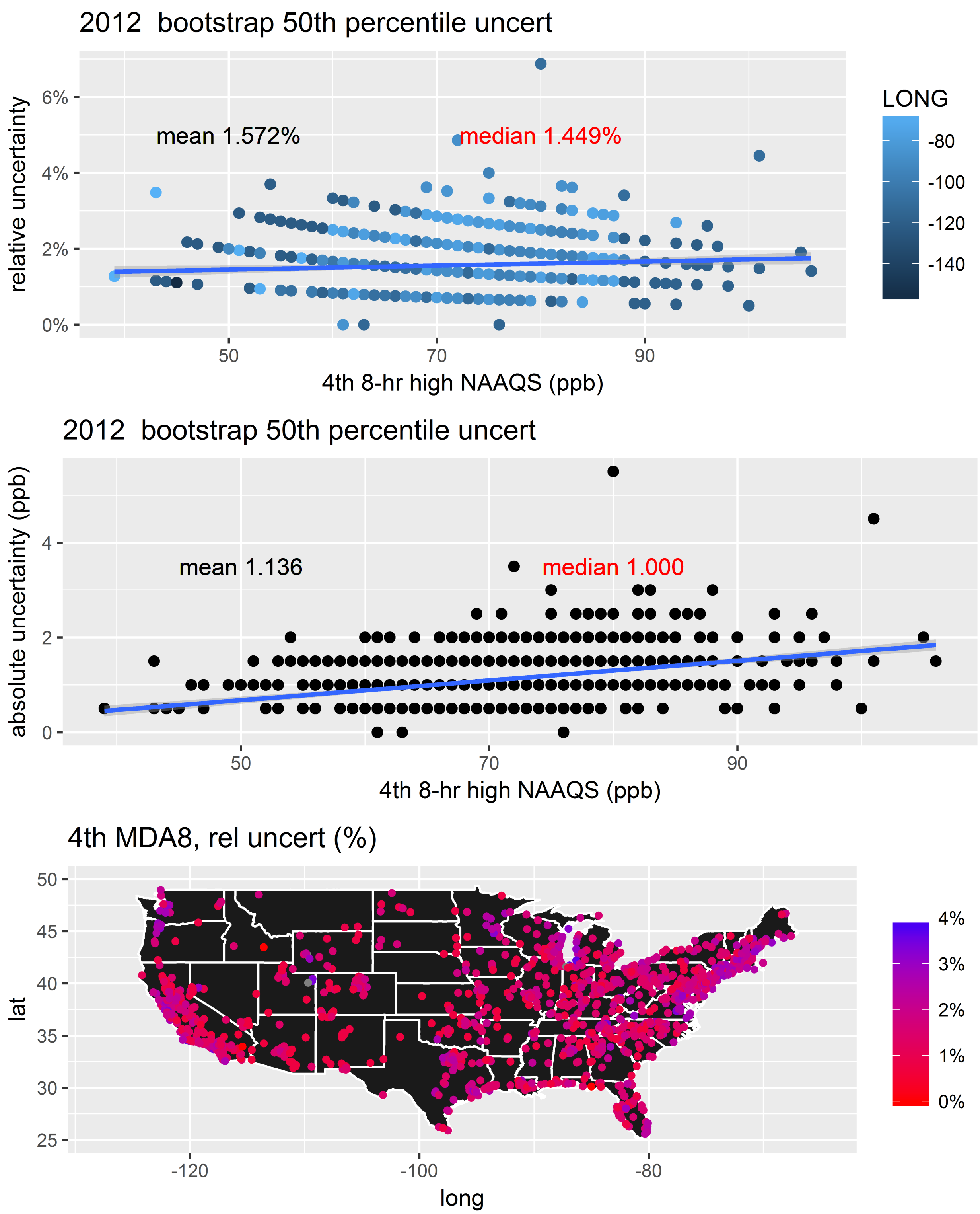


Figure 13: Bootstrap results from the 50% CIs for the 2012 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2012 ozone DV at each site.

2.5 2009-2011 ozone bootstrap results

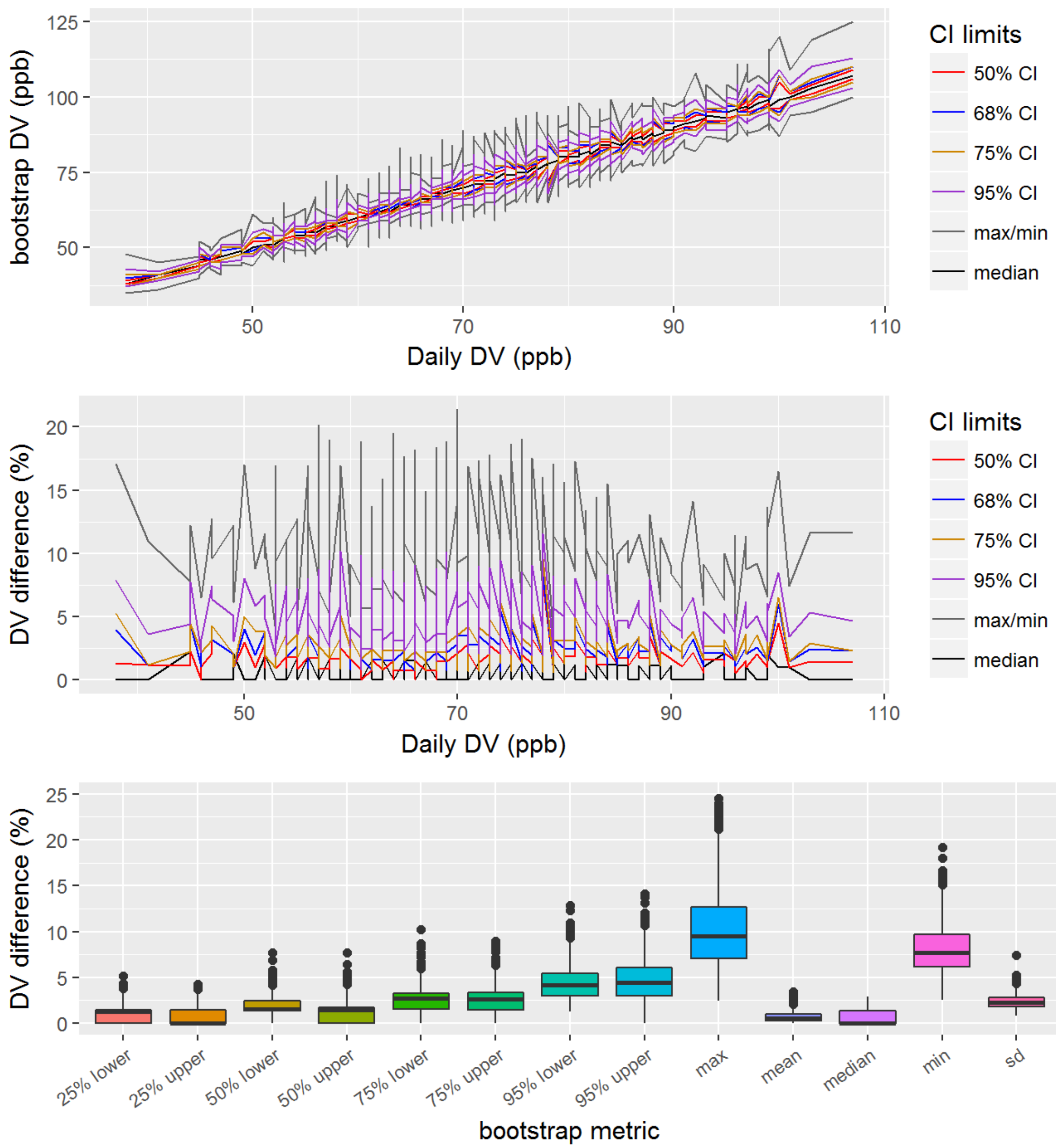


Figure 14: Bootstrap results for the ozone 2011 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

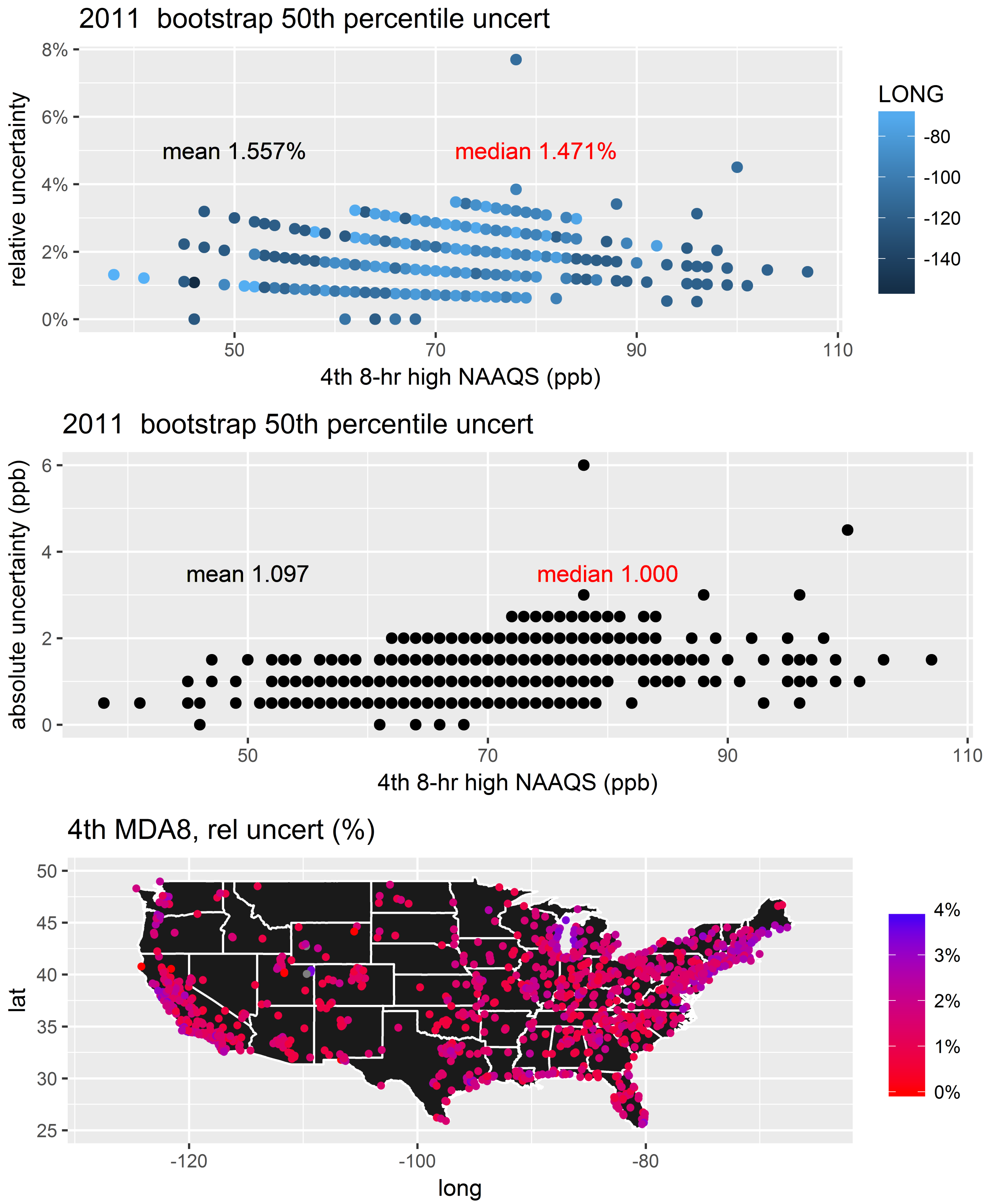


Figure 15: Bootstrap results from the 50% CIs for the 2011 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2011 ozone DV at each site.

2.6 2008-2010 ozone bootstrap results

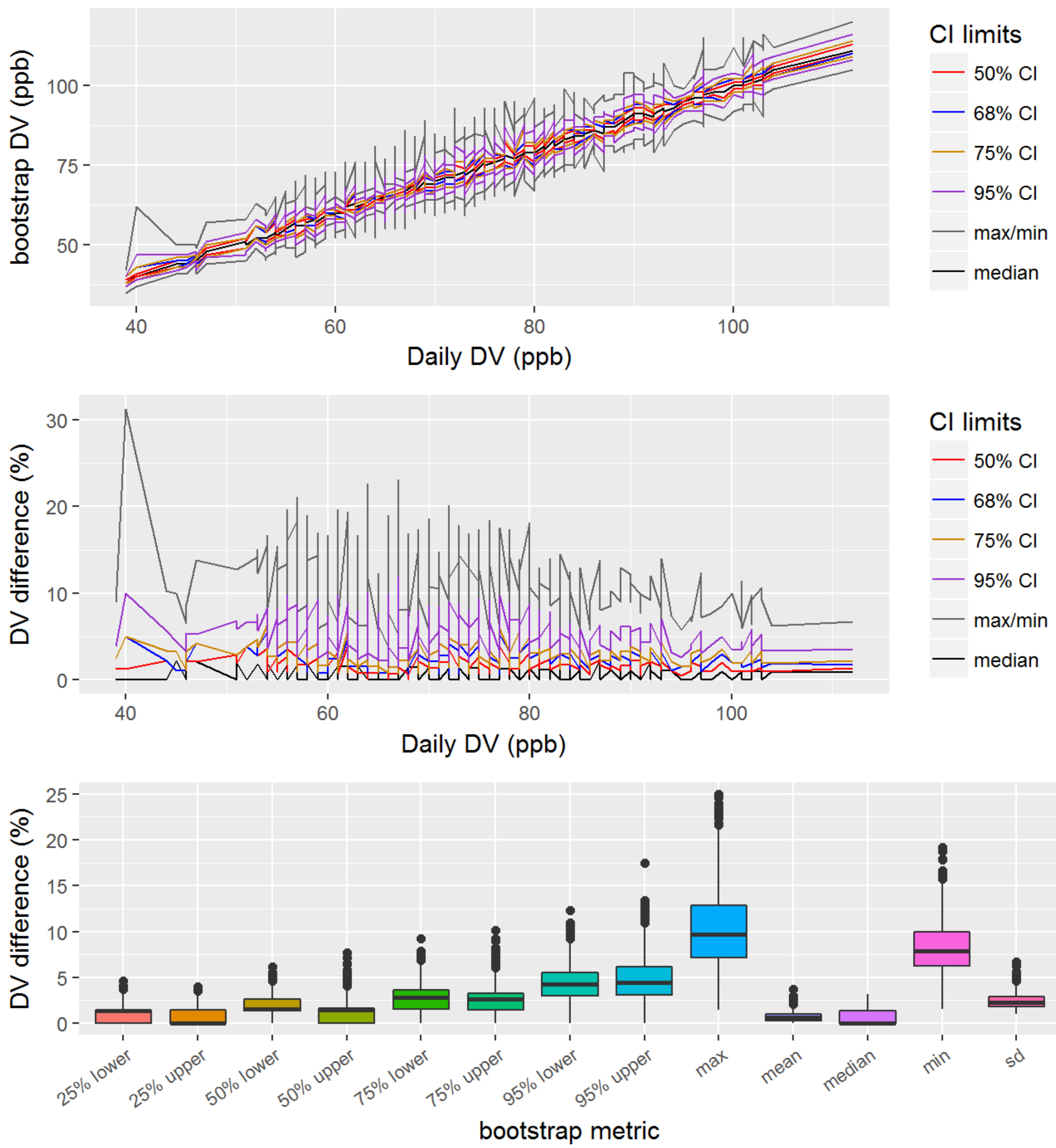


Figure 16: Bootstrap results for the ozone 2010 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

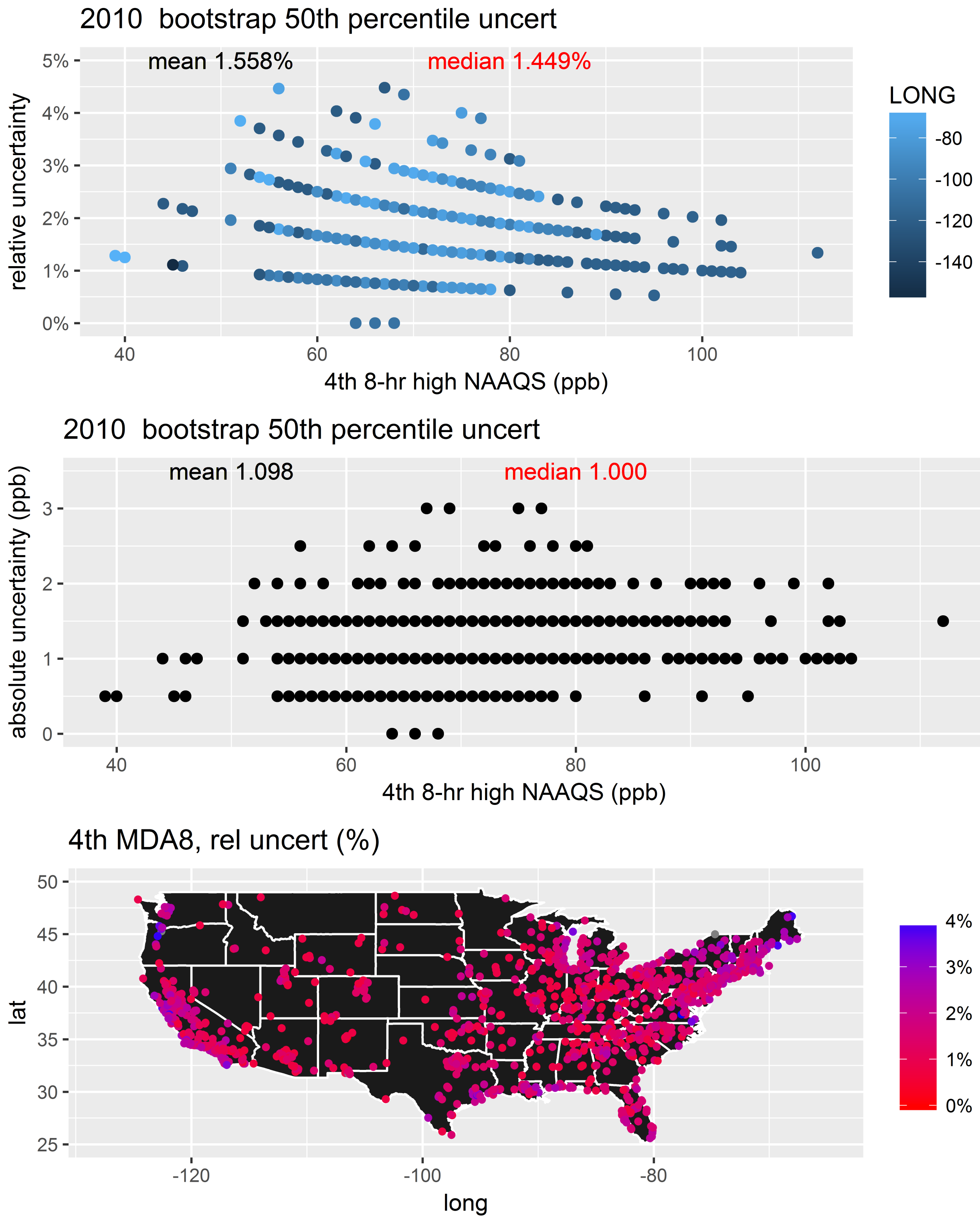


Figure 17: Bootstrap results from the 50% CIs for the 2010 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2010 ozone DV at each site.

2.7 2007-2009 ozone bootstrap results

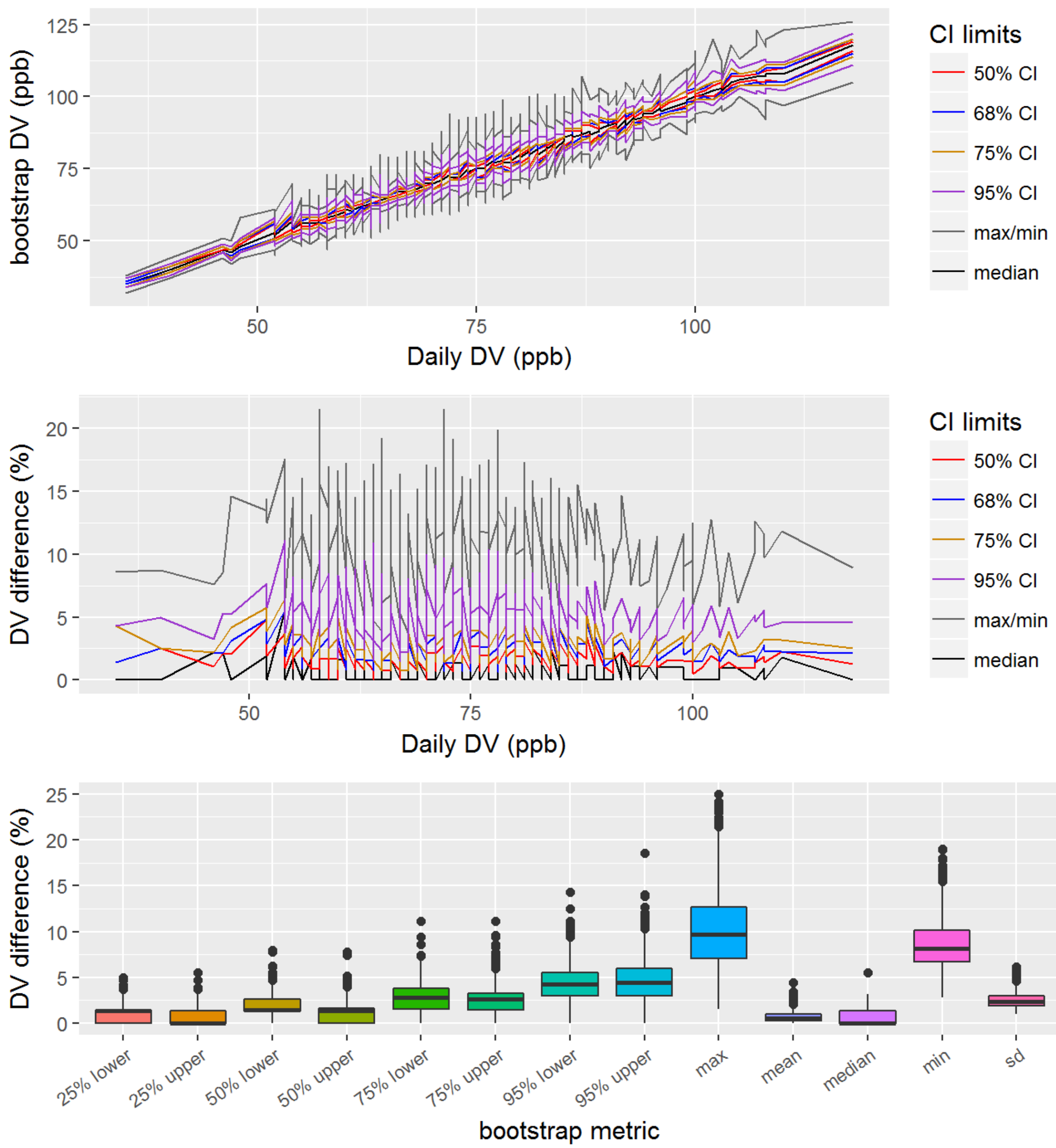


Figure 18: Bootstrap results for the ozone 2009 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

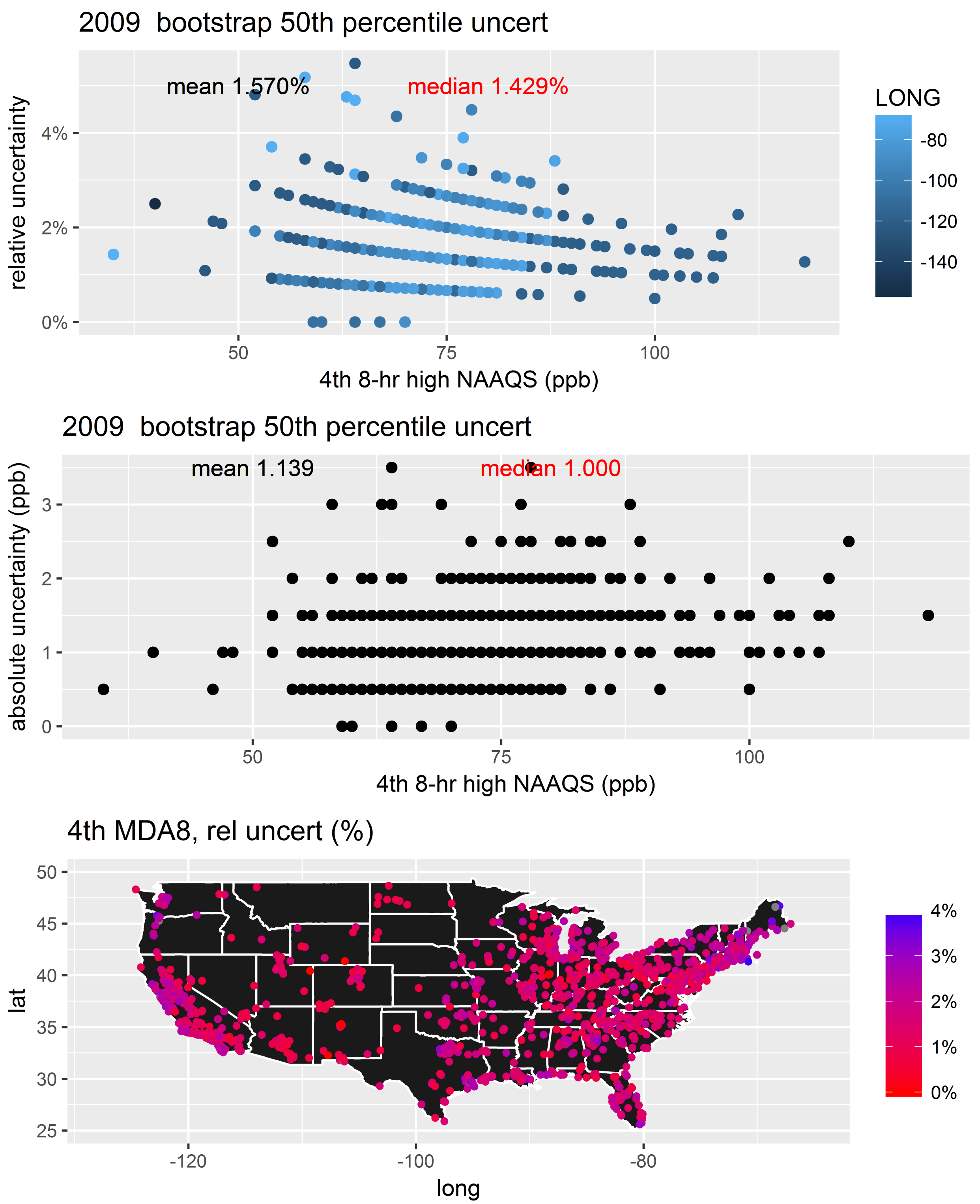


Figure 19: Bootstrap results from the 50% CIs for the 2009 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2009 ozone DV at each site.

2.8 2006-2008 ozone bootstrap results

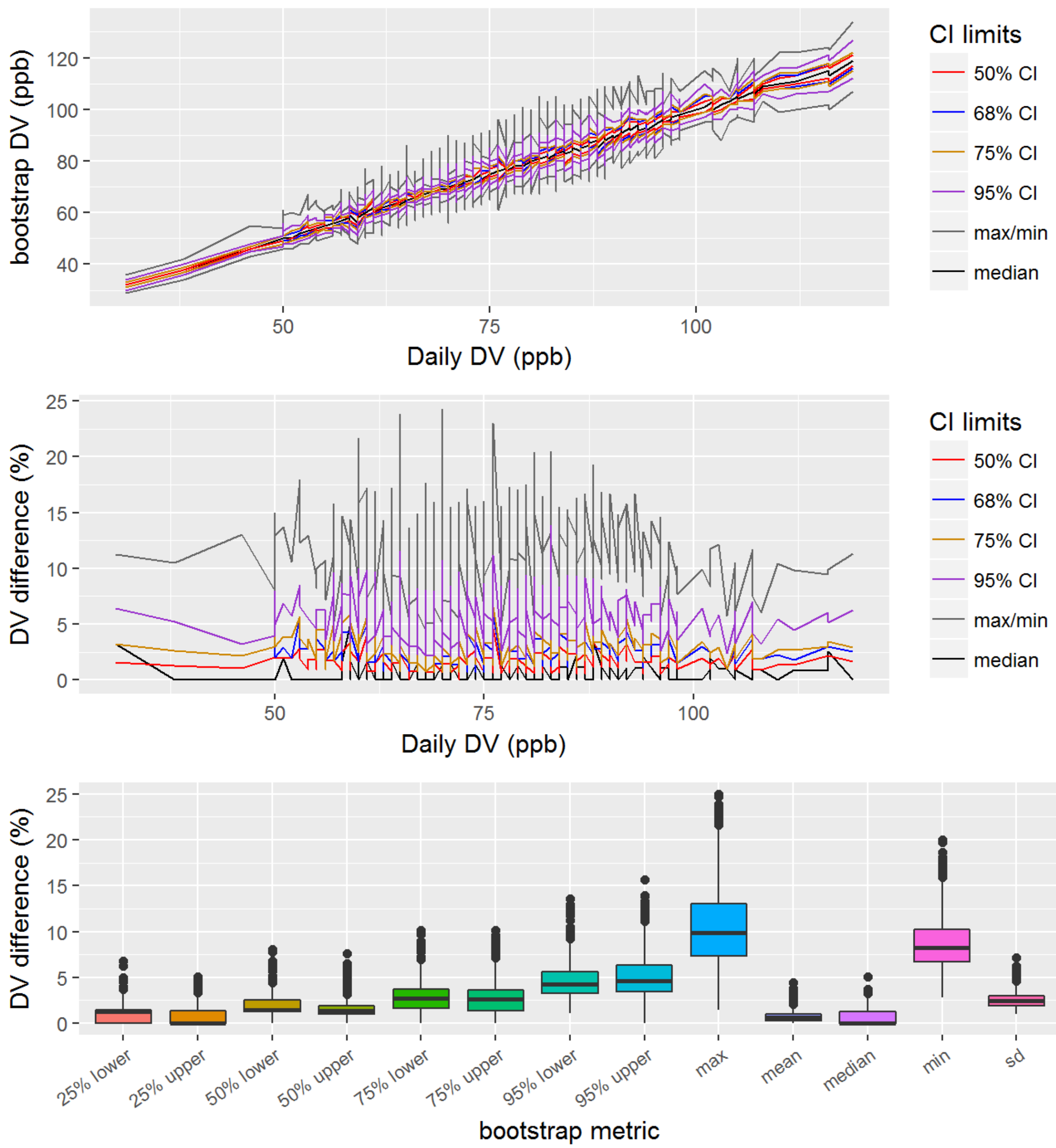


Figure 20: Bootstrap results for the ozone 2008 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

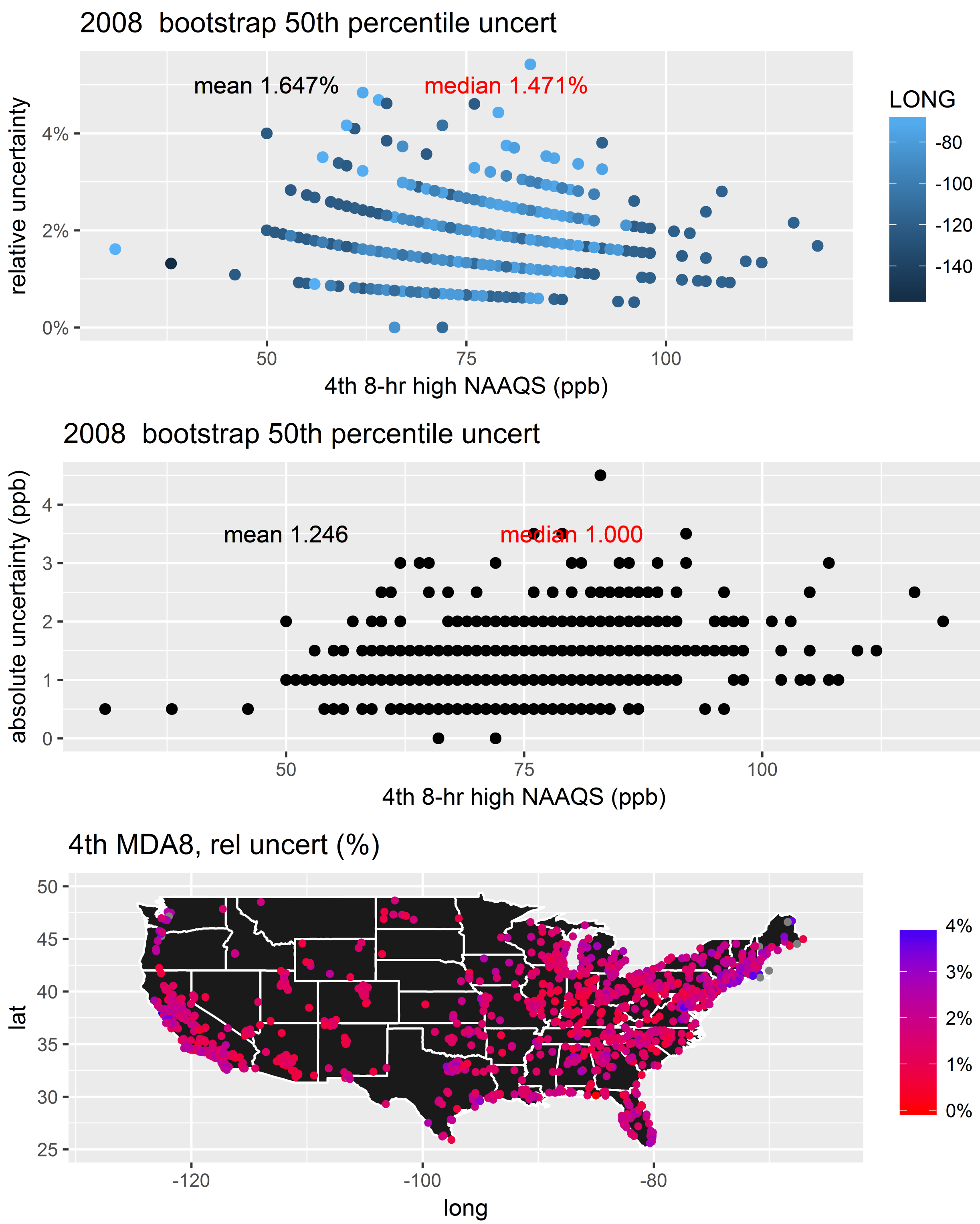


Figure 21: Bootstrap results from the 50% CIs for the 2008 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2008 ozone DV at each site.

2.9 2005-2007 ozone bootstrap results

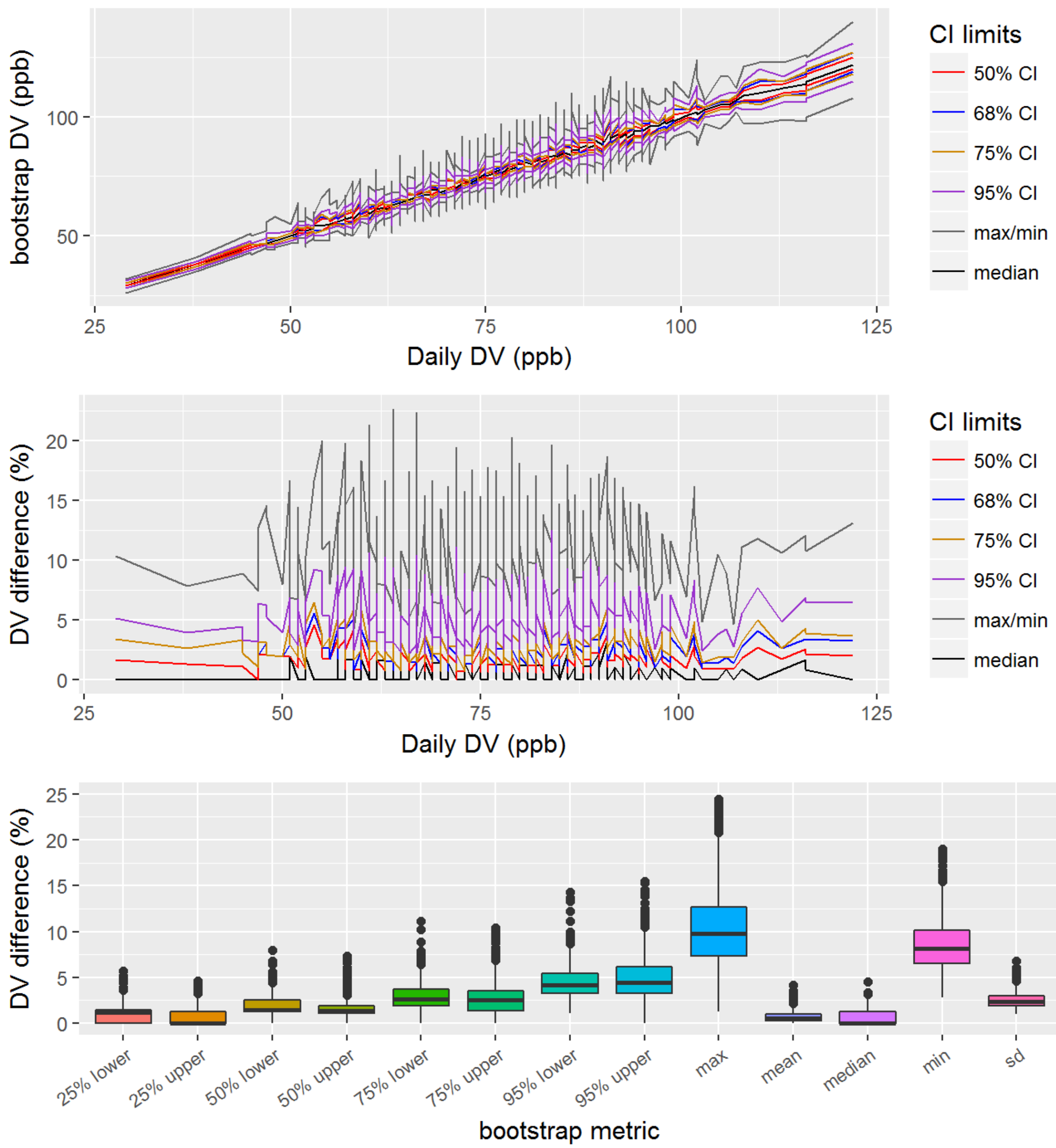


Figure 22: Bootstrap results for the ozone 2007 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

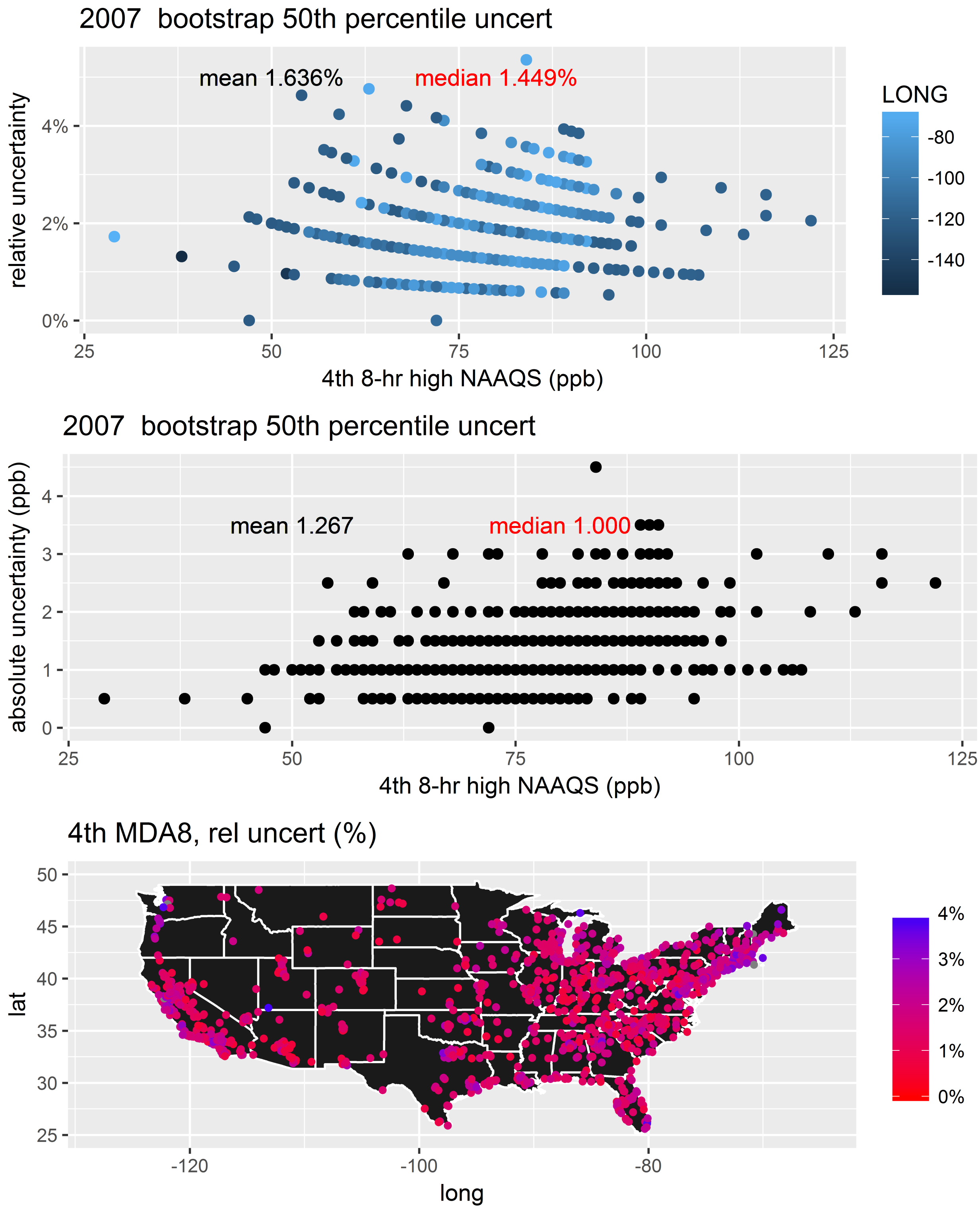


Figure 23: Bootstrap results from the 50% CIs for the 2007 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2007 ozone DV at each site.

2.10 2004-2006 ozone bootstrap results

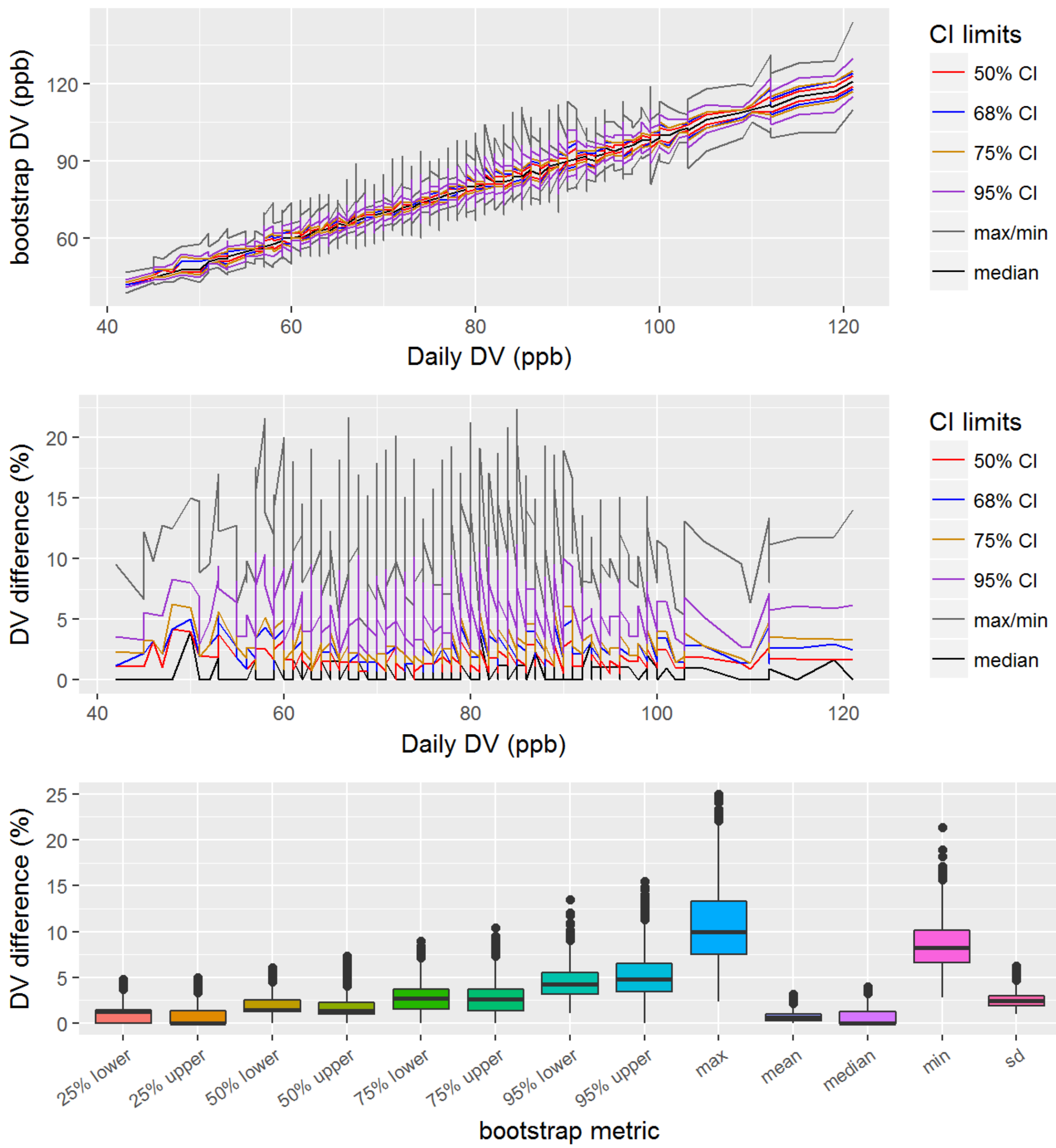


Figure 24: Bootstrap results for the ozone 2006 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

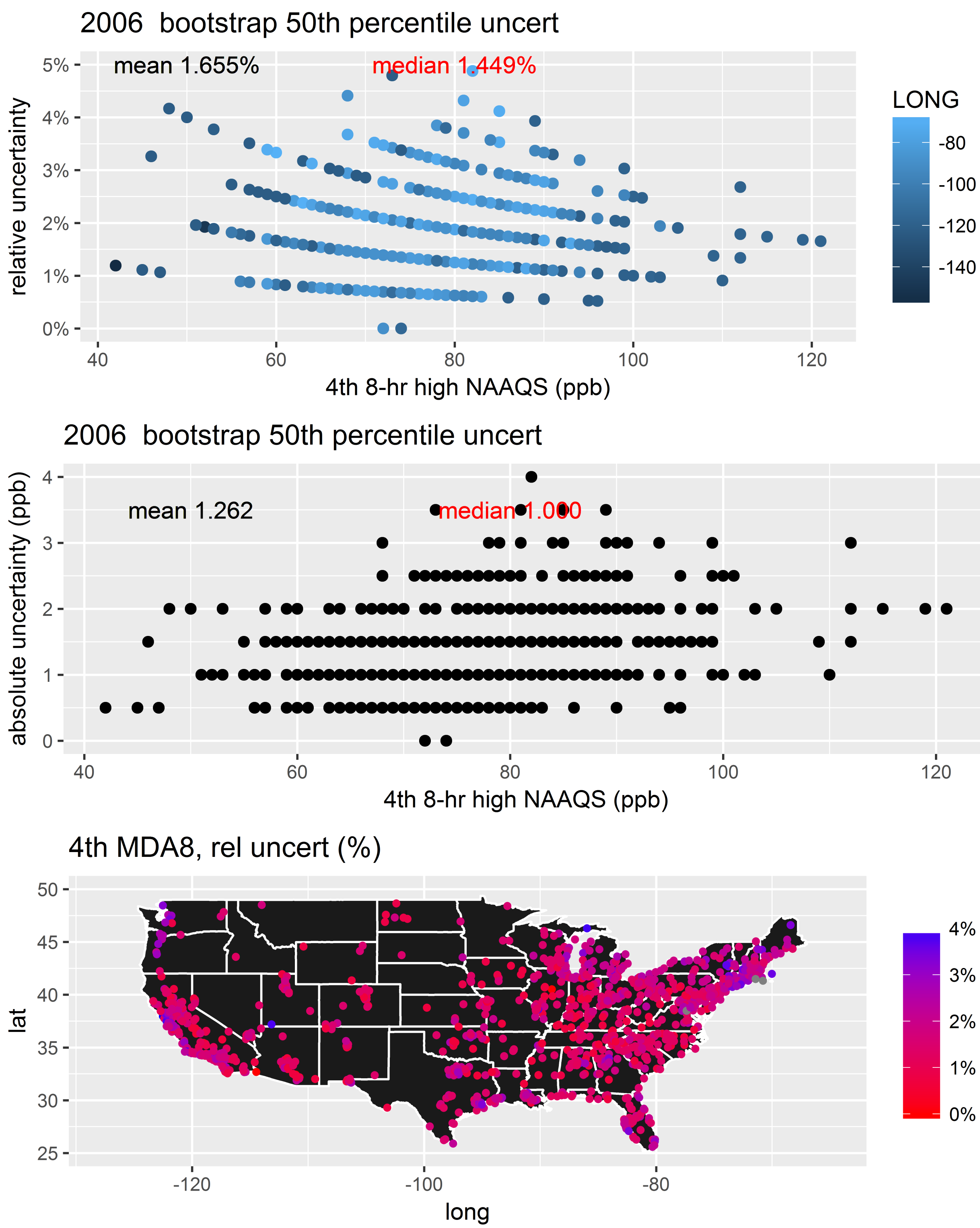


Figure 25: Bootstrap results from the 50% CIs for the 2006 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2006 ozone DV at each site.

2.11 2003-2005 ozone bootstrap results

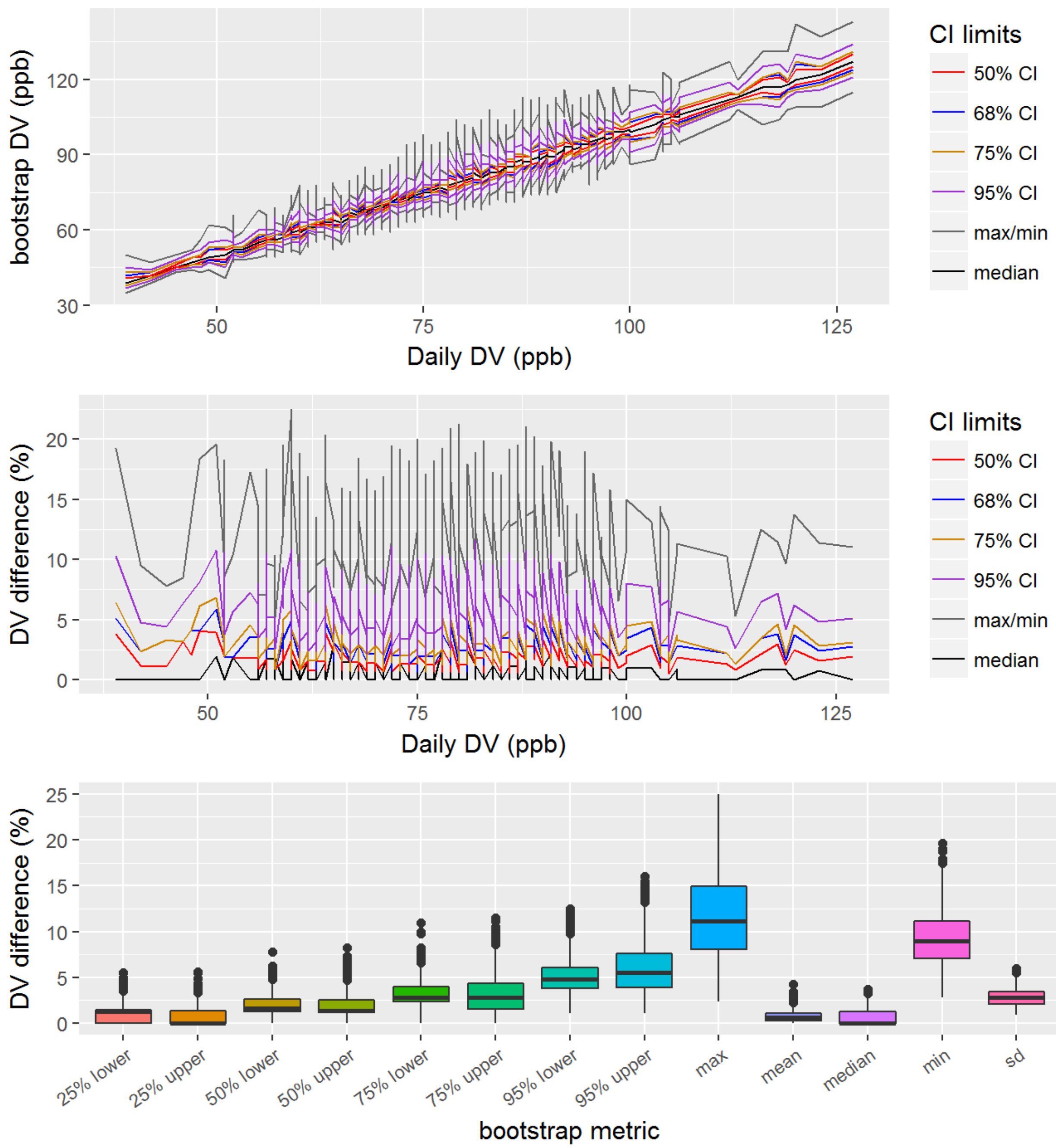


Figure 26: Bootstrap results for the ozone 2005 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

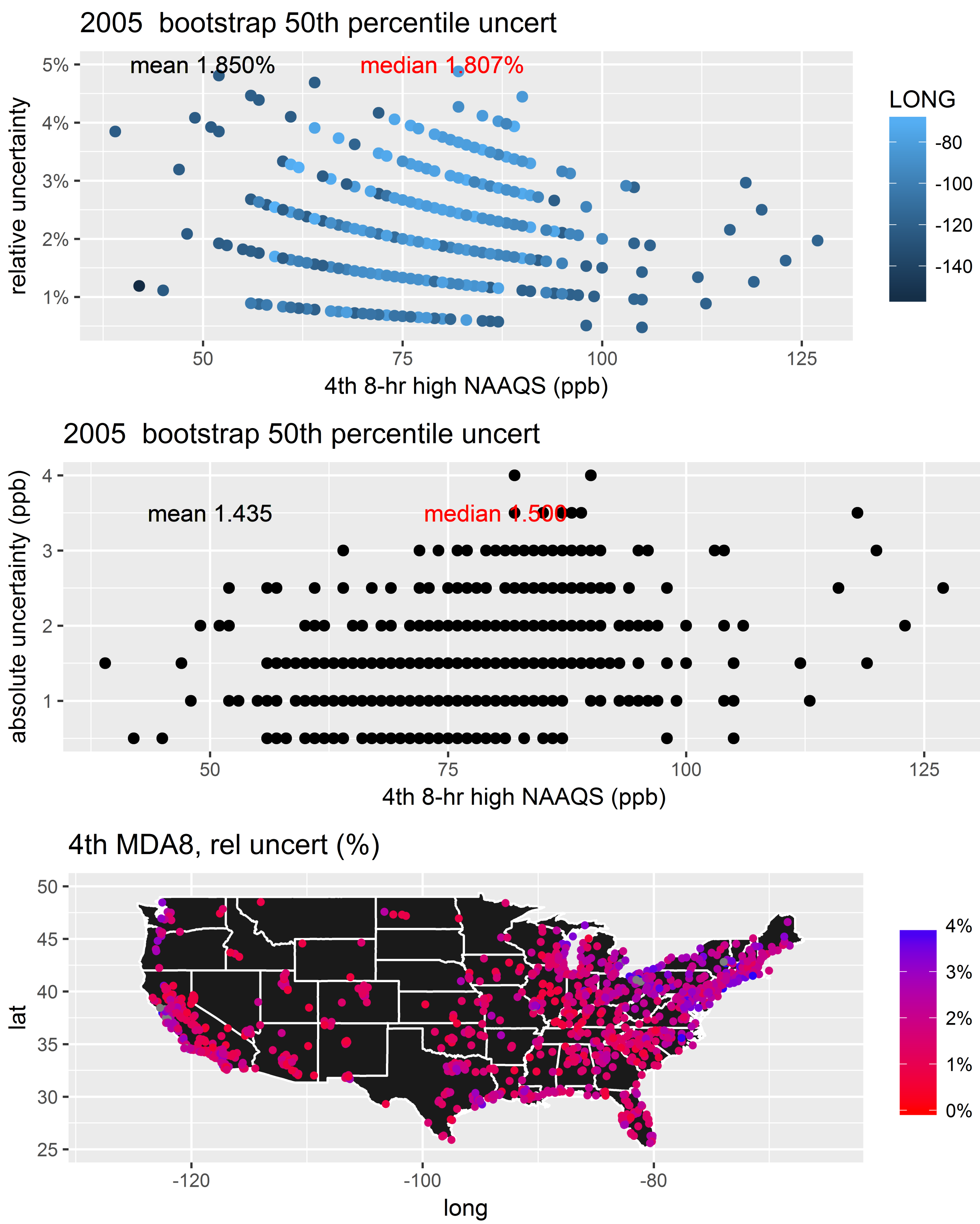


Figure 27: Bootstrap results from the 50% CIs for the 2005 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2005 ozone DV at each site.

2.12 2002-2004 ozone bootstrap results

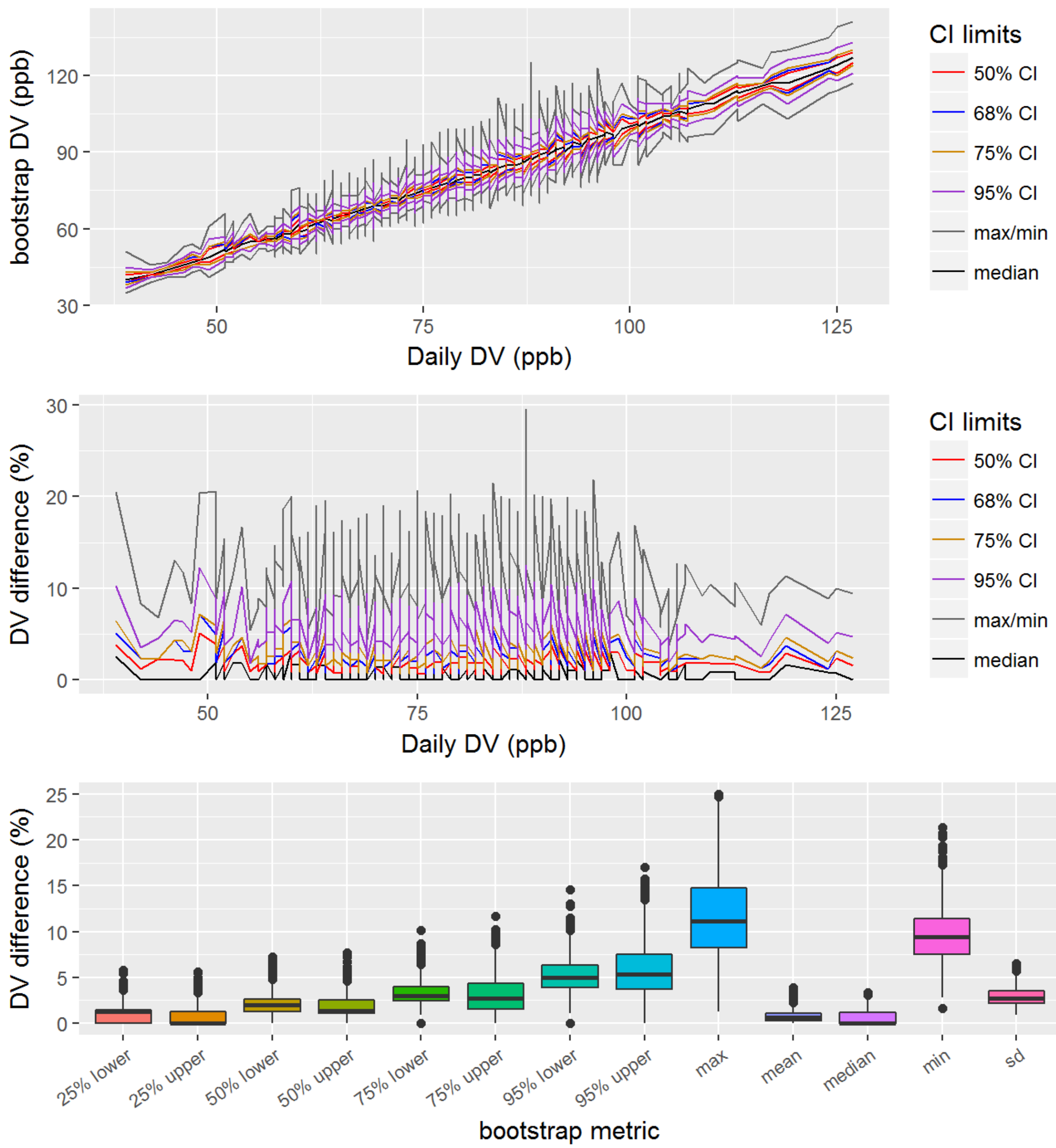


Figure 28: Bootstrap results for the ozone 2004 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

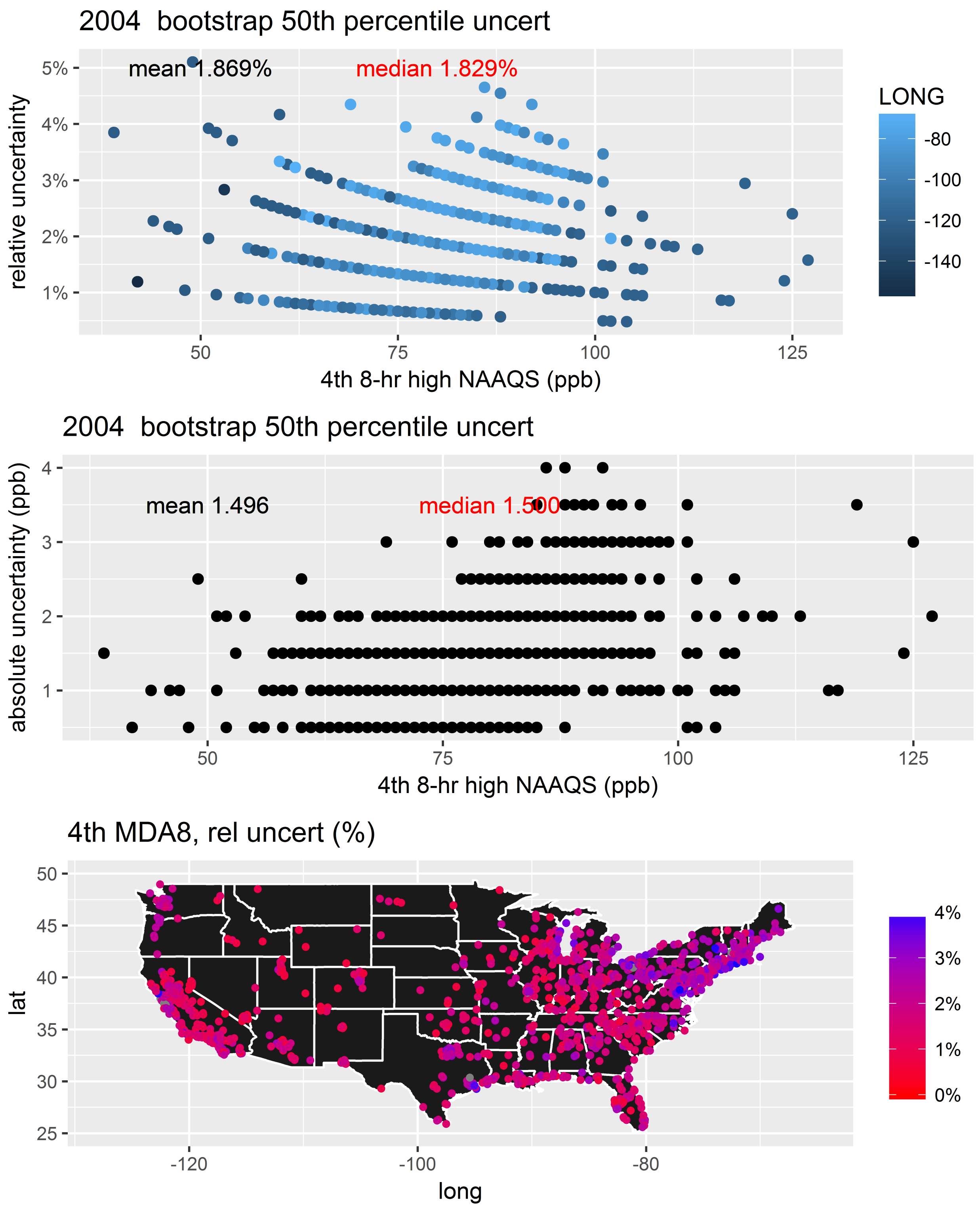


Figure 29: Bootstrap results from the 50% CIs for the 2004 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2004 ozone DV at each site.

2.13 2001-2003 ozone bootstrap results

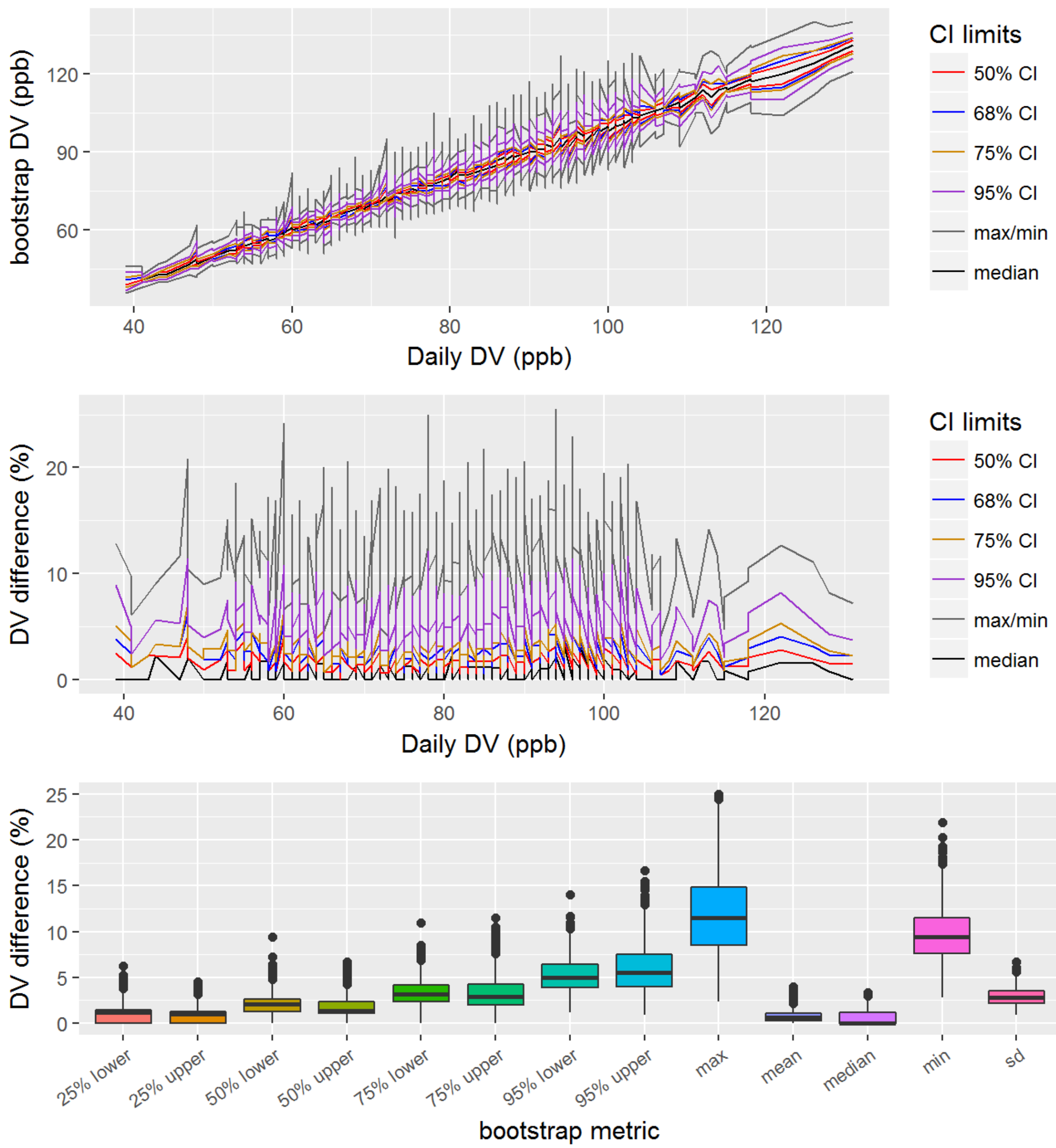


Figure 30: Bootstrap results for the ozone 2003 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

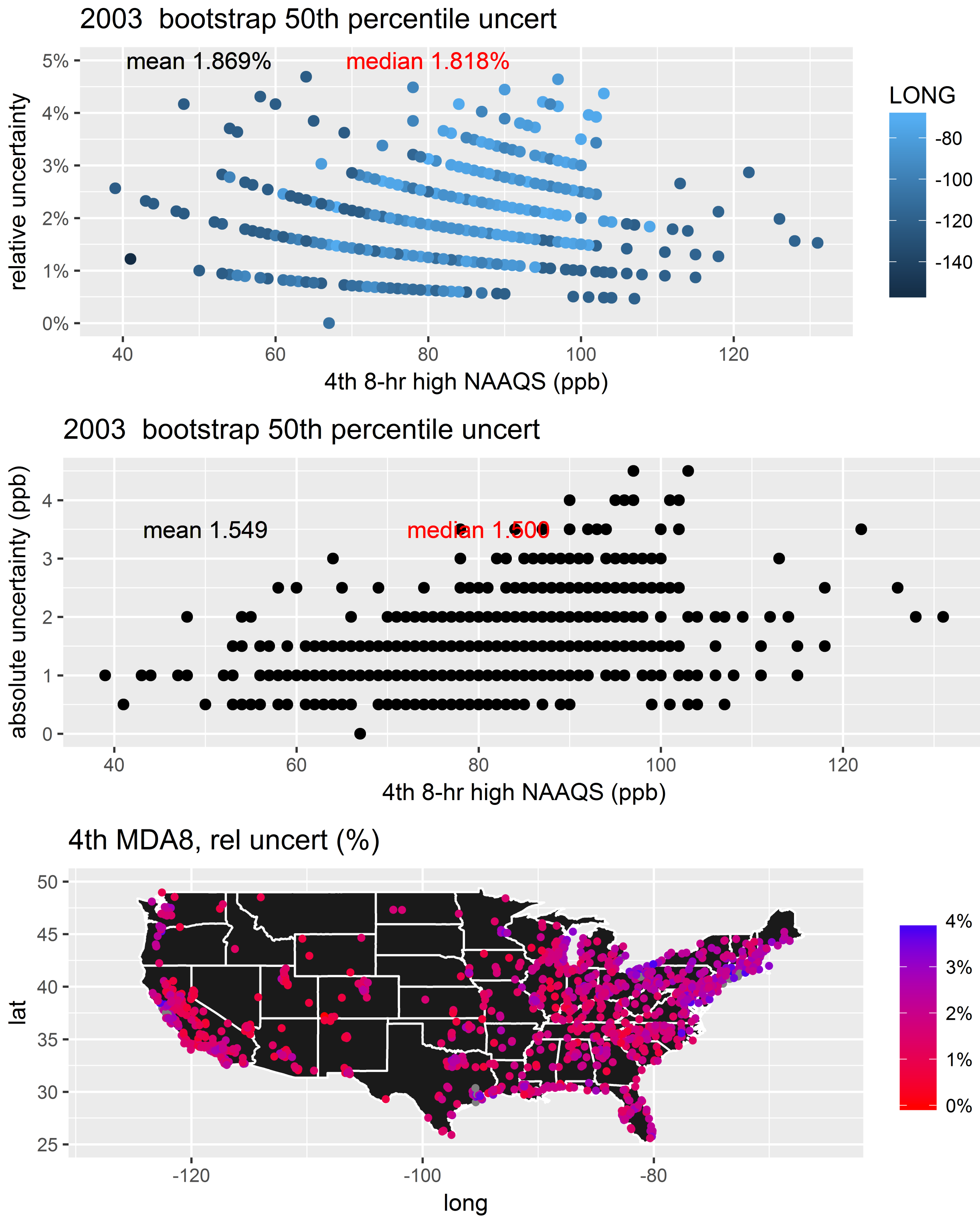


Figure 31: Bootstrap results from the 50% CIs for the 2003 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2003 ozone DV at each site.

2.14 2000-2002 ozone bootstrap results

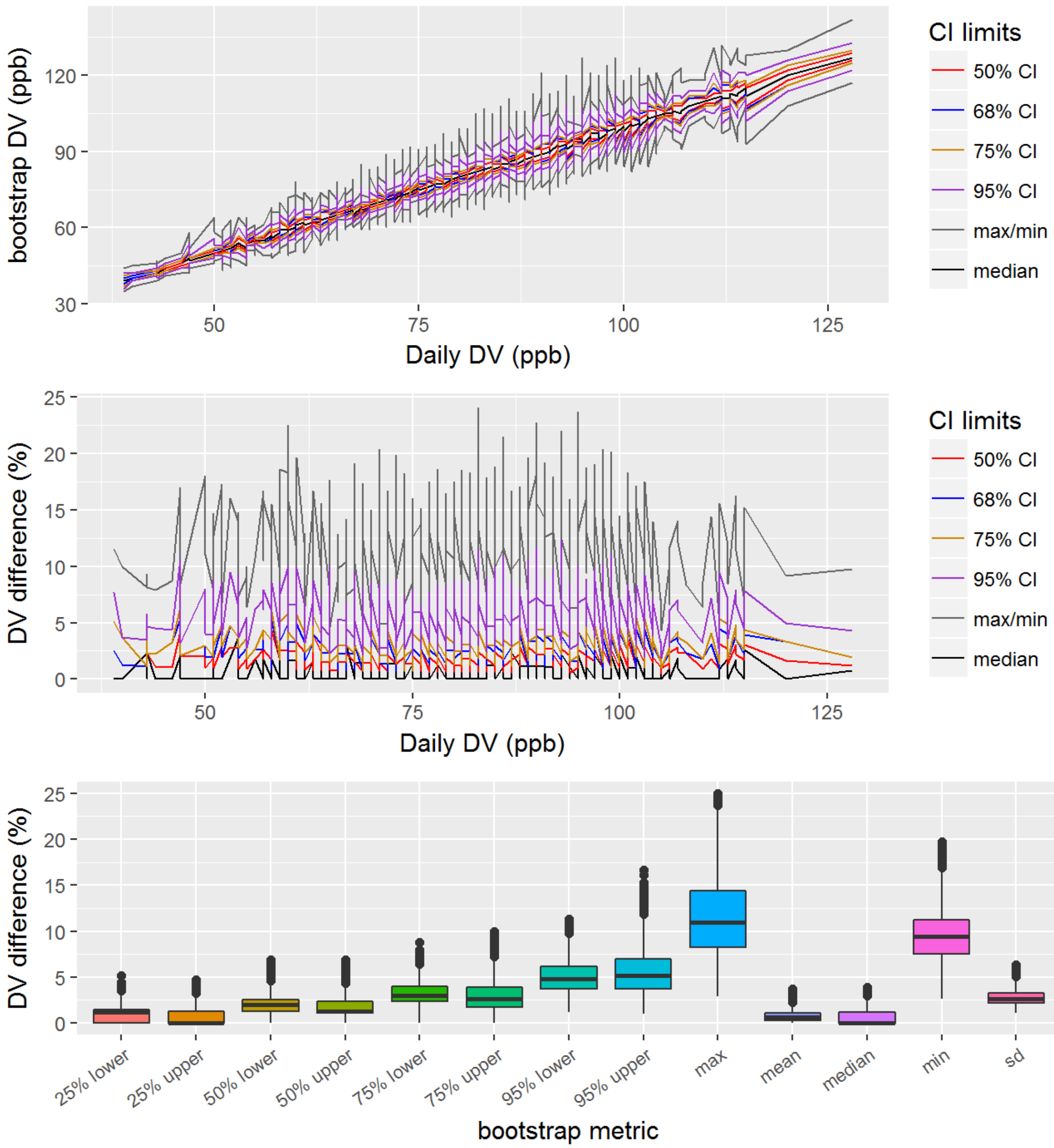


Figure 32: Bootstrap results for the ozone 2002 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

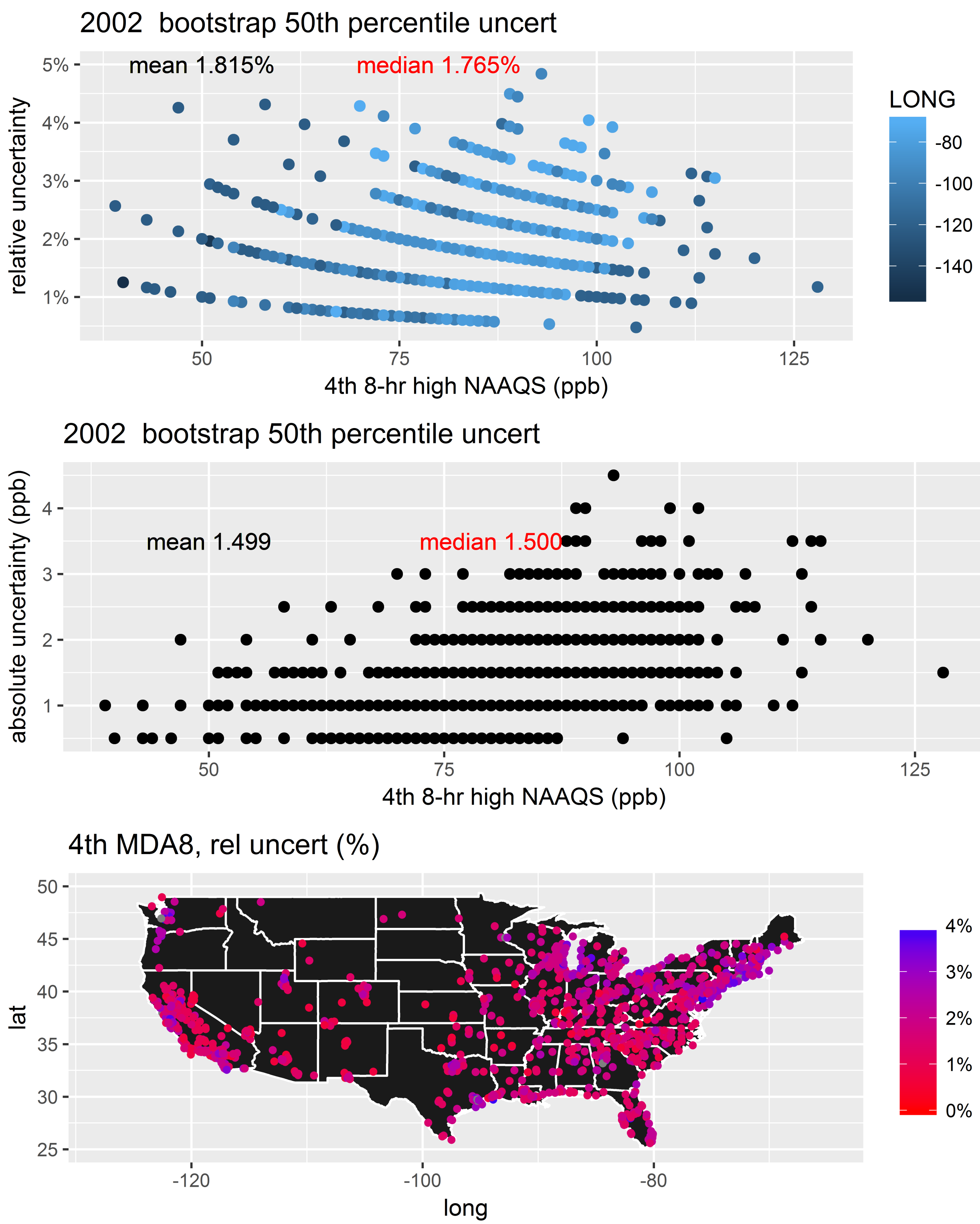


Figure 33: Bootstrap results from the 50% CIs for the 2002 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2002 ozone DV at each site.

3 Air quality variability results for years 2002-2013 for $PM_{2.5}$

Bootstrap results for $PM_{2.5}$ data from the years 2000-2015. Each section containts a single DV period,*e.g.*, the results for 2015 include data from 2013-2015.

3.1 2013-2015 $PM_{2.5}$ bootstrap results

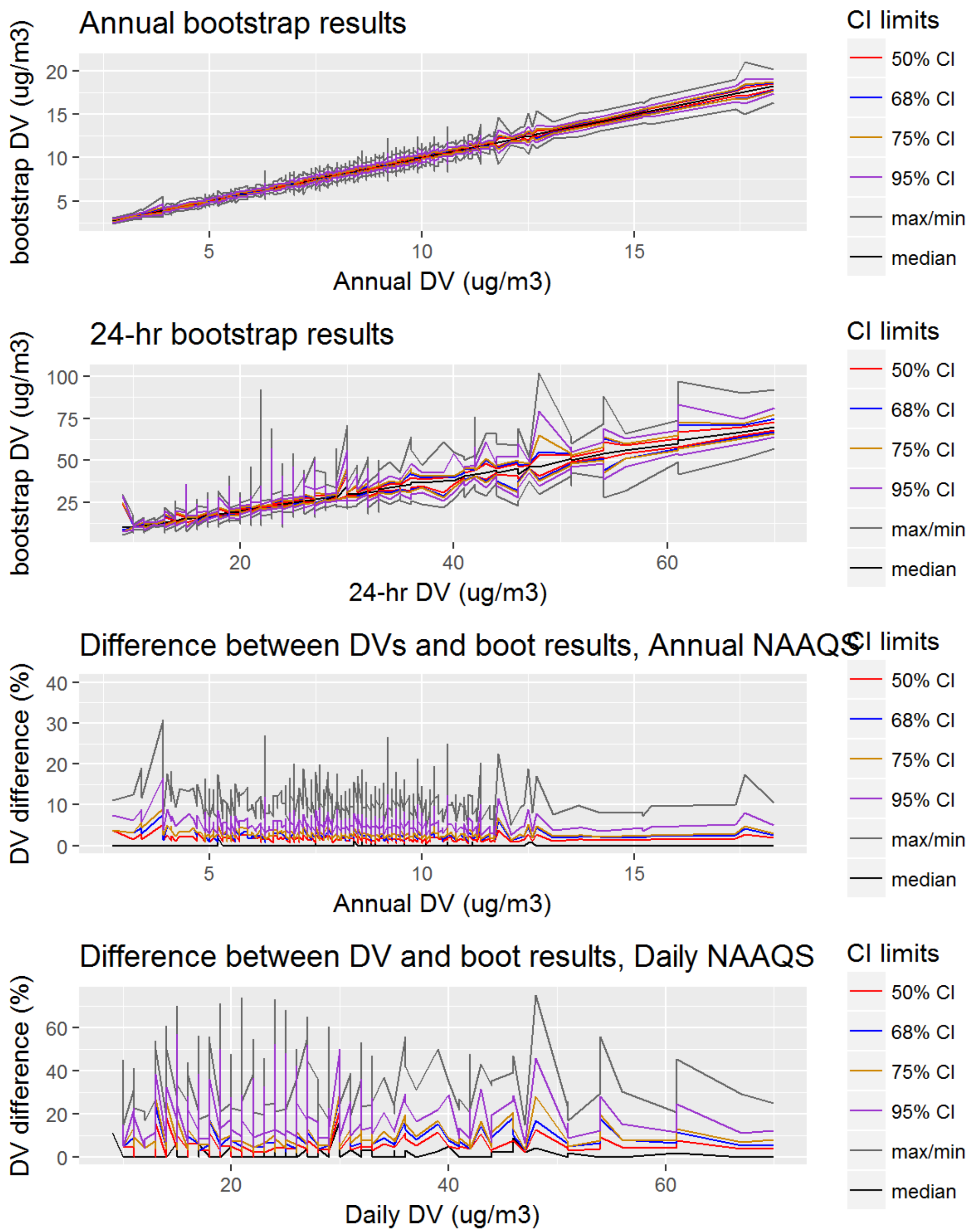


Figure 34: Bootstrap results for the 2015 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

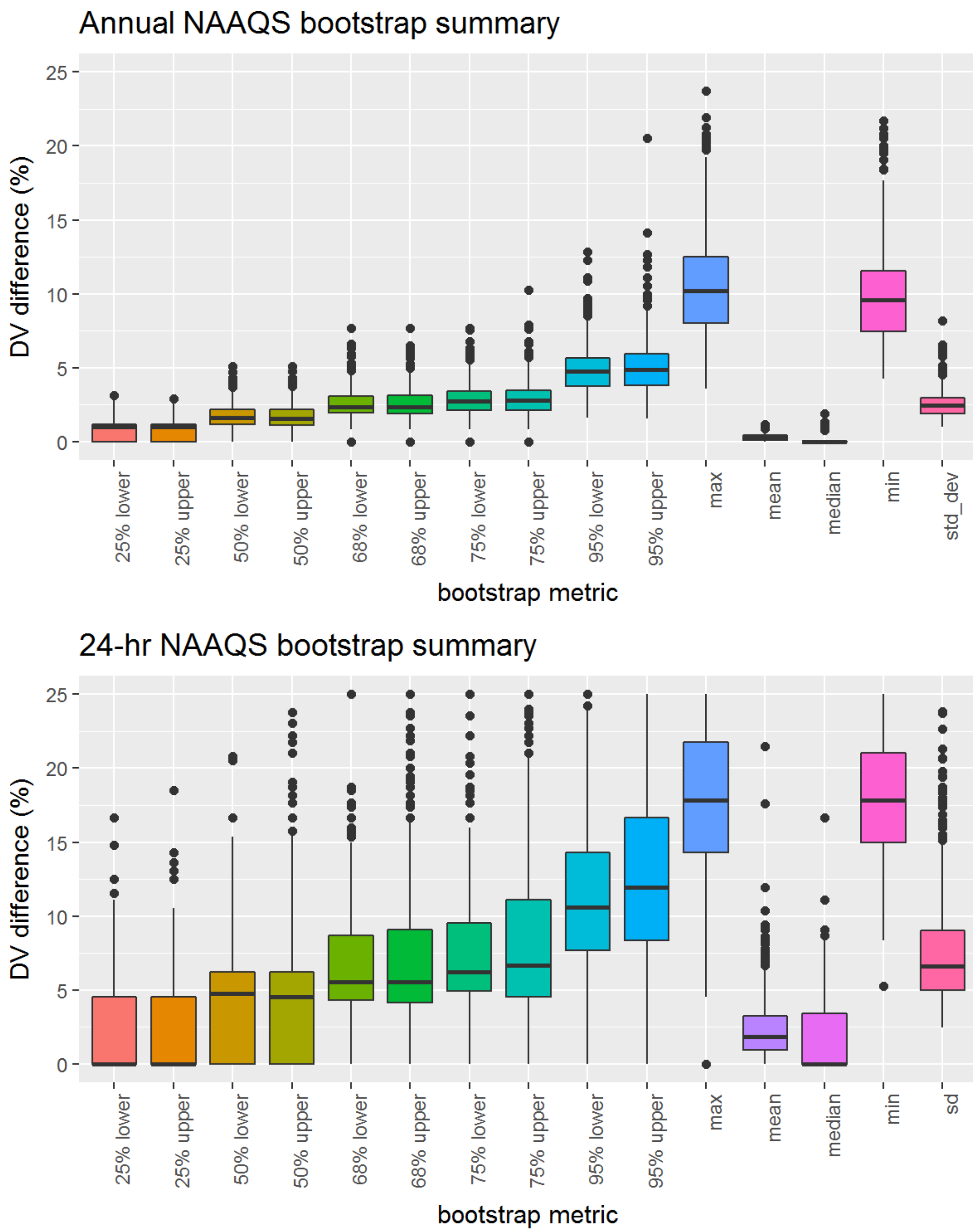


Figure 35: Bootstrap results for the 2015 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

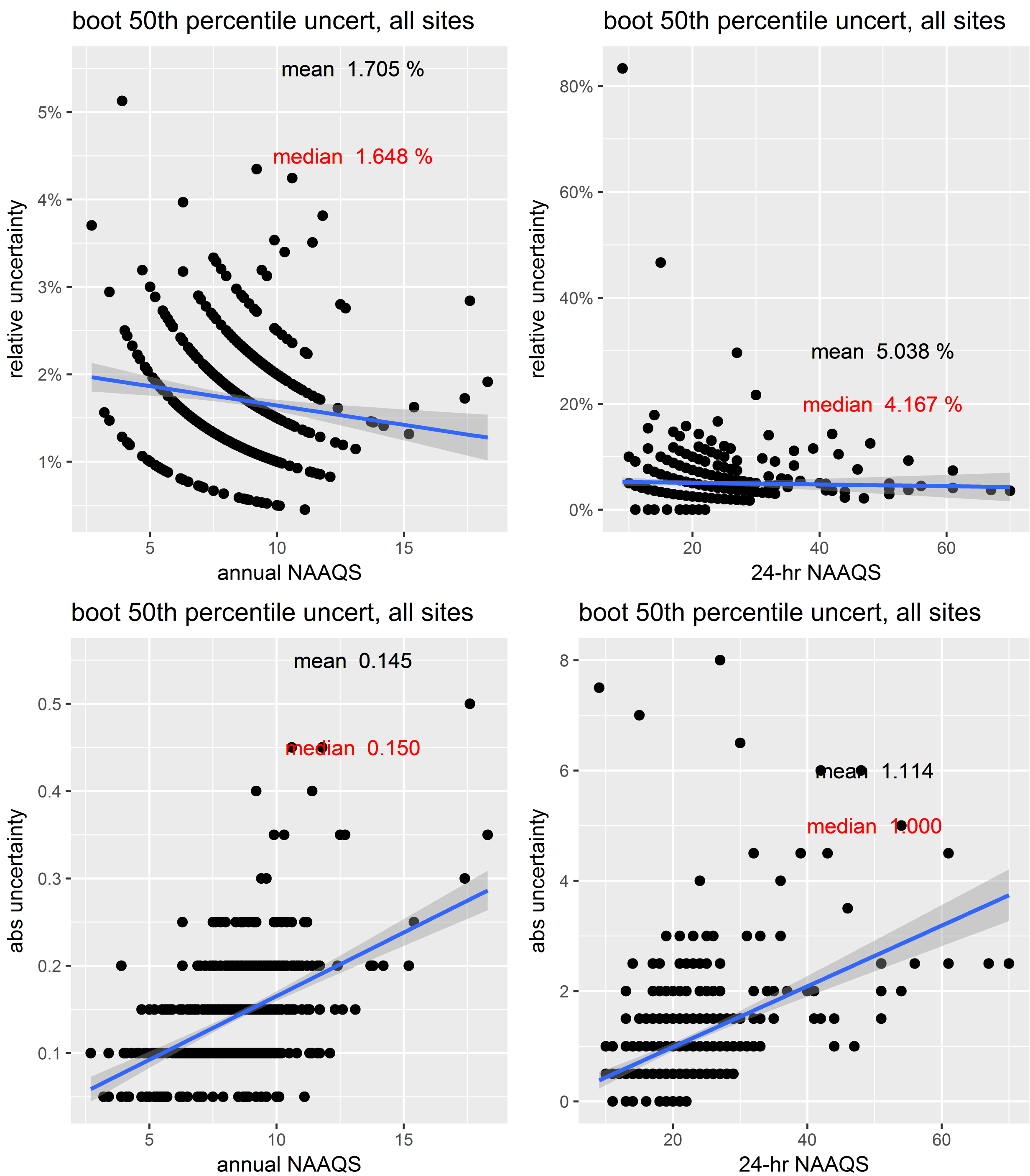


Figure 36: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

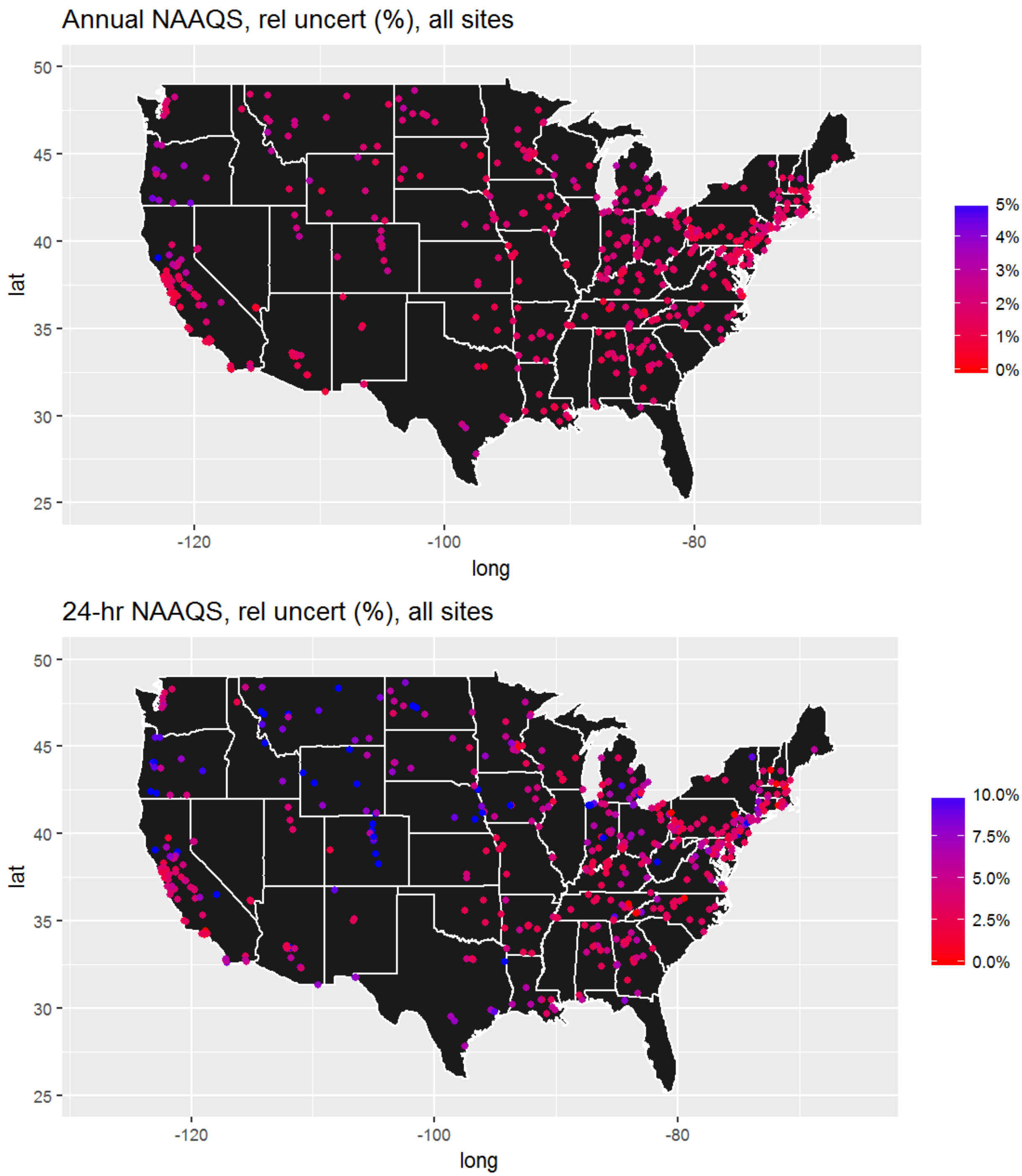


Figure 37: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2015 $PM_{2.5}$ DVs.

3.2 2012-2014 $PM_{2.5}$ bootstrap results

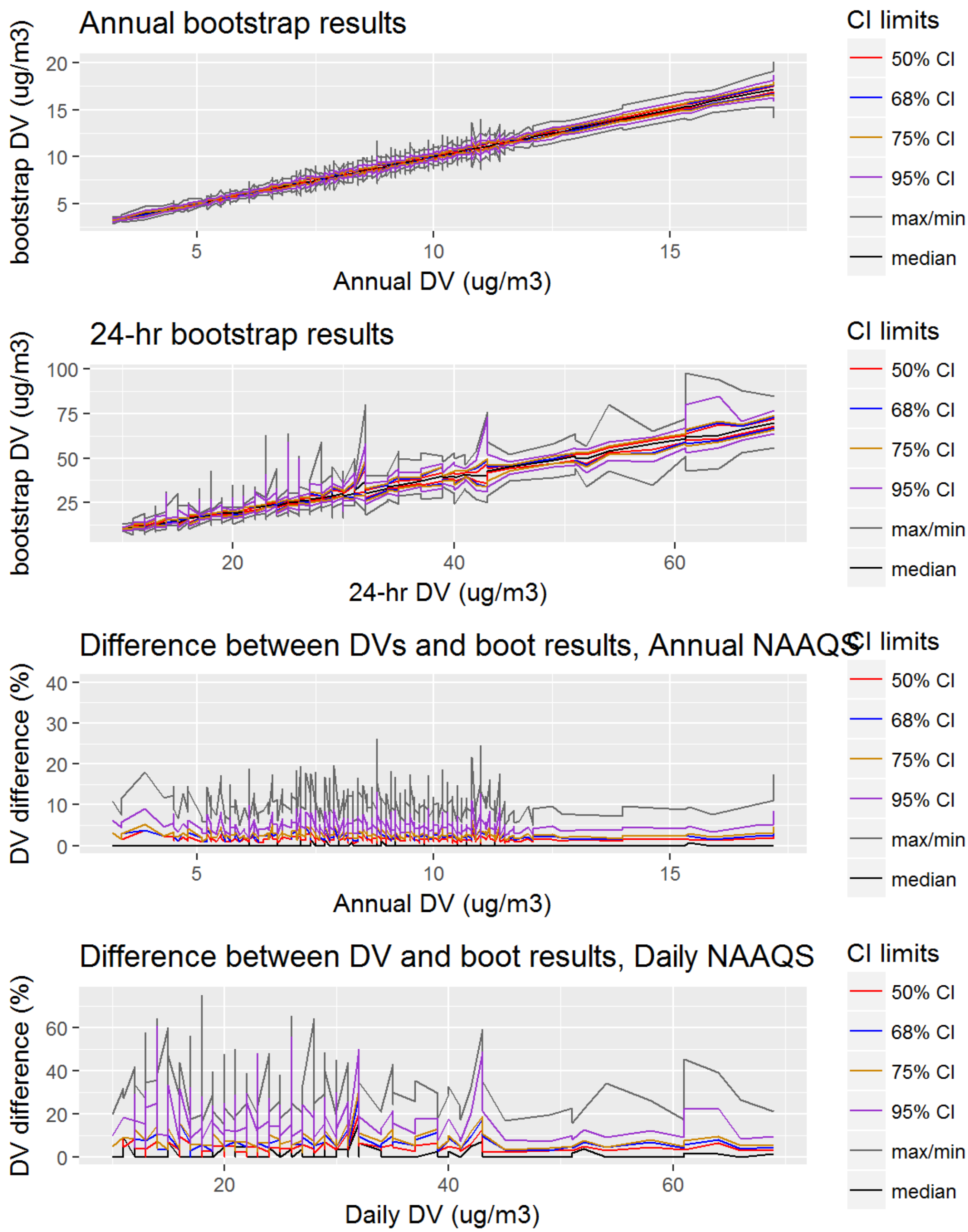


Figure 38: Bootstrap results for the 2014 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

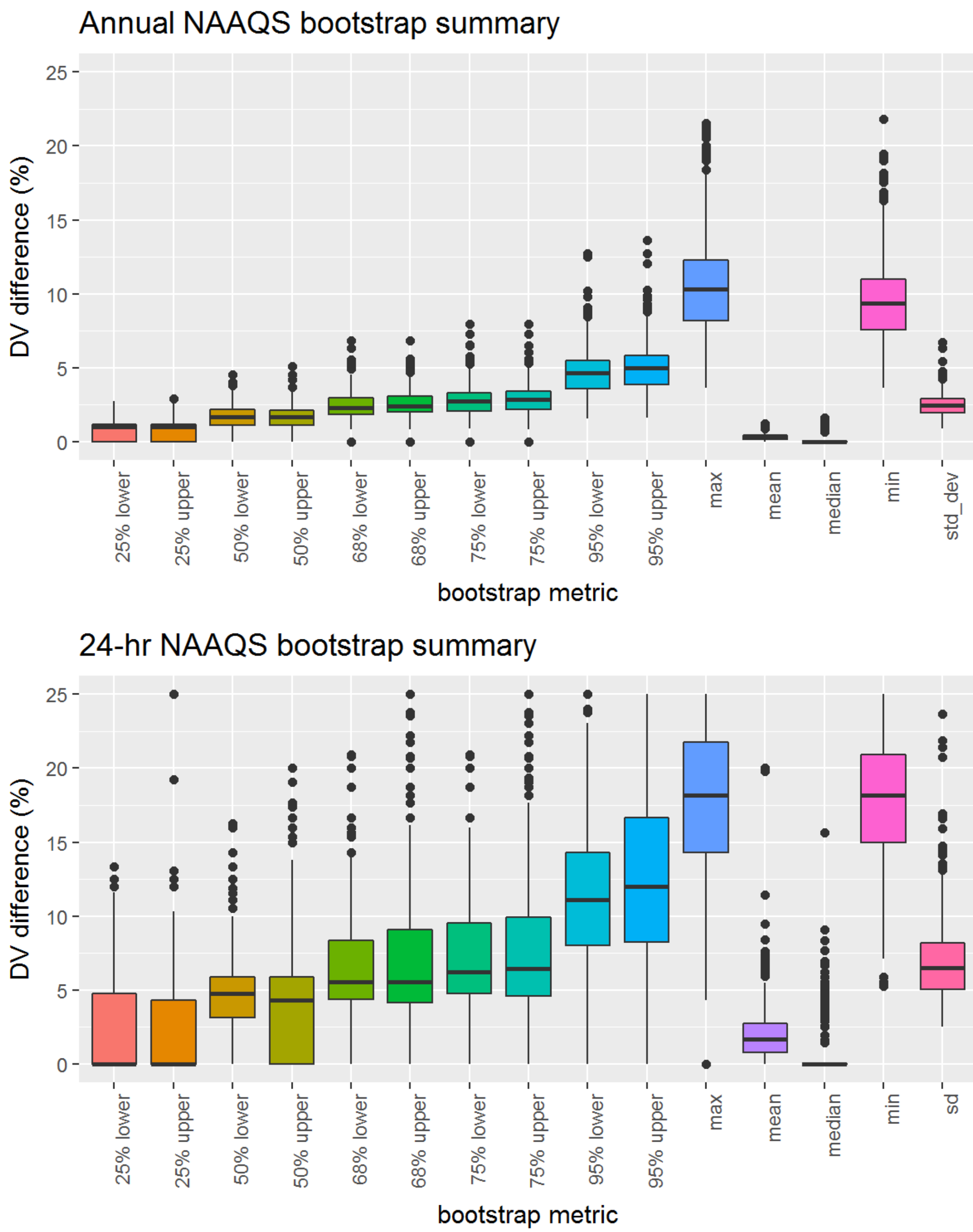


Figure 39: Bootstrap results for the 2014 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

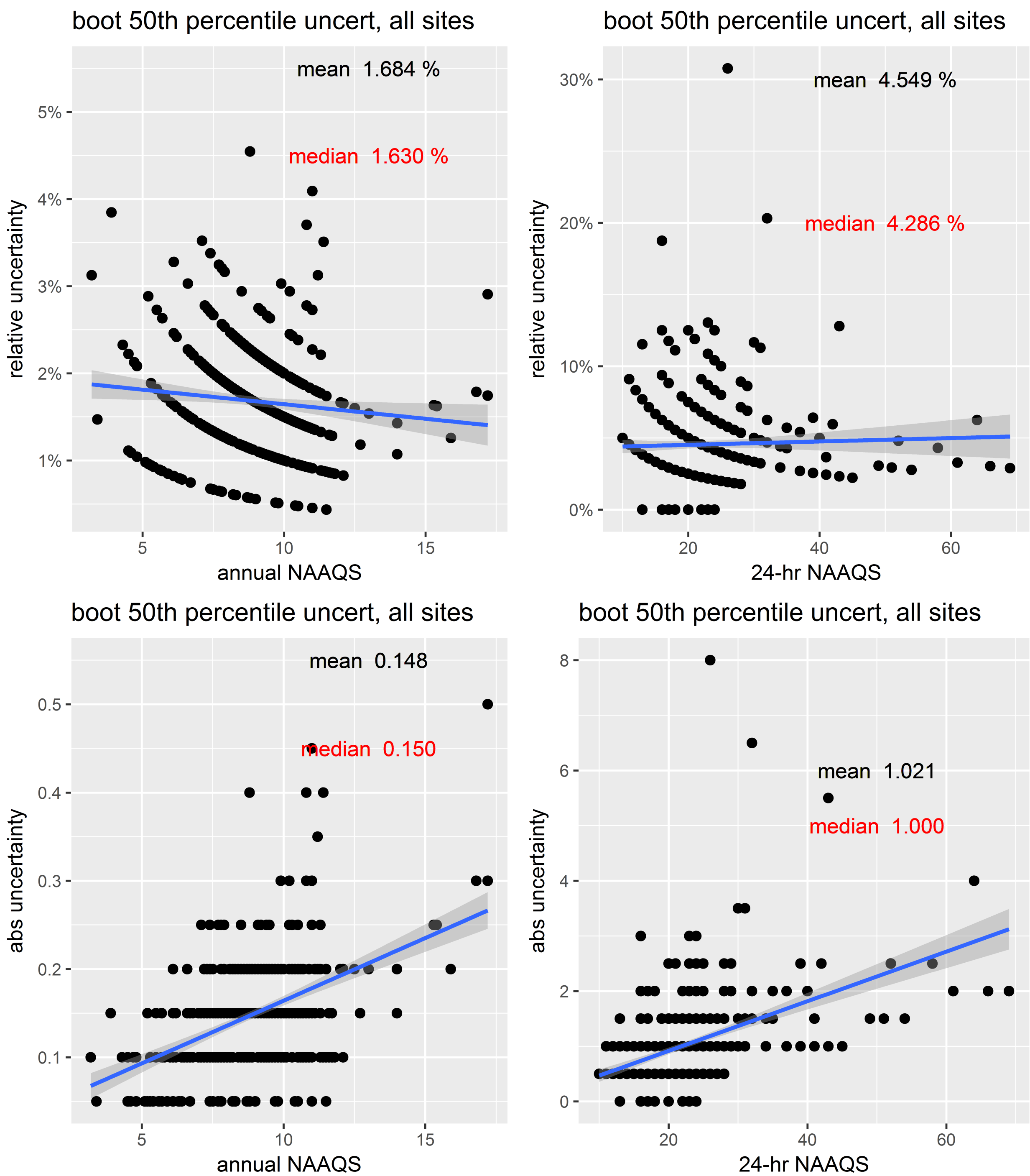


Figure 40: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

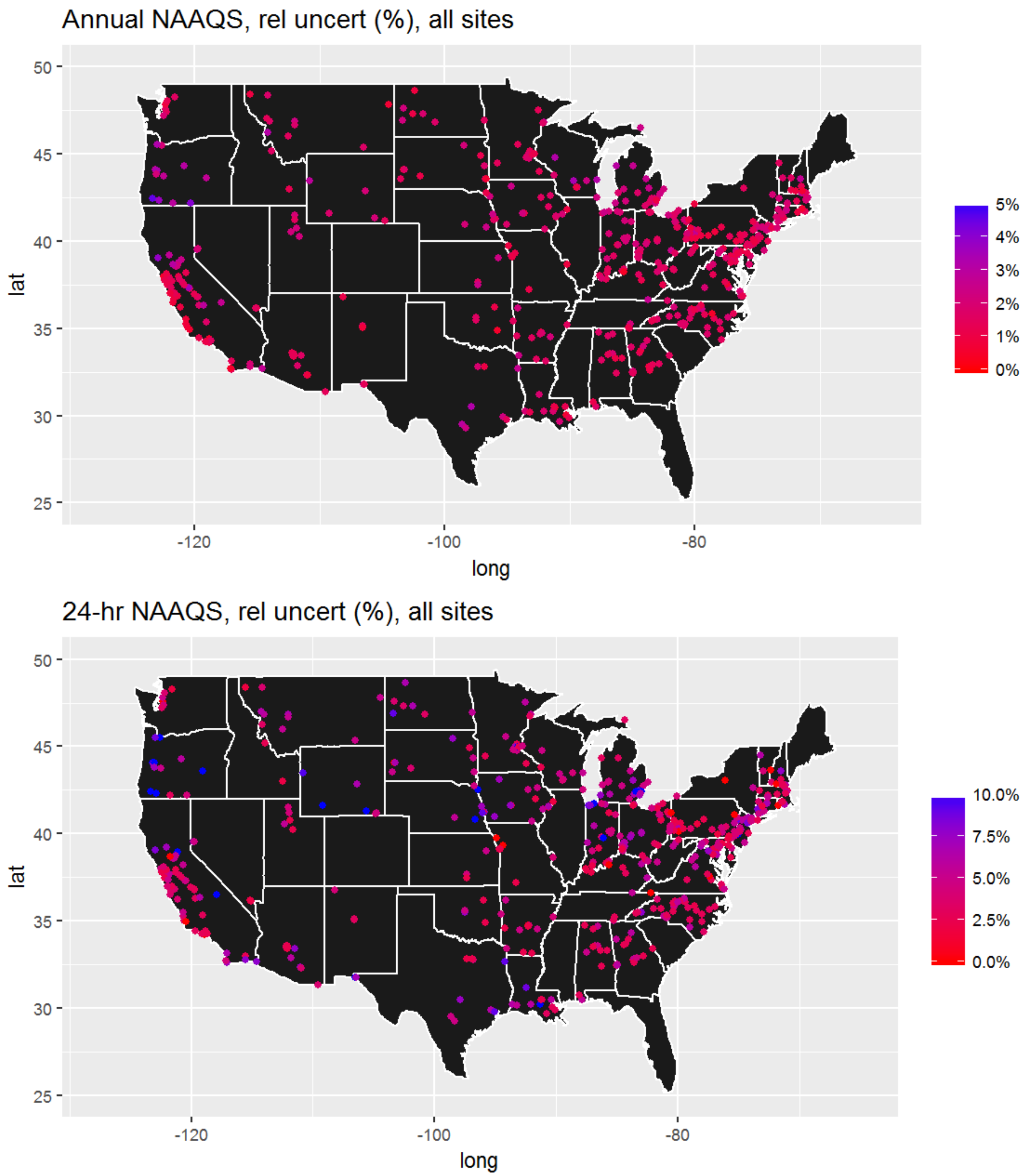


Figure 41: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2014 $PM_{2.5}$ DVs.

3.3 2011-2013 $PM_{2.5}$ bootstrap results

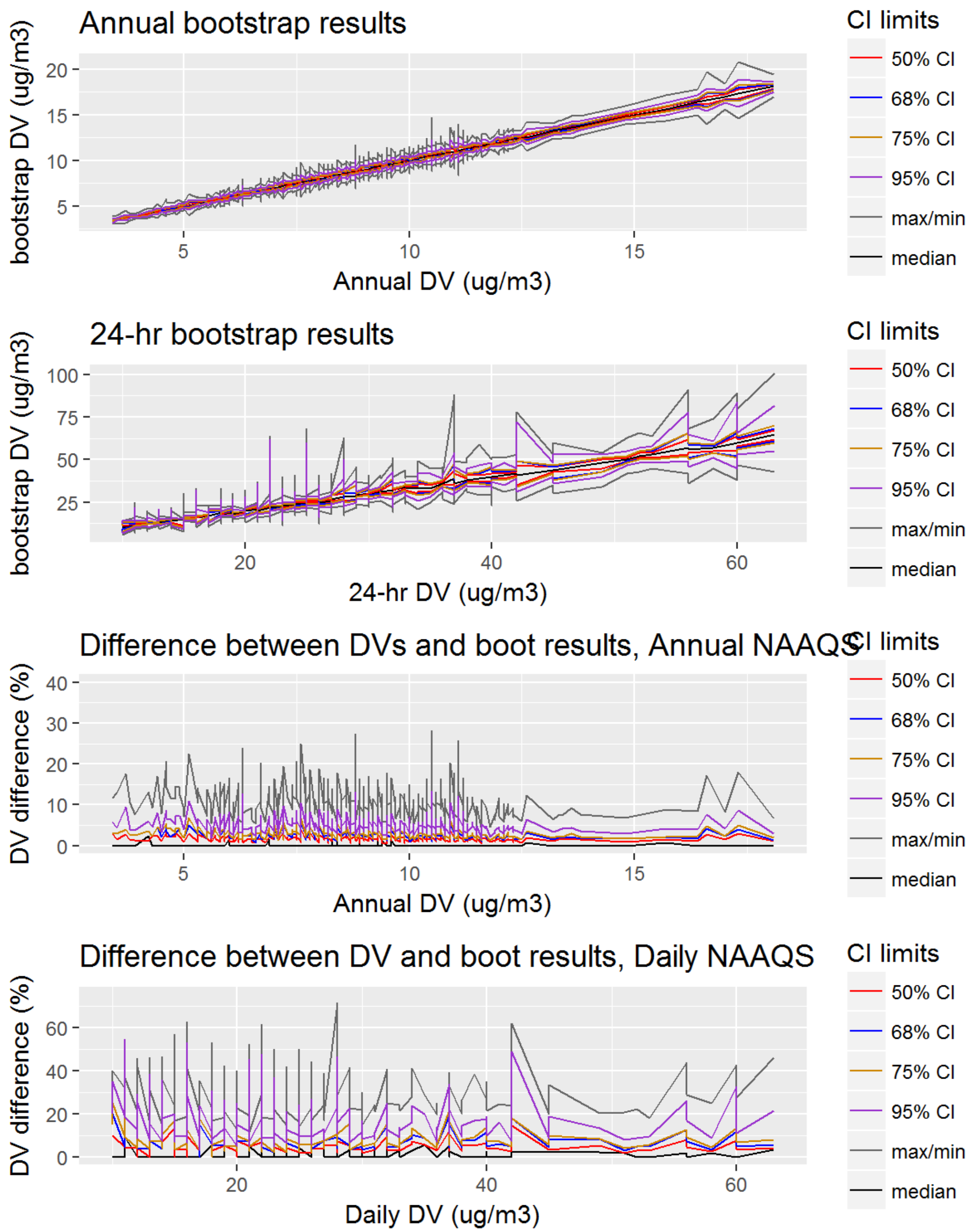


Figure 42: Bootstrap results for the 2013 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

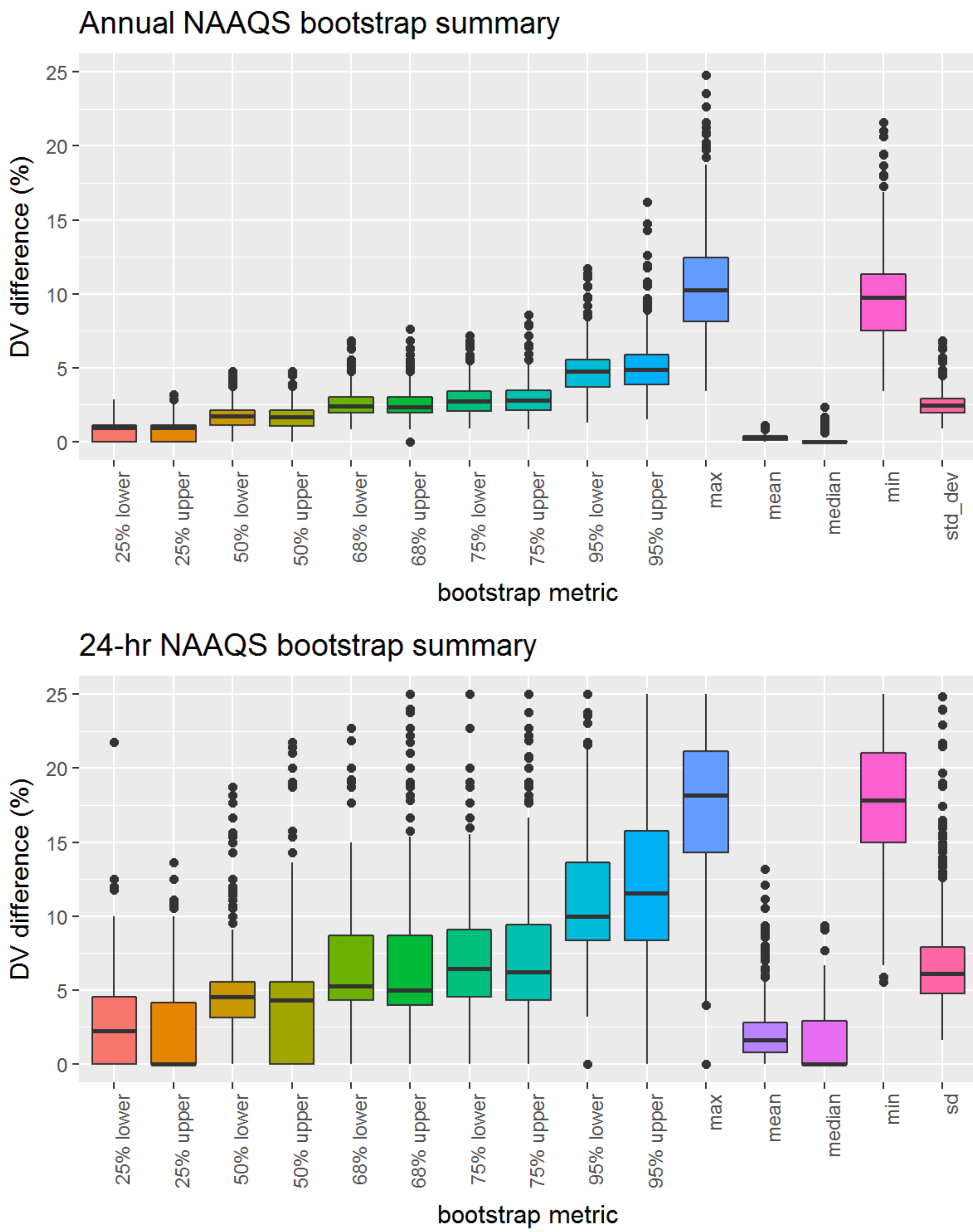


Figure 43: Bootstrap results for the 2013 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

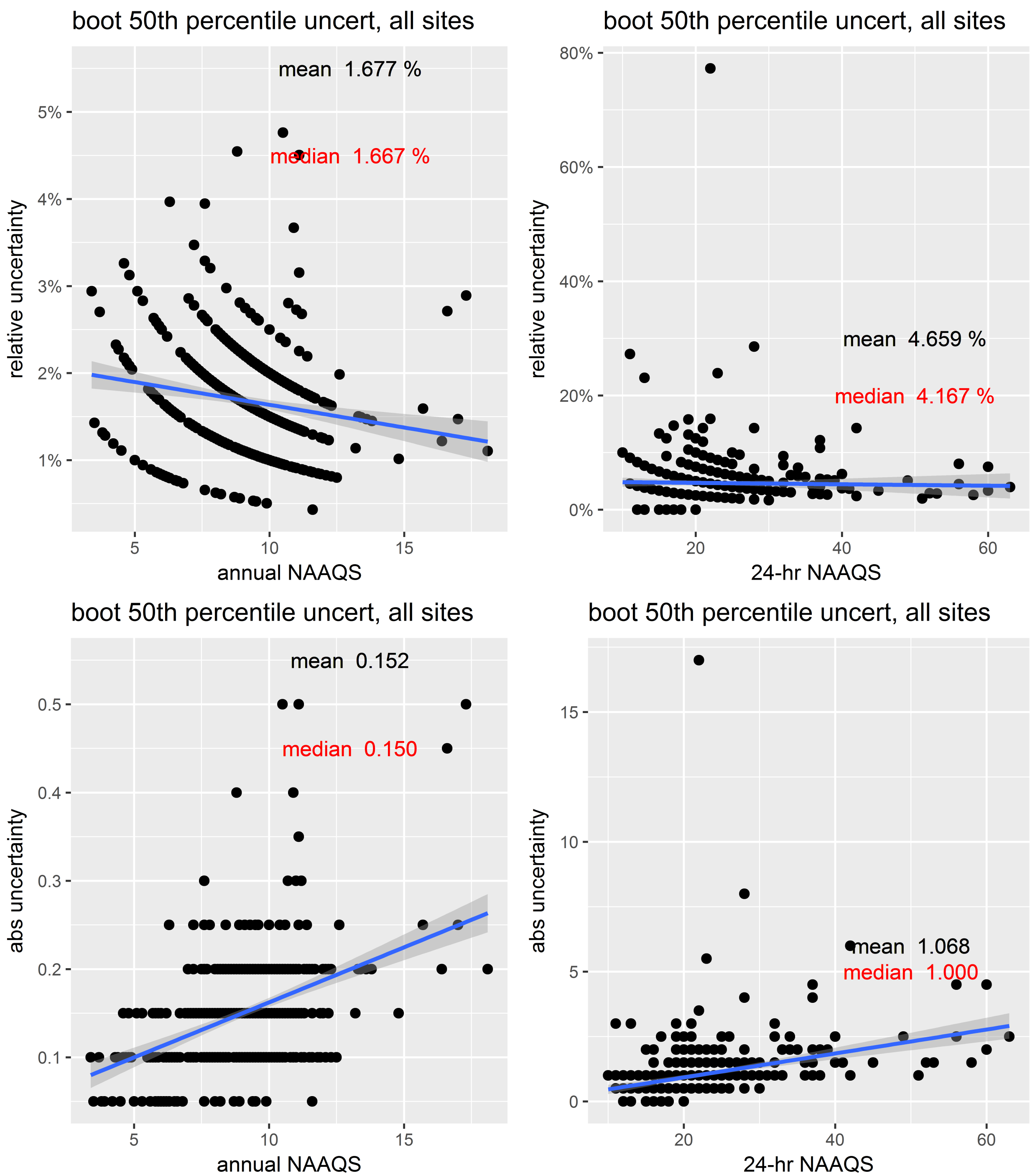


Figure 44: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

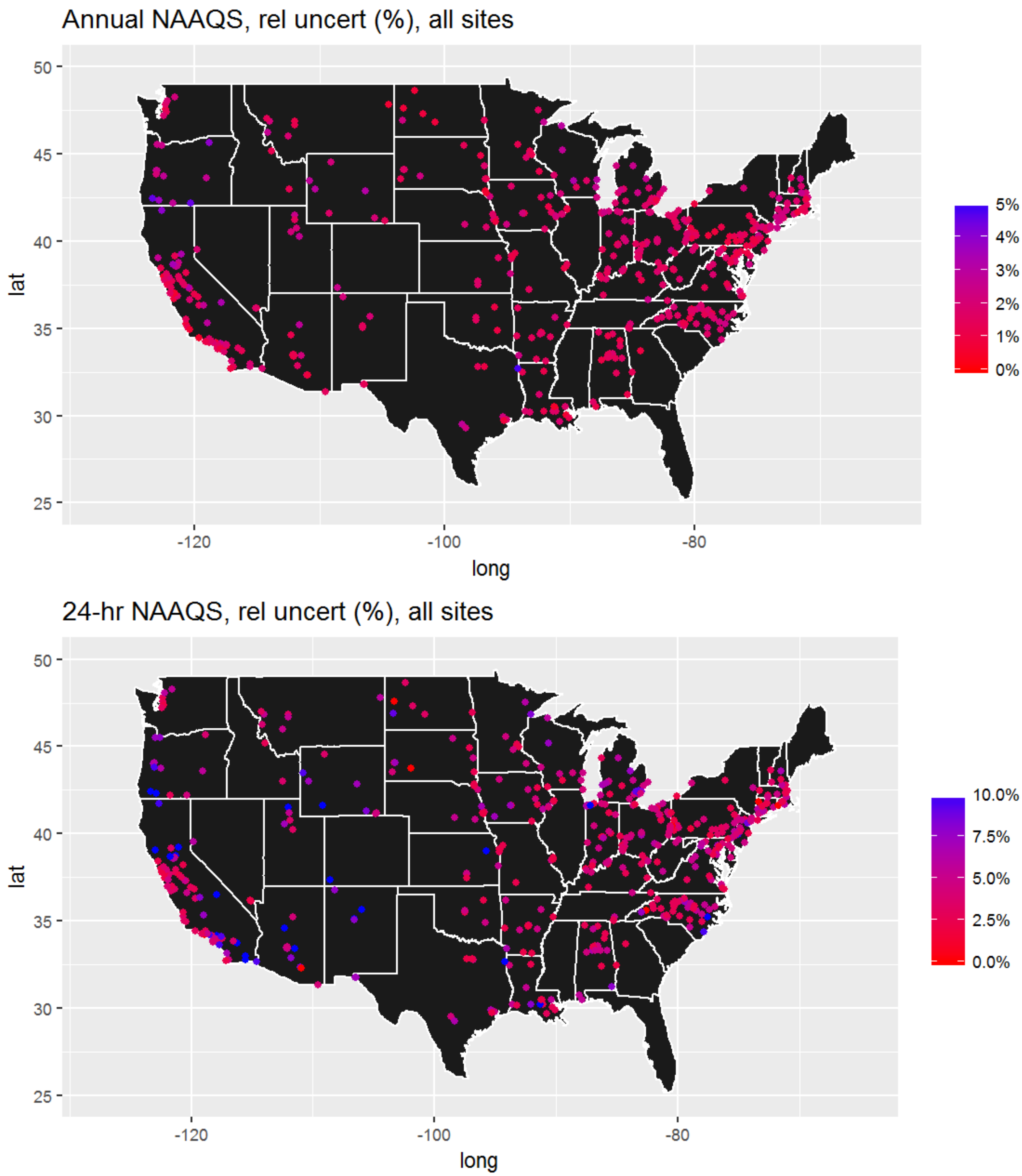


Figure 45: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2013 $PM_{2.5}$ DVs.

3.4 2010-2012 $PM_{2.5}$ bootstrap results

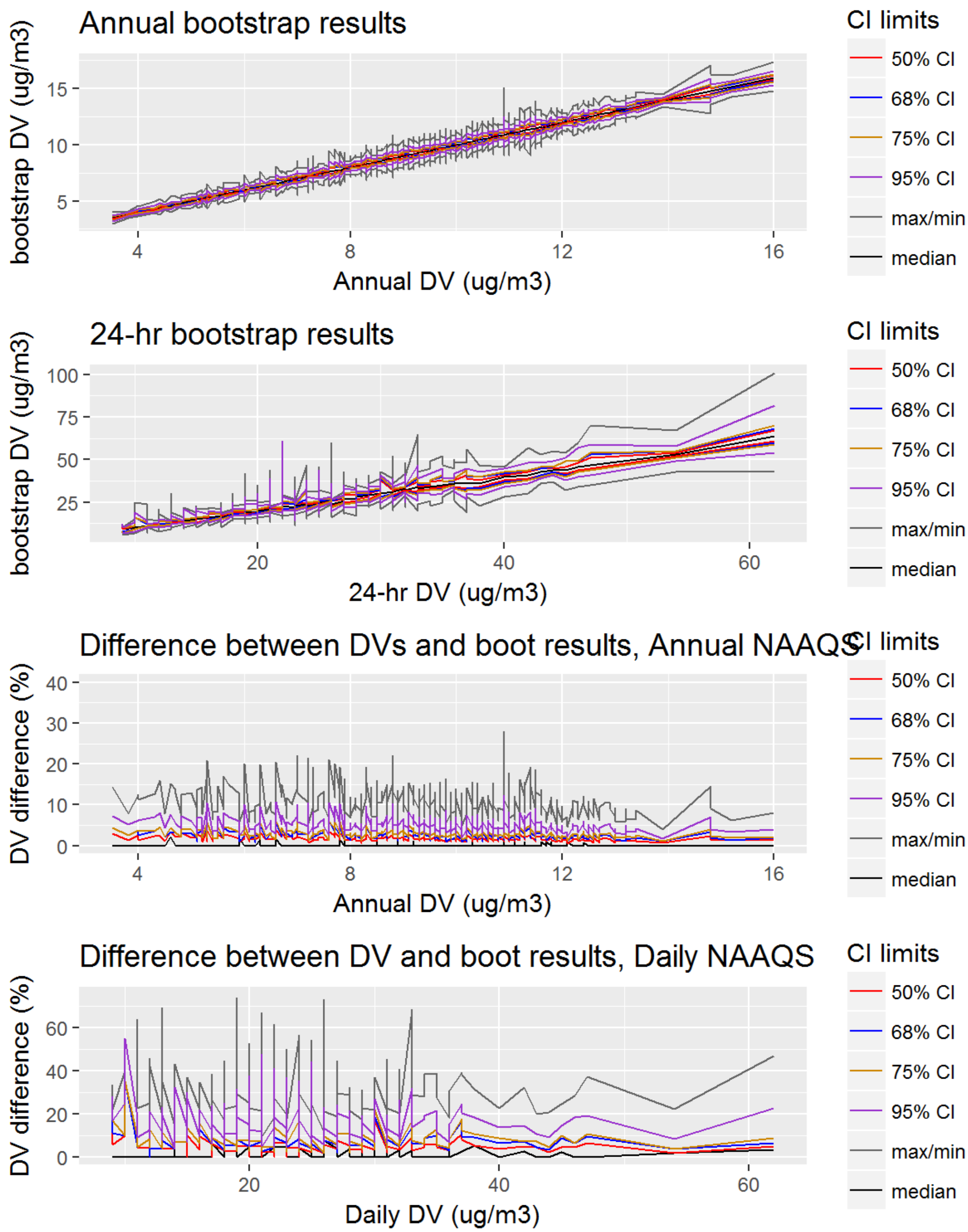


Figure 46: Bootstrap results for the 2012 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

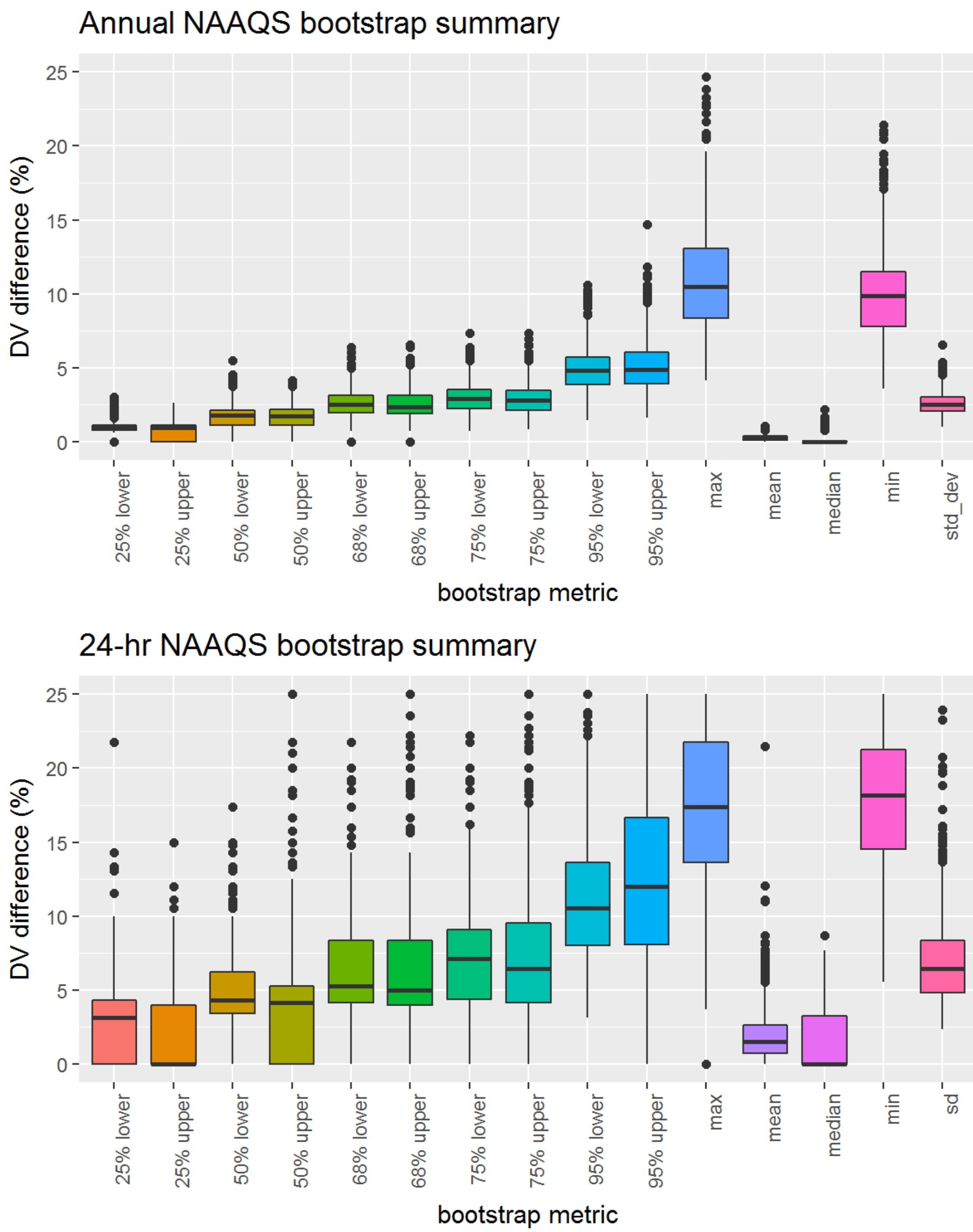


Figure 47: Bootstrap results for the 2012 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

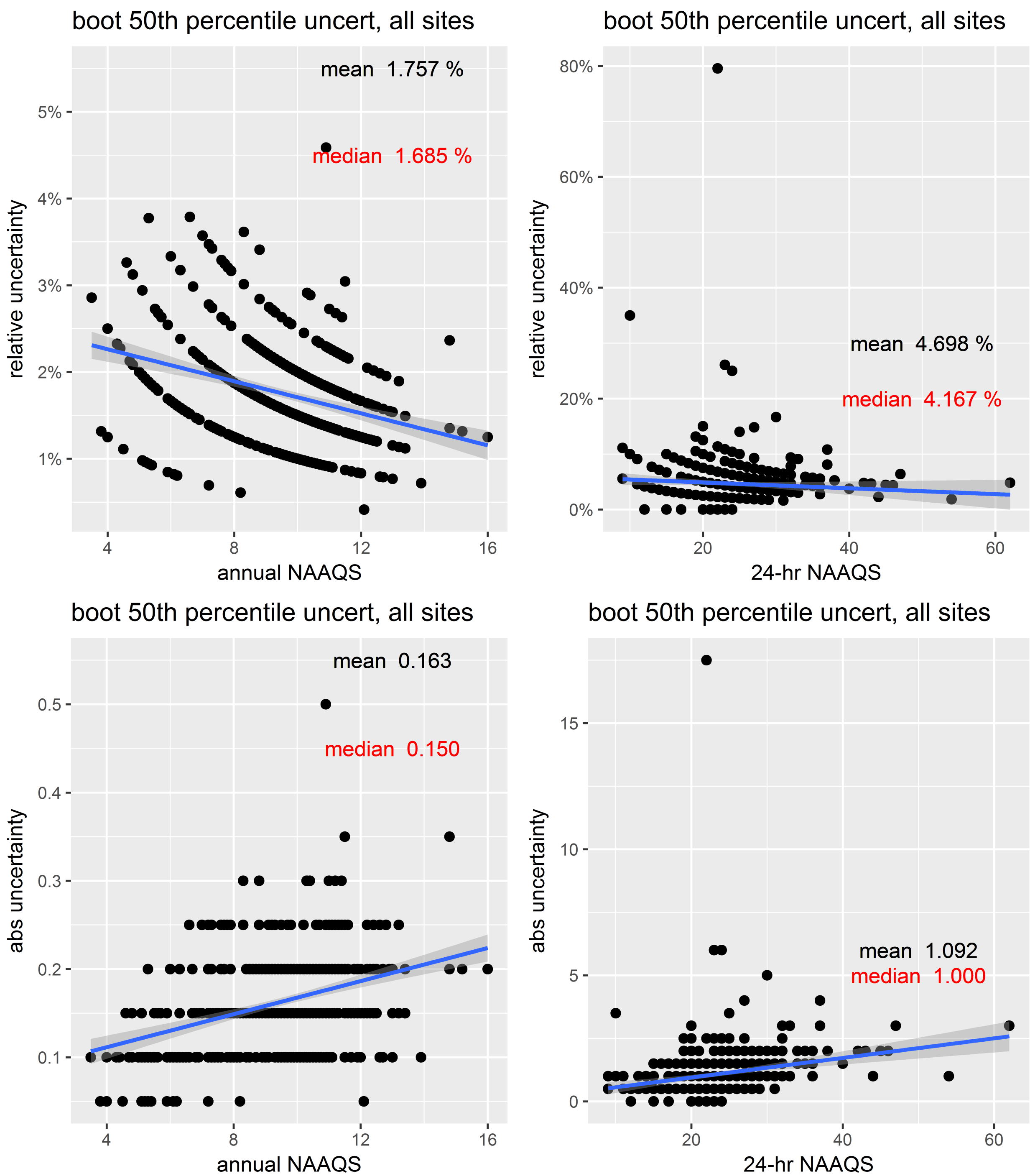


Figure 48: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

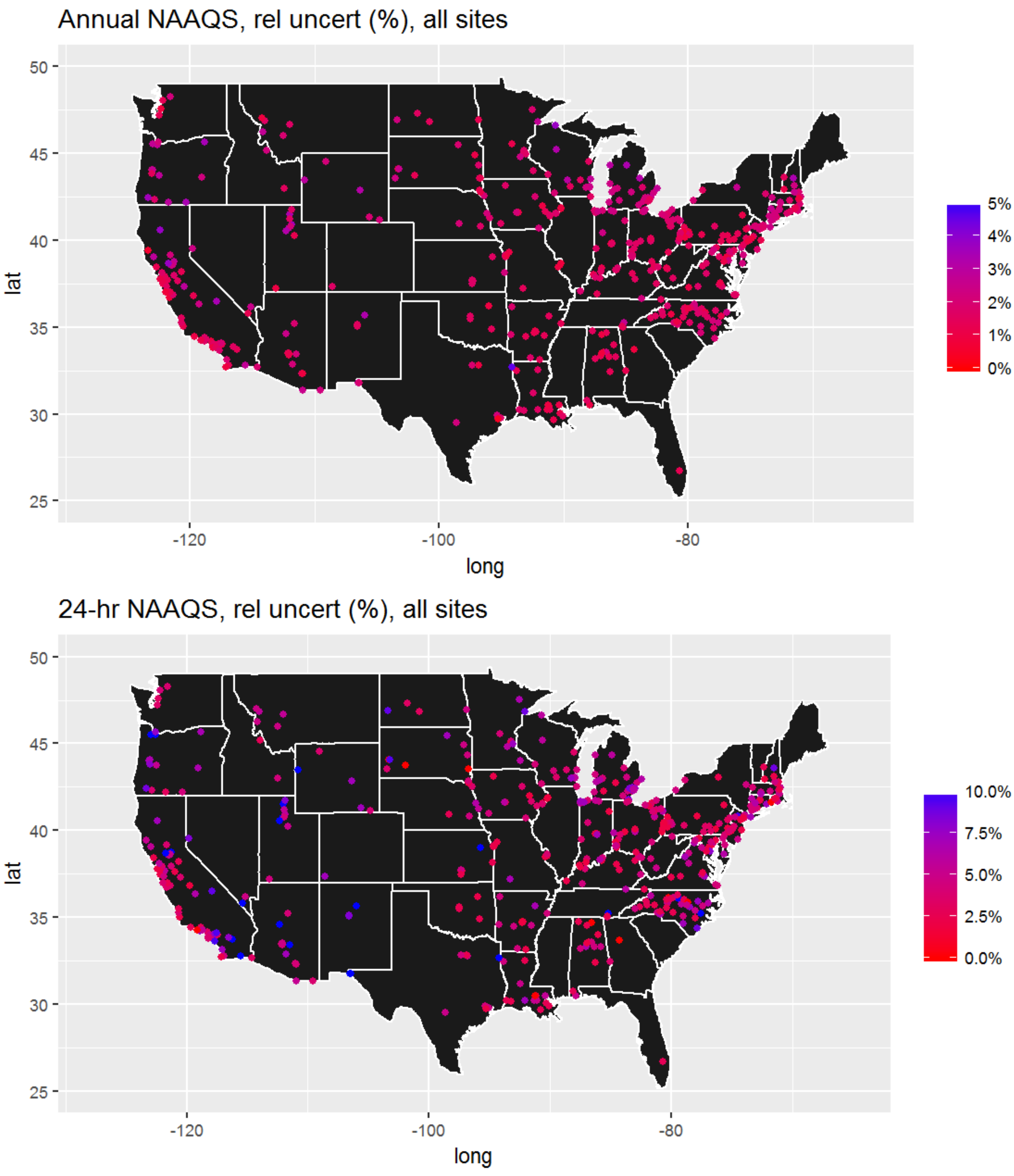


Figure 49: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2012 $PM_{2.5}$ DVs.

3.5 2009-2011 $PM_{2.5}$ bootstrap results

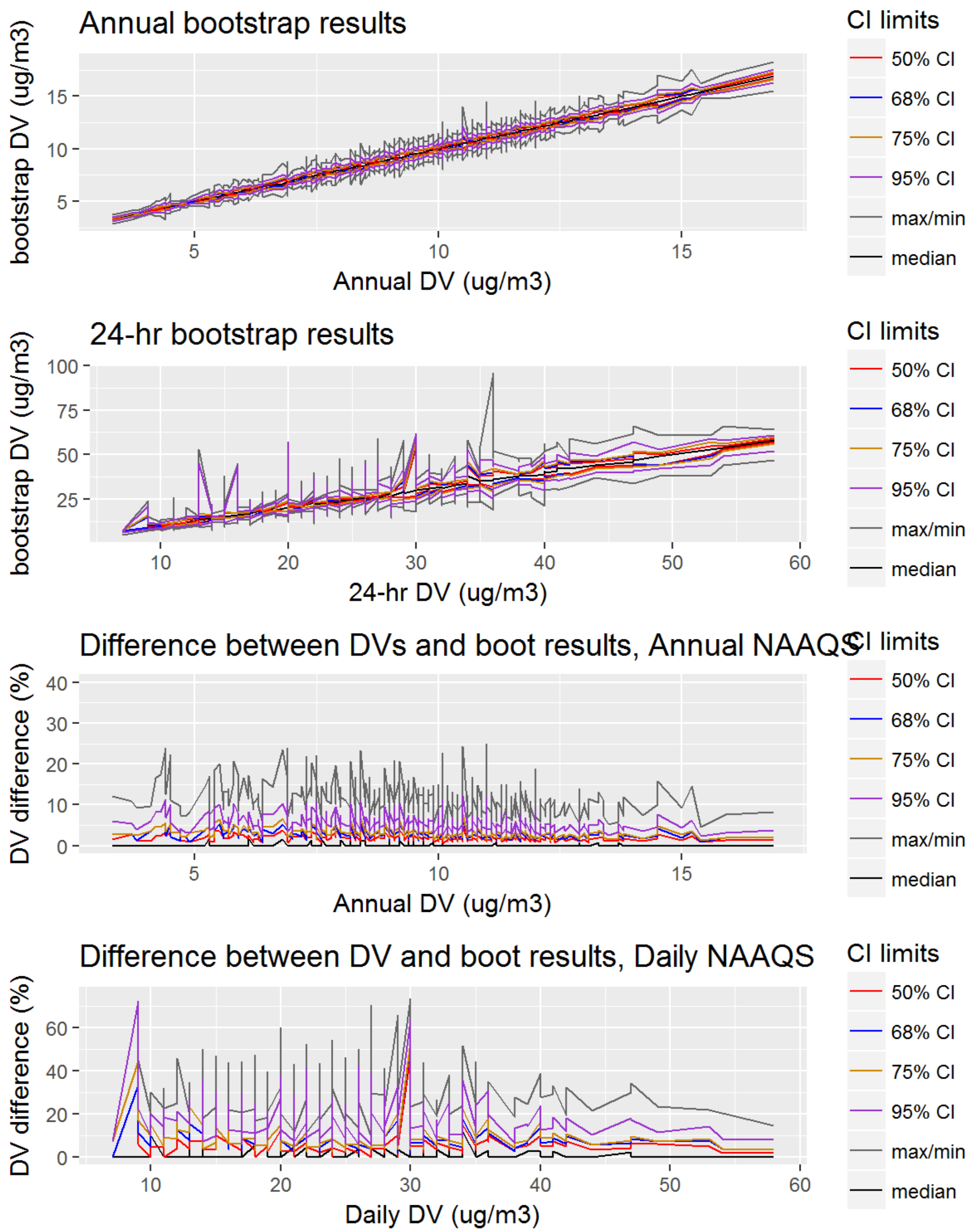


Figure 50: Bootstrap results for the 2011 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

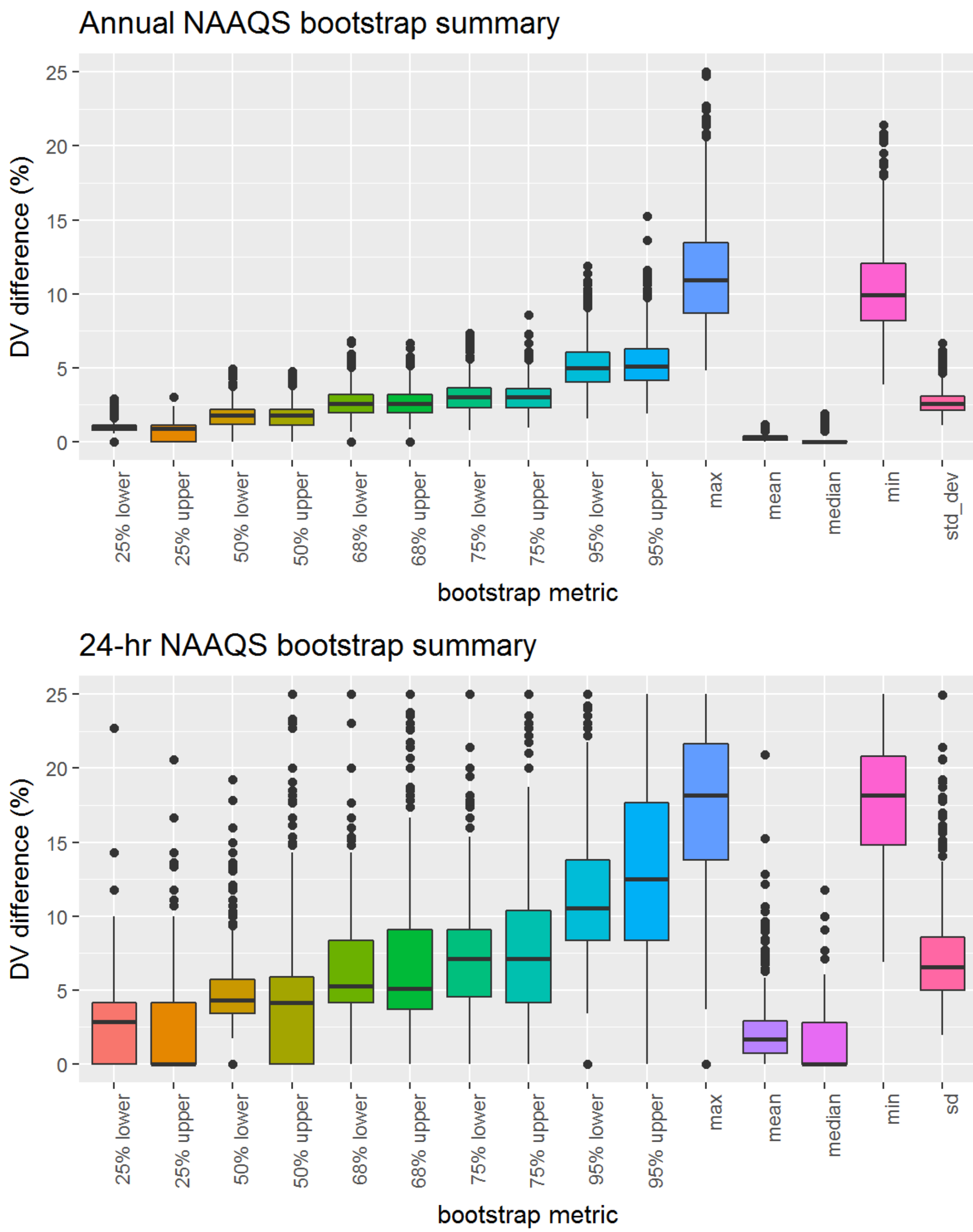


Figure 51: Bootstrap results for the 2011 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

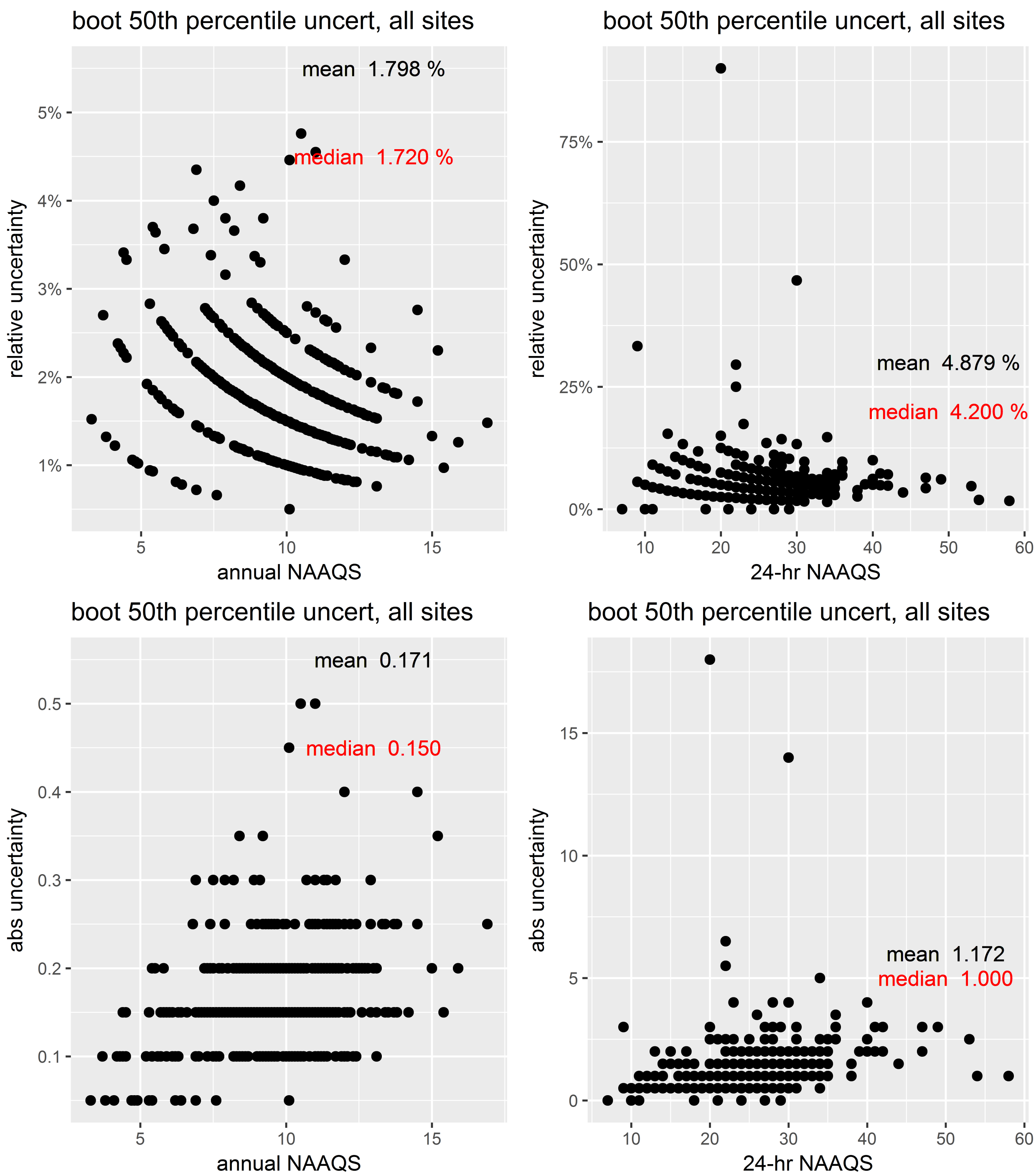


Figure 52: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

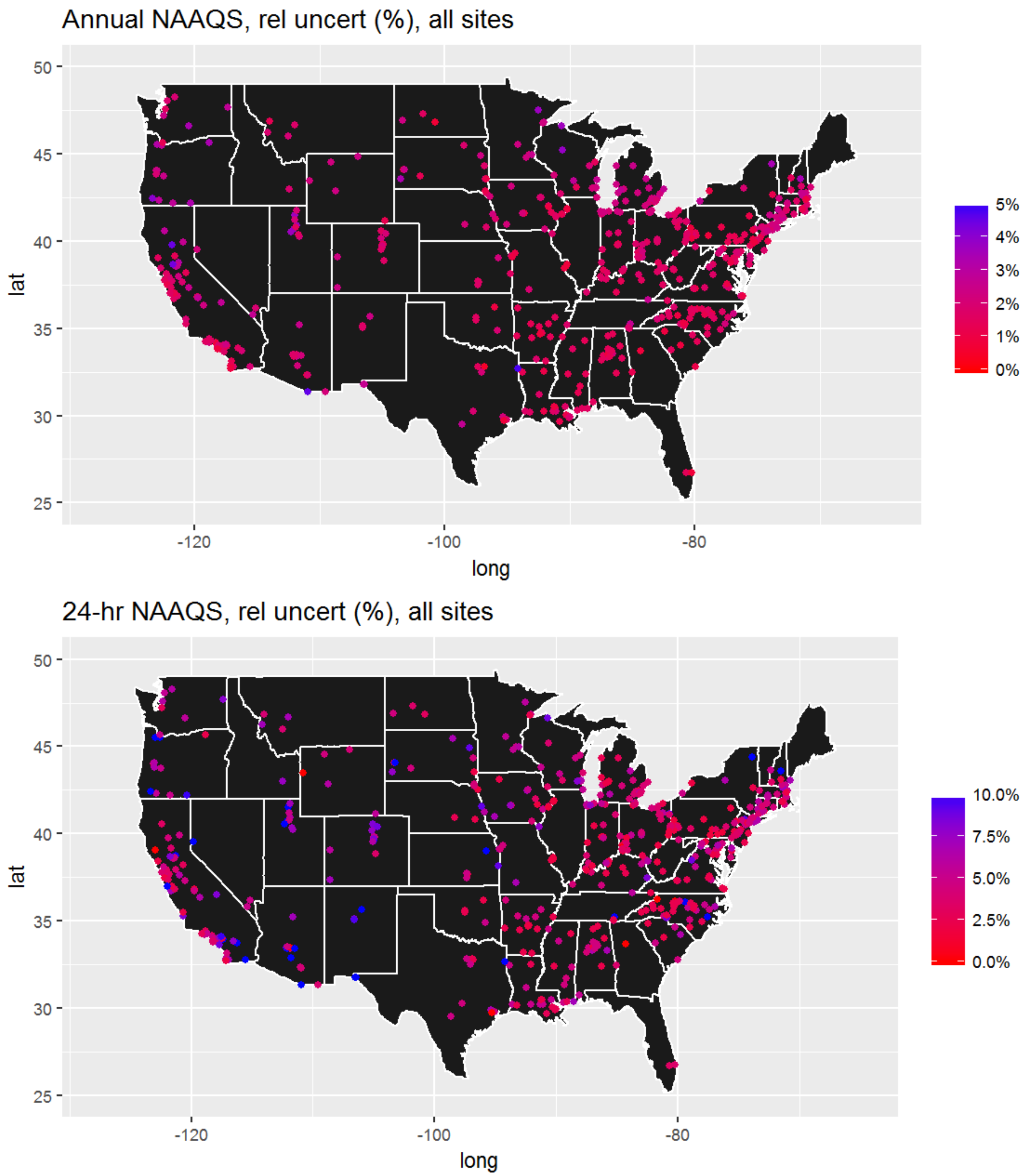


Figure 53: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2011 $PM_{2.5}$ DVs.

3.6 2008-2010 $PM_{2.5}$ bootstrap results

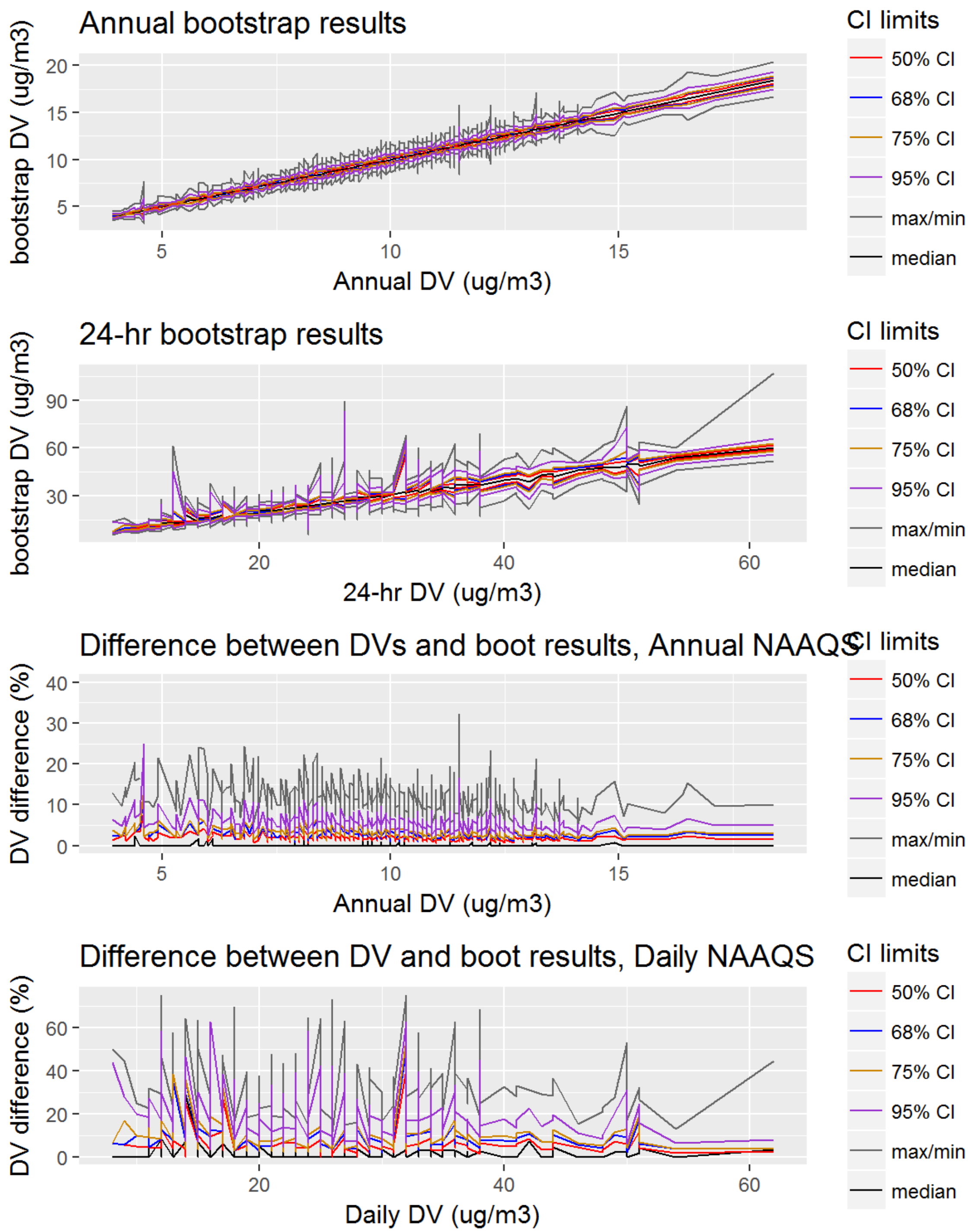


Figure 54: Bootstrap results for the 2010 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

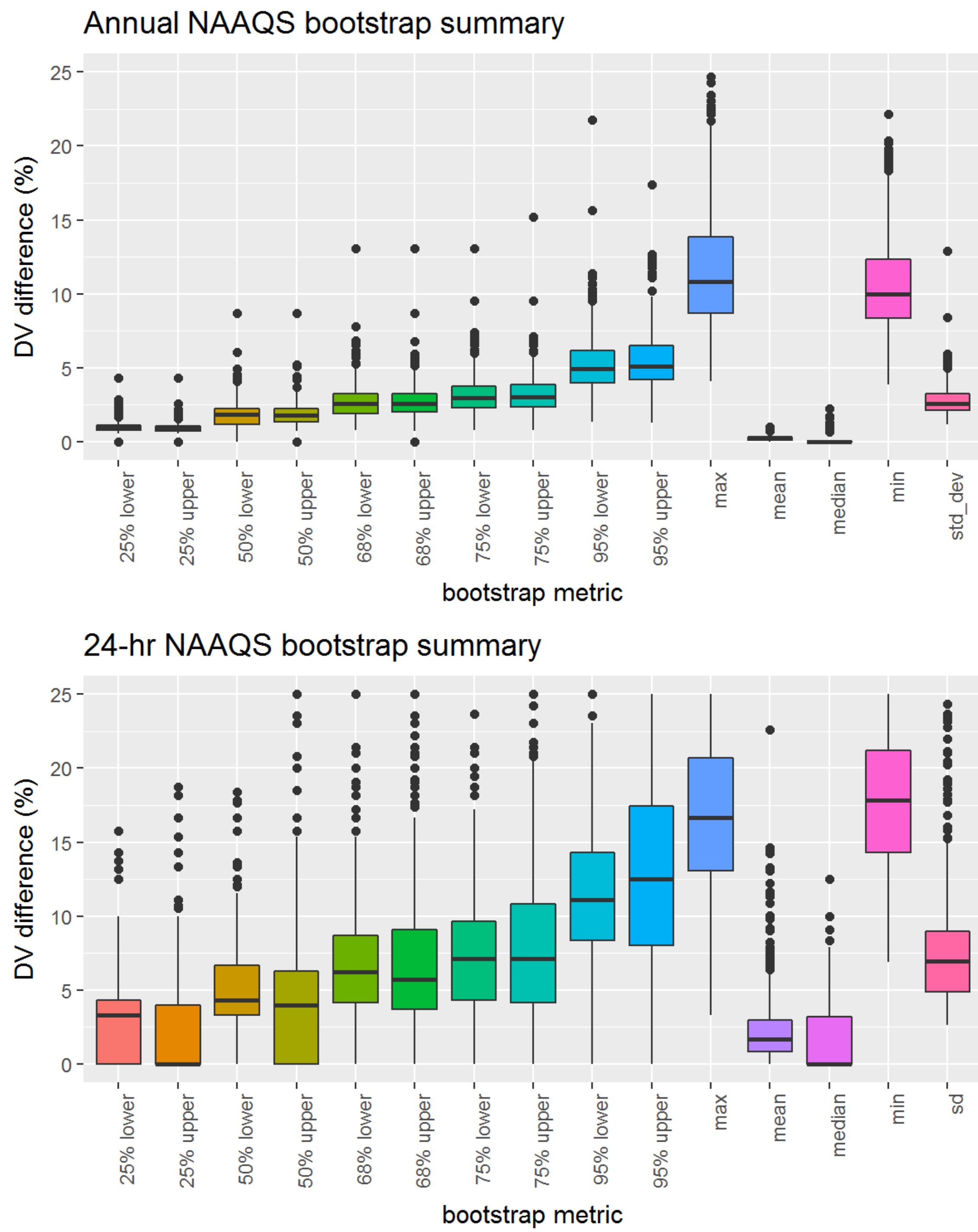


Figure 55: Bootstrap results for the 2010 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

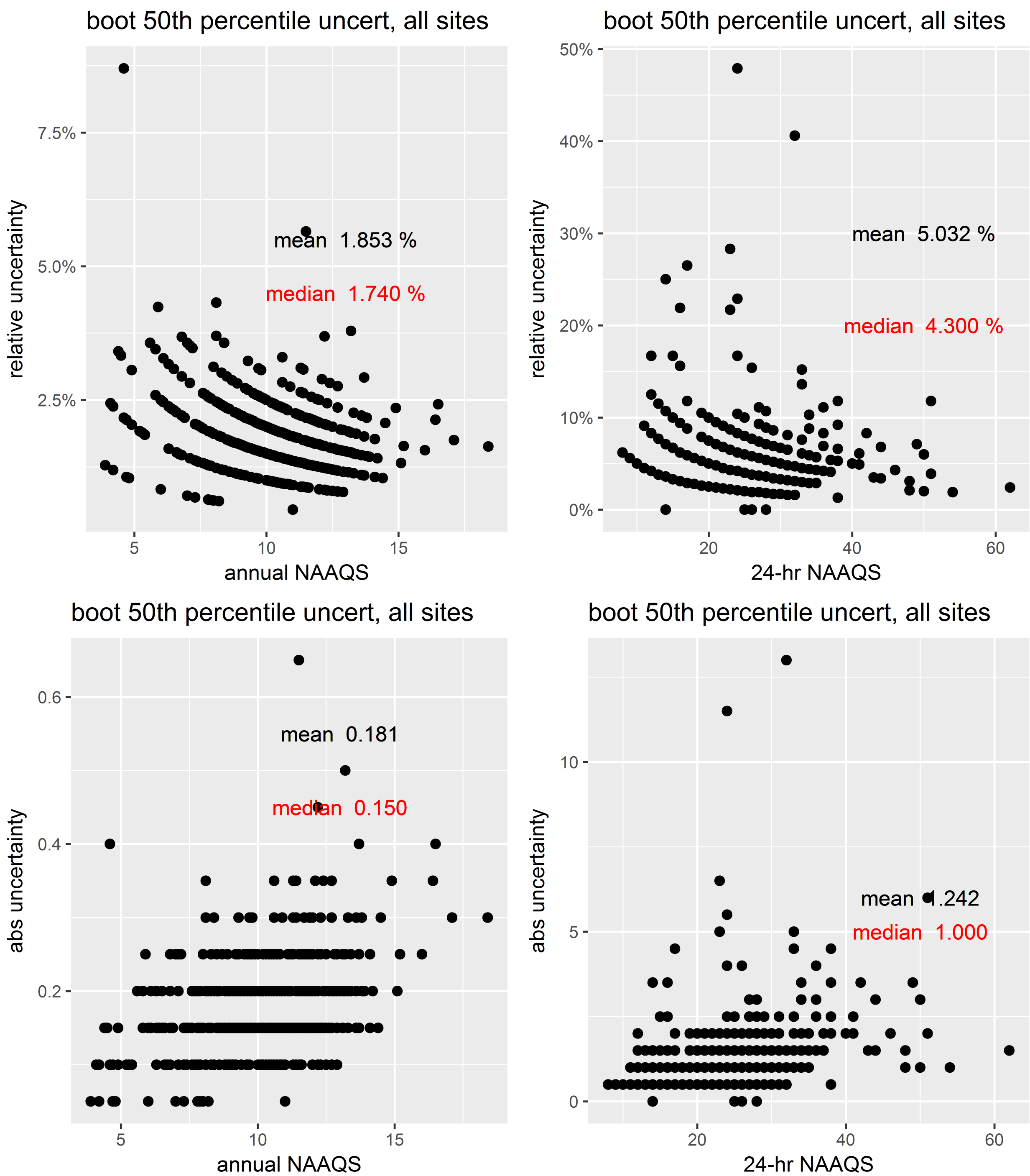


Figure 56: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

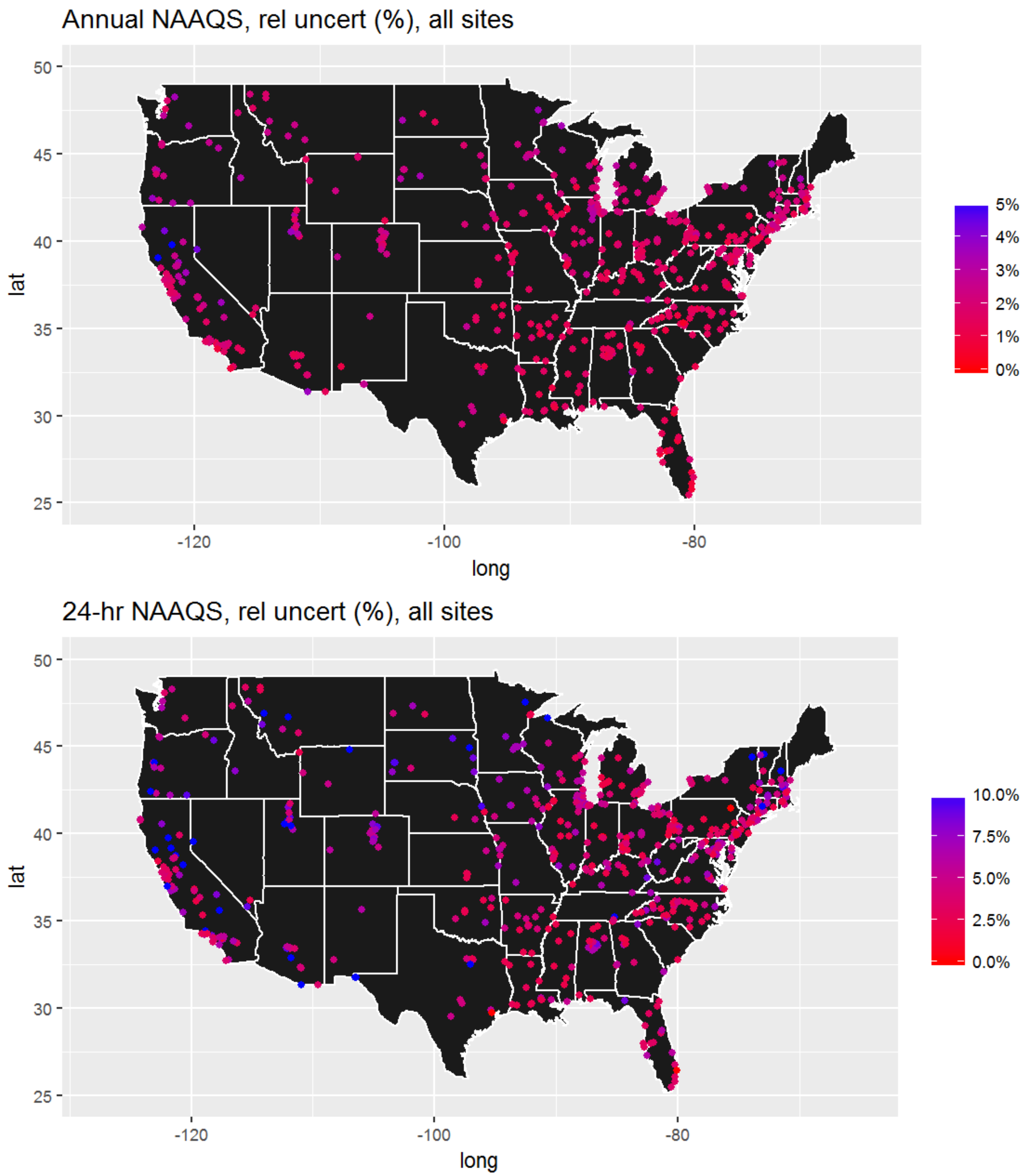


Figure 57: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2010 $PM_{2.5}$ DVs.

3.7 2007-2009 $PM_{2.5}$ bootstrap results

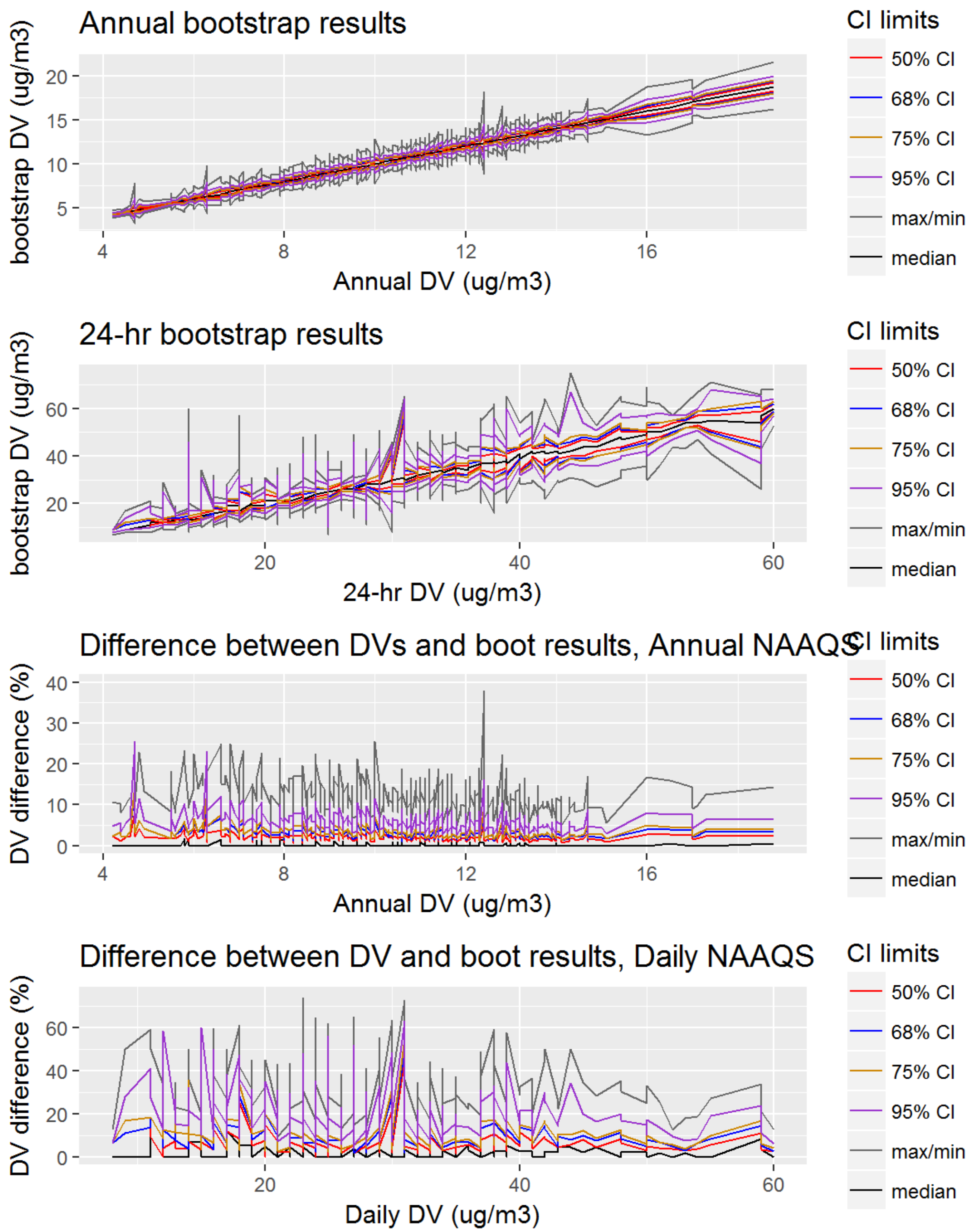


Figure 58: Bootstrap results for the 2009 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

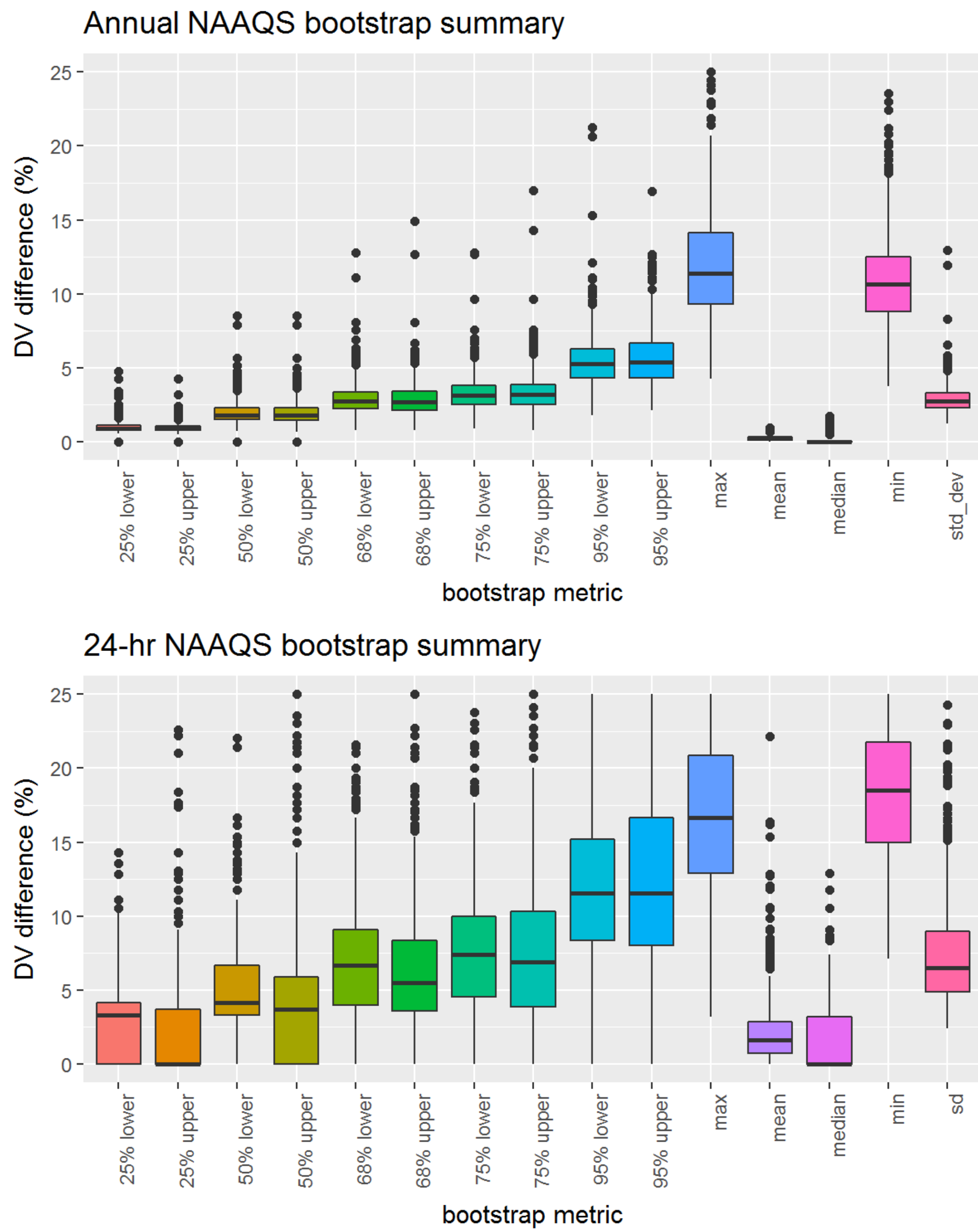


Figure 59: Bootstrap results for the 2009 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

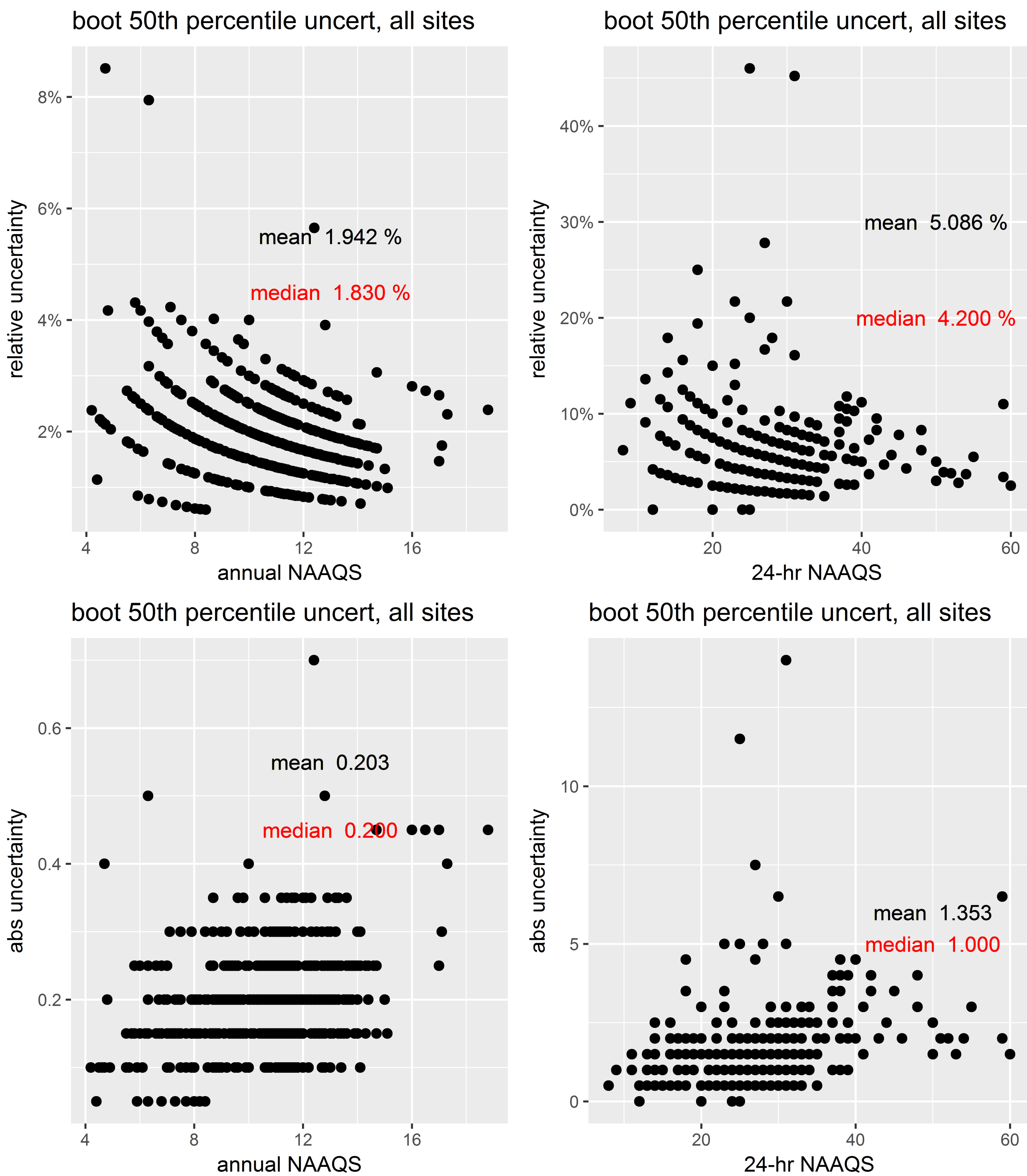


Figure 60: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

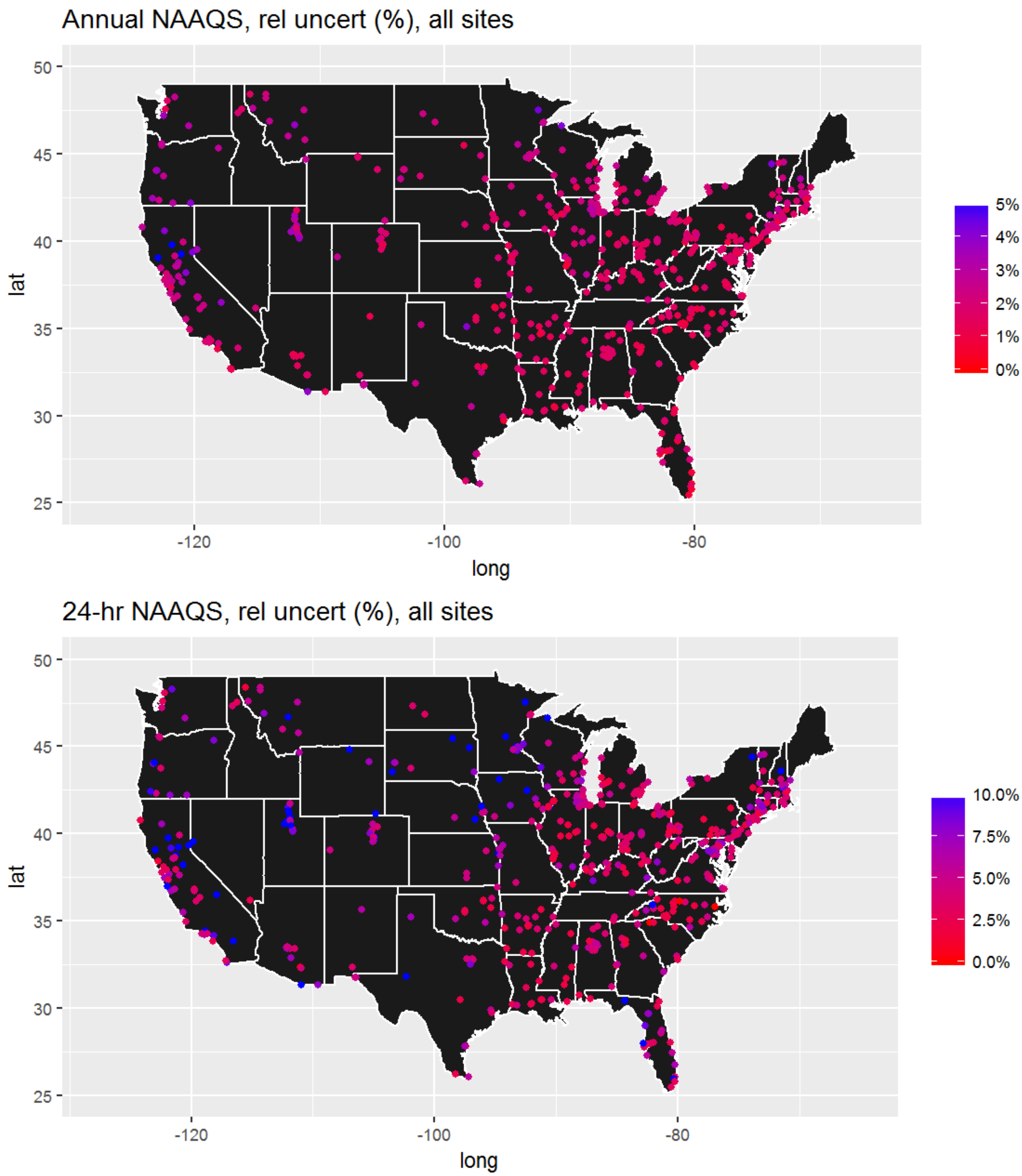


Figure 61: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2009 $PM_{2.5}$ DVs.

3.8 2006-2008 $PM_{2.5}$ bootstrap results

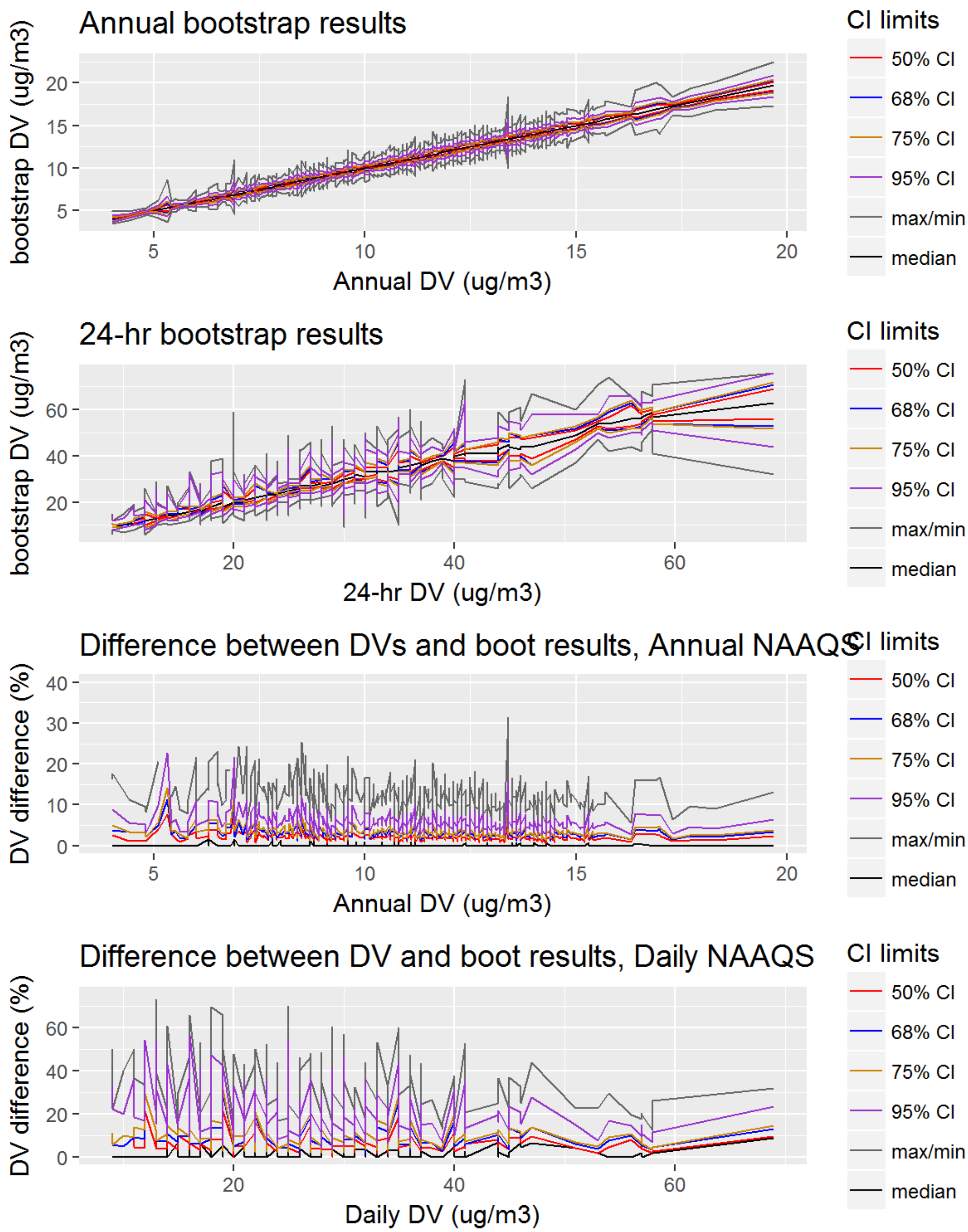


Figure 62: Bootstrap results for the 2008 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

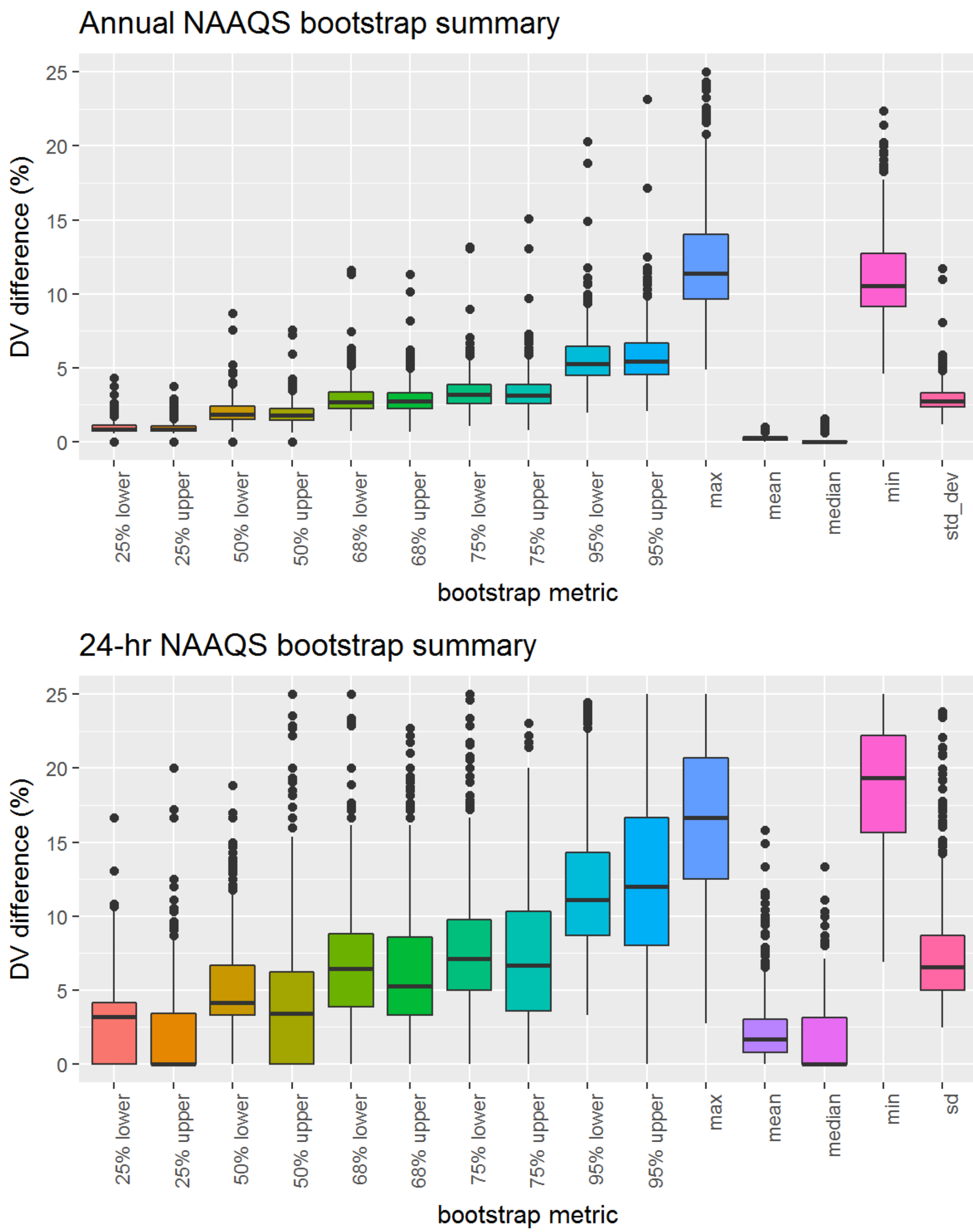


Figure 63: Bootstrap results for the 2008 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

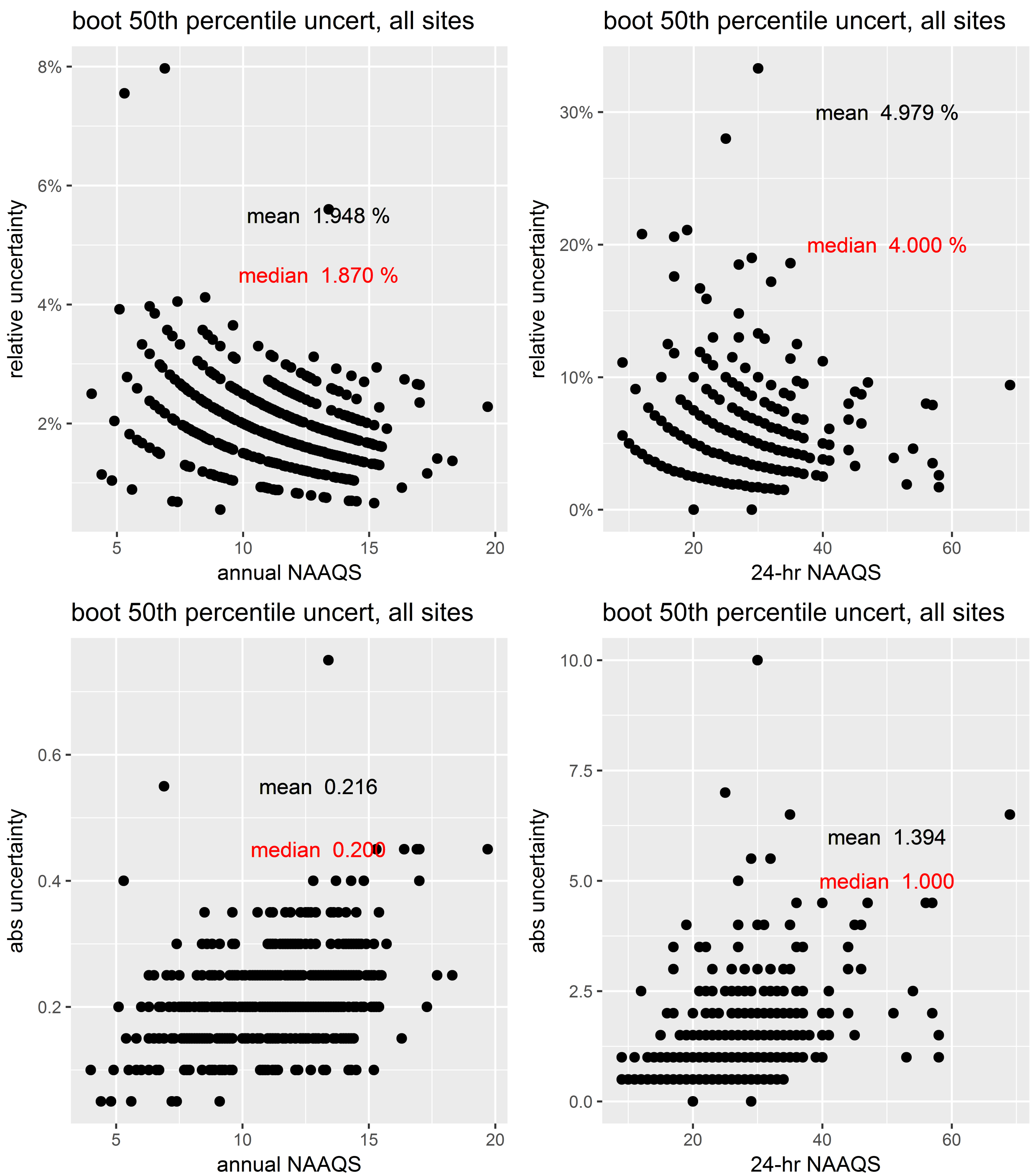


Figure 64: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

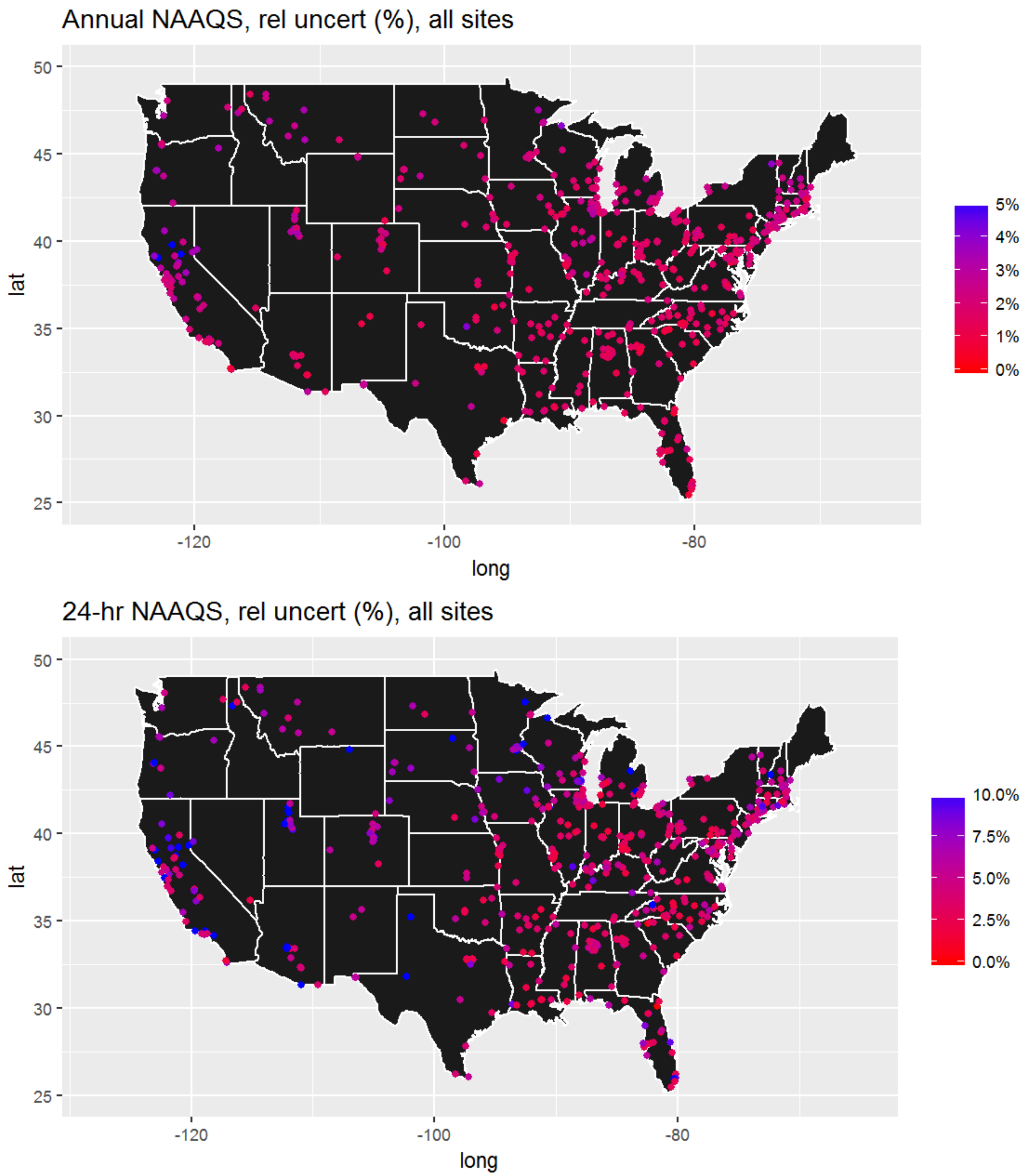


Figure 65: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2008 $PM_{2.5}$ DVs.

3.9 2005-2007 $PM_{2.5}$ bootstrap results

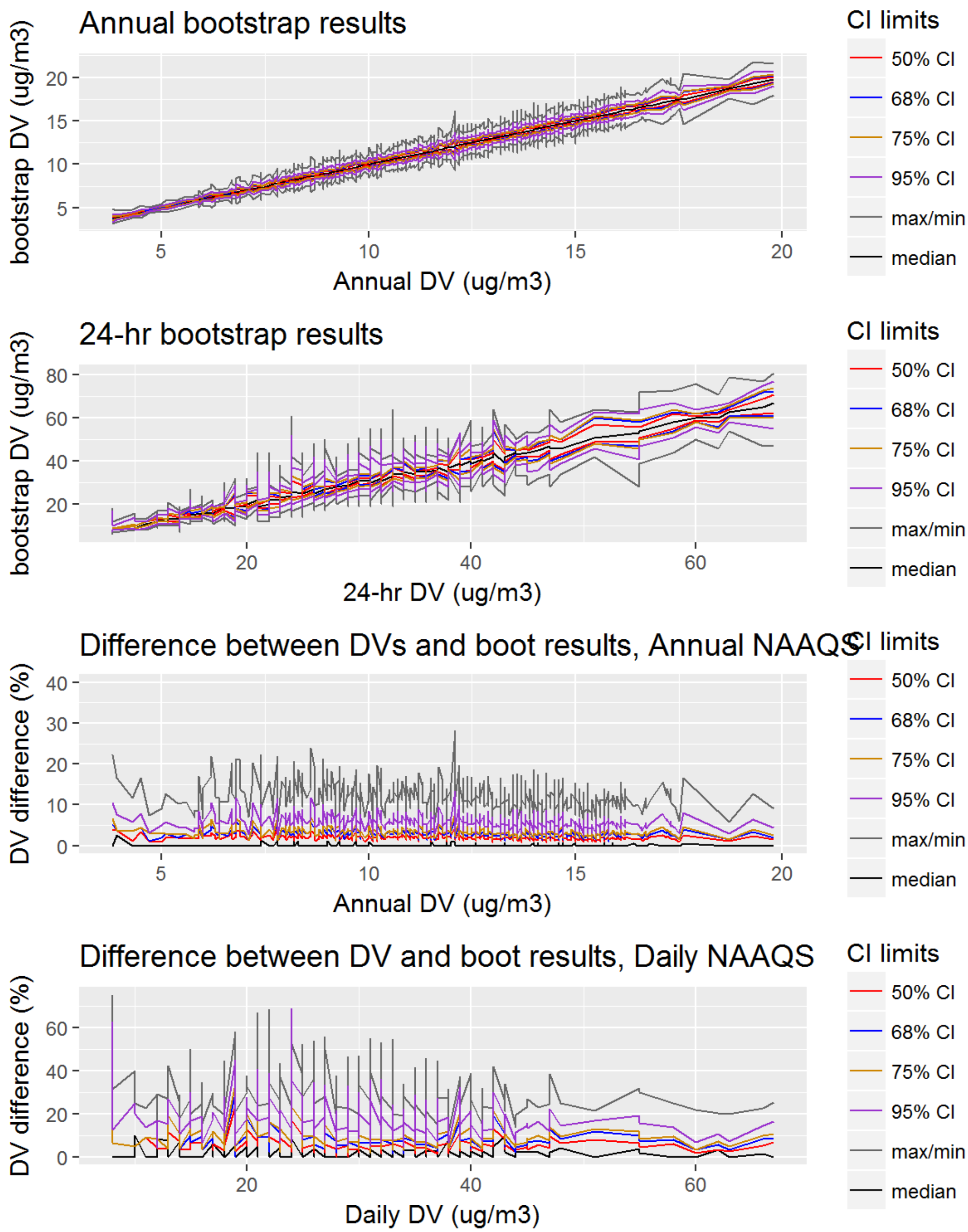
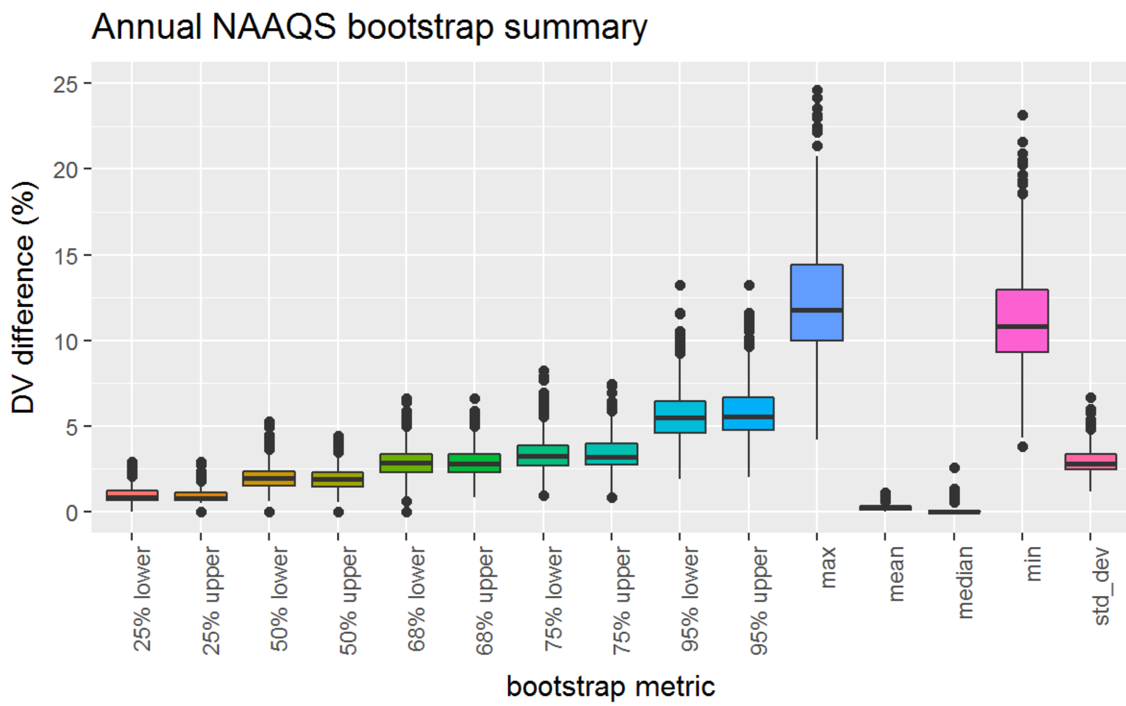


Figure 66: Bootstrap results for the 2007 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.



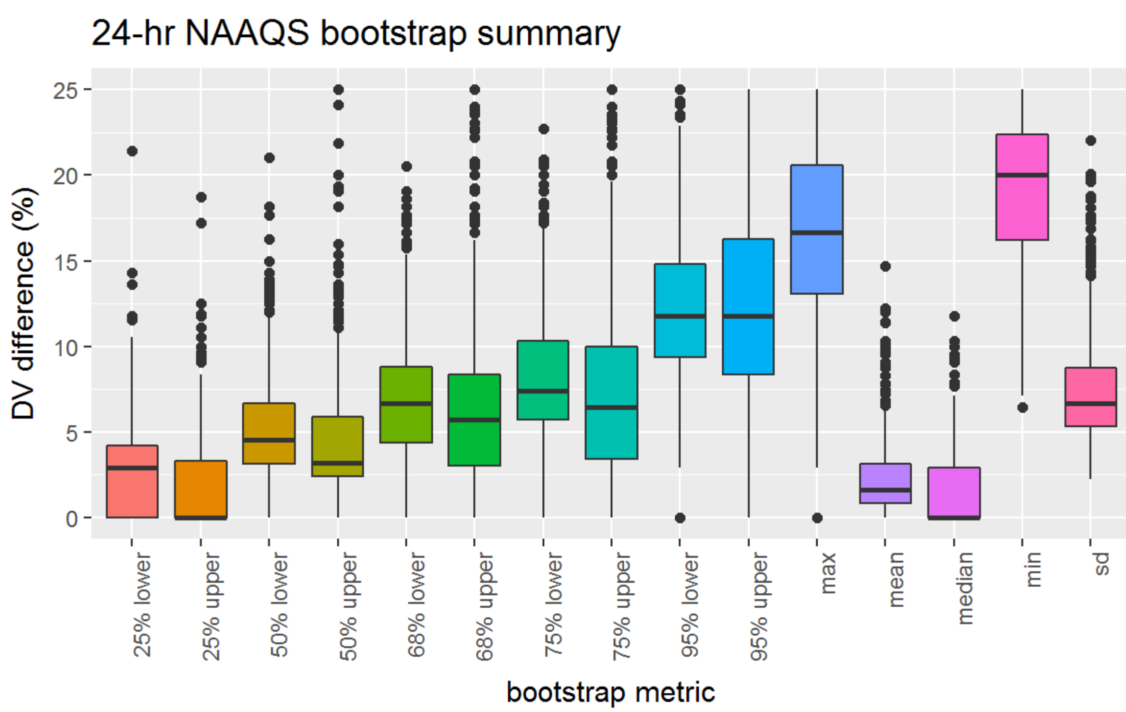


Figure 67: Bootstrap results for the 2007 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

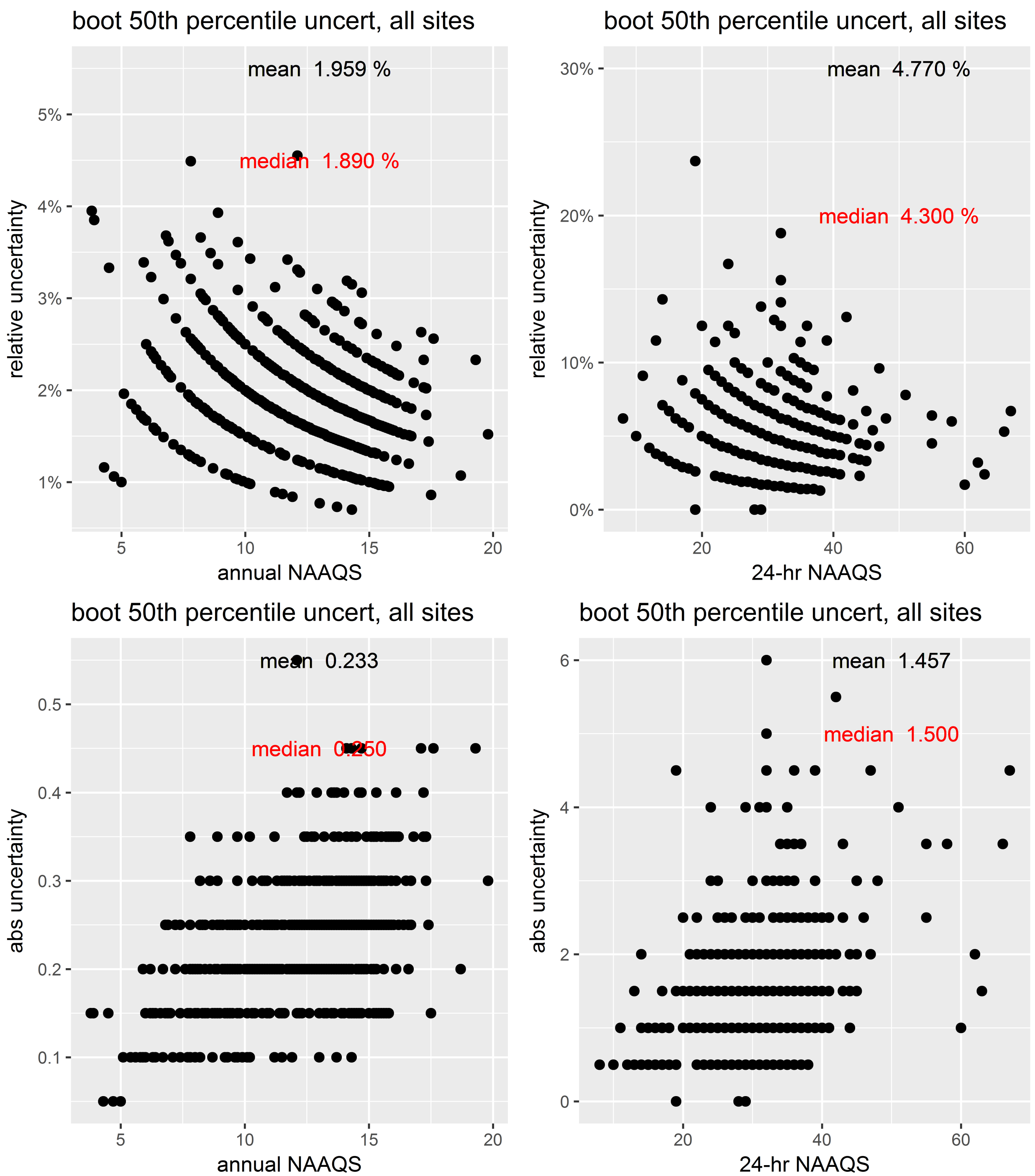


Figure 68: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

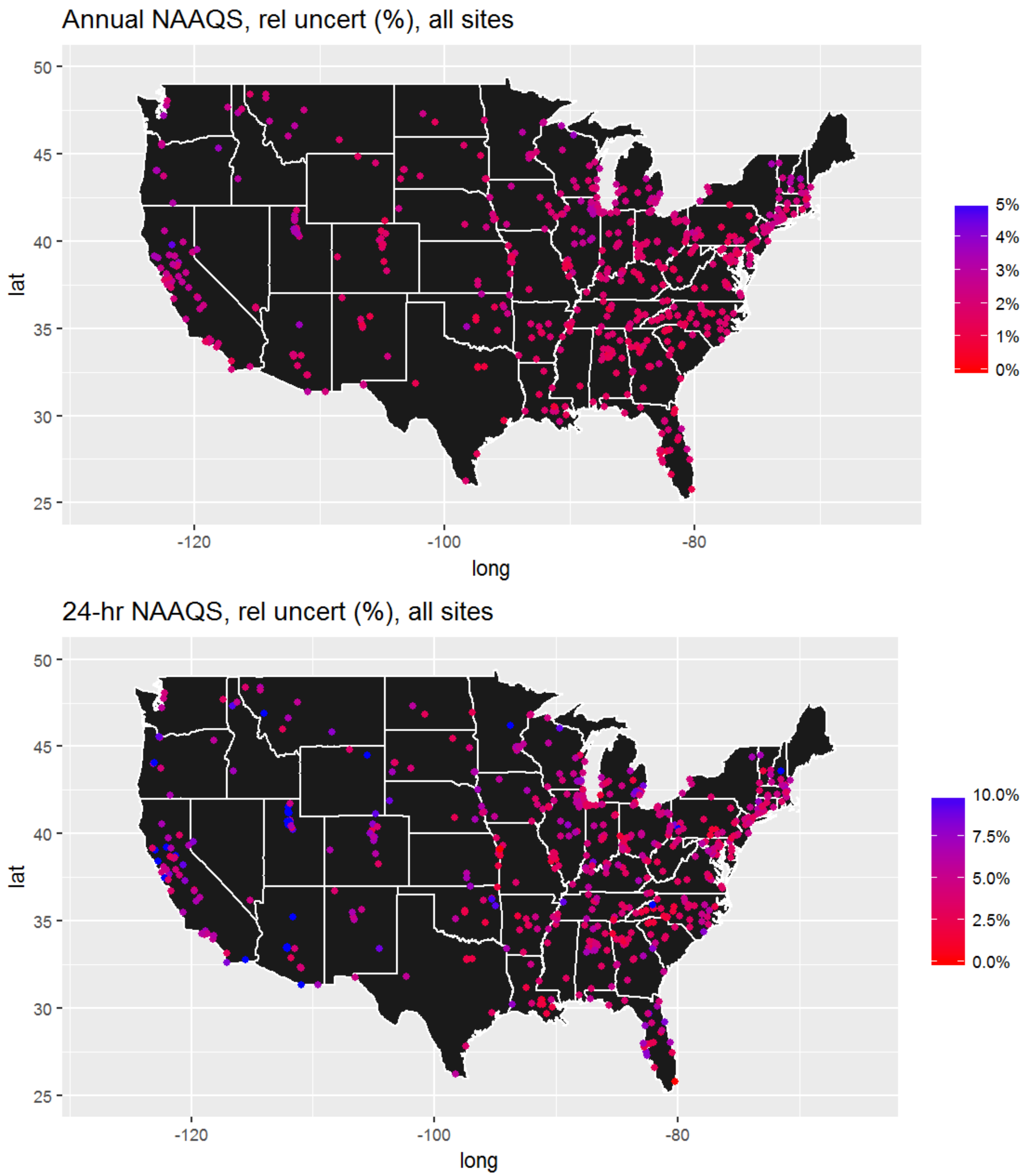


Figure 69: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2007 $PM_{2.5}$ DVs.

3.10 2004-2006 $PM_{2.5}$ bootstrap results

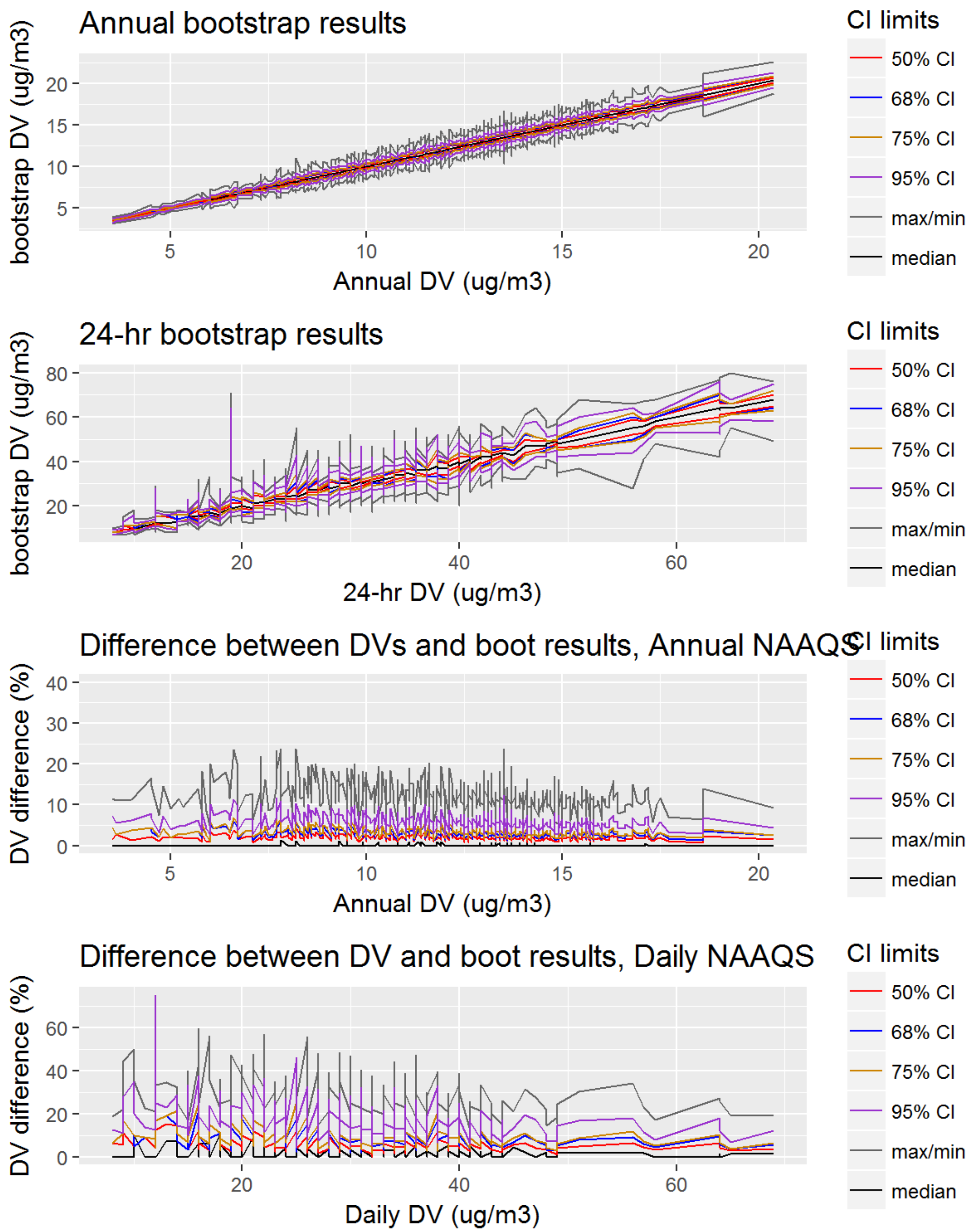


Figure 70: Bootstrap results for the 2006 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

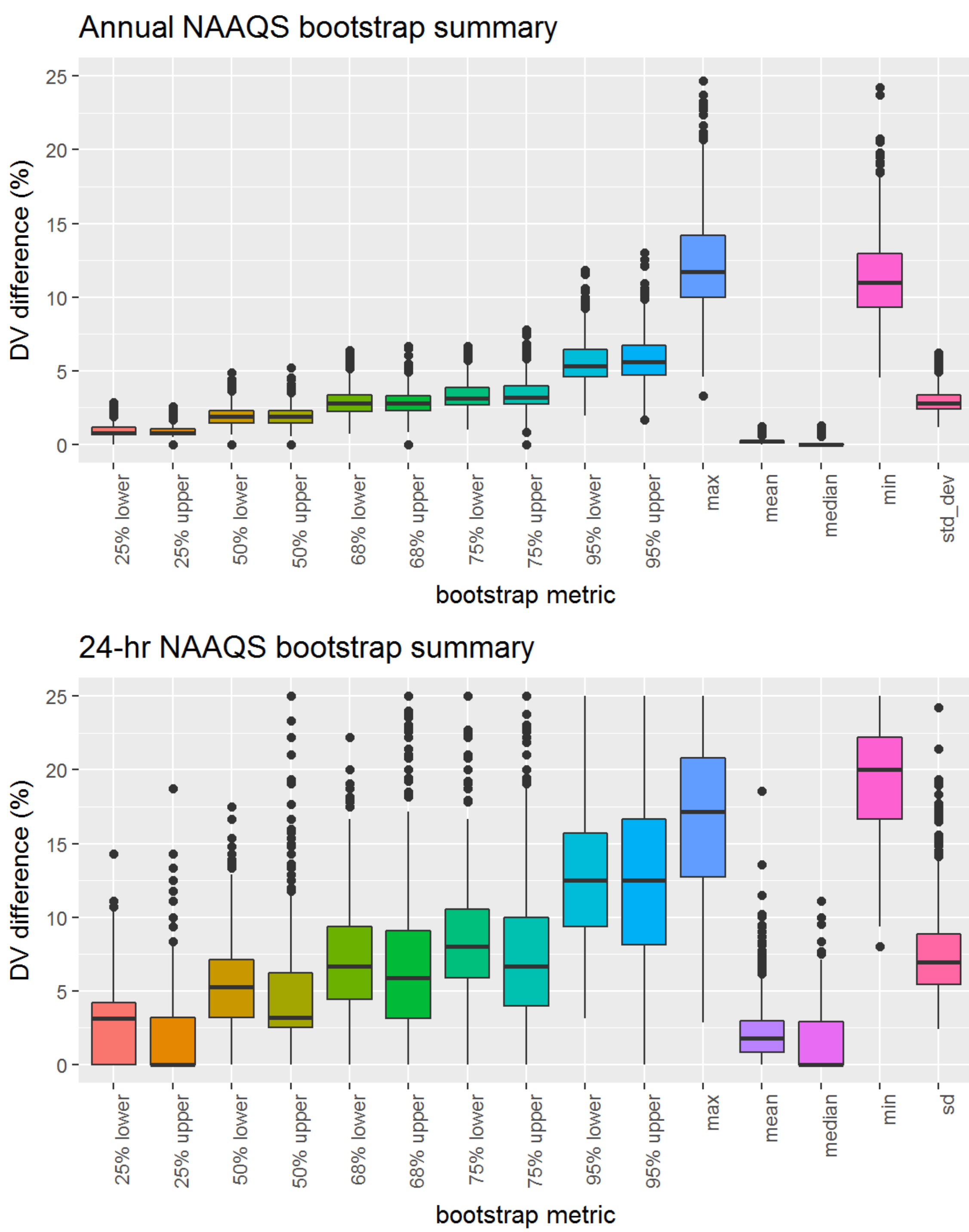


Figure 71: Bootstrap results for the 2006 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

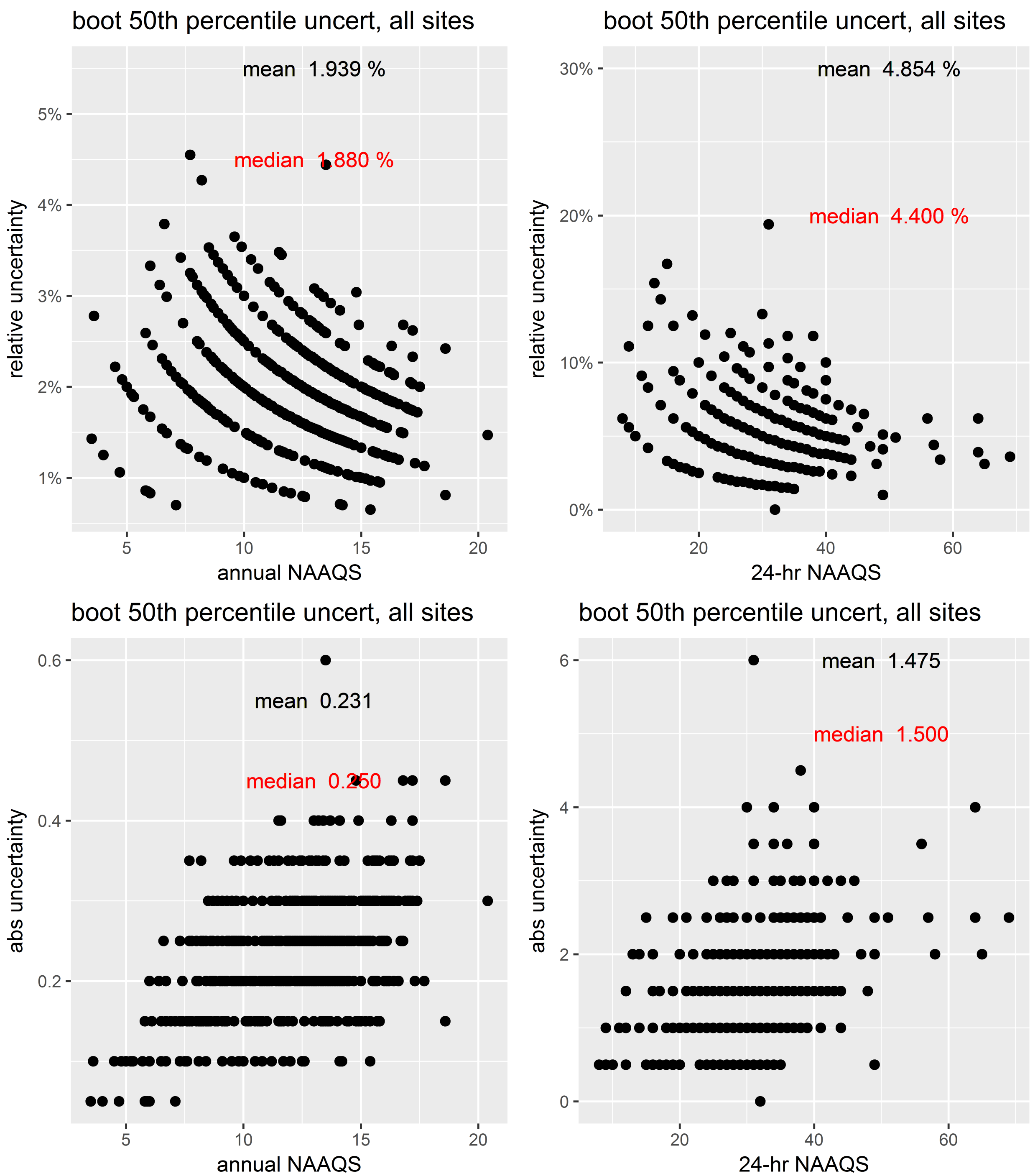


Figure 72: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

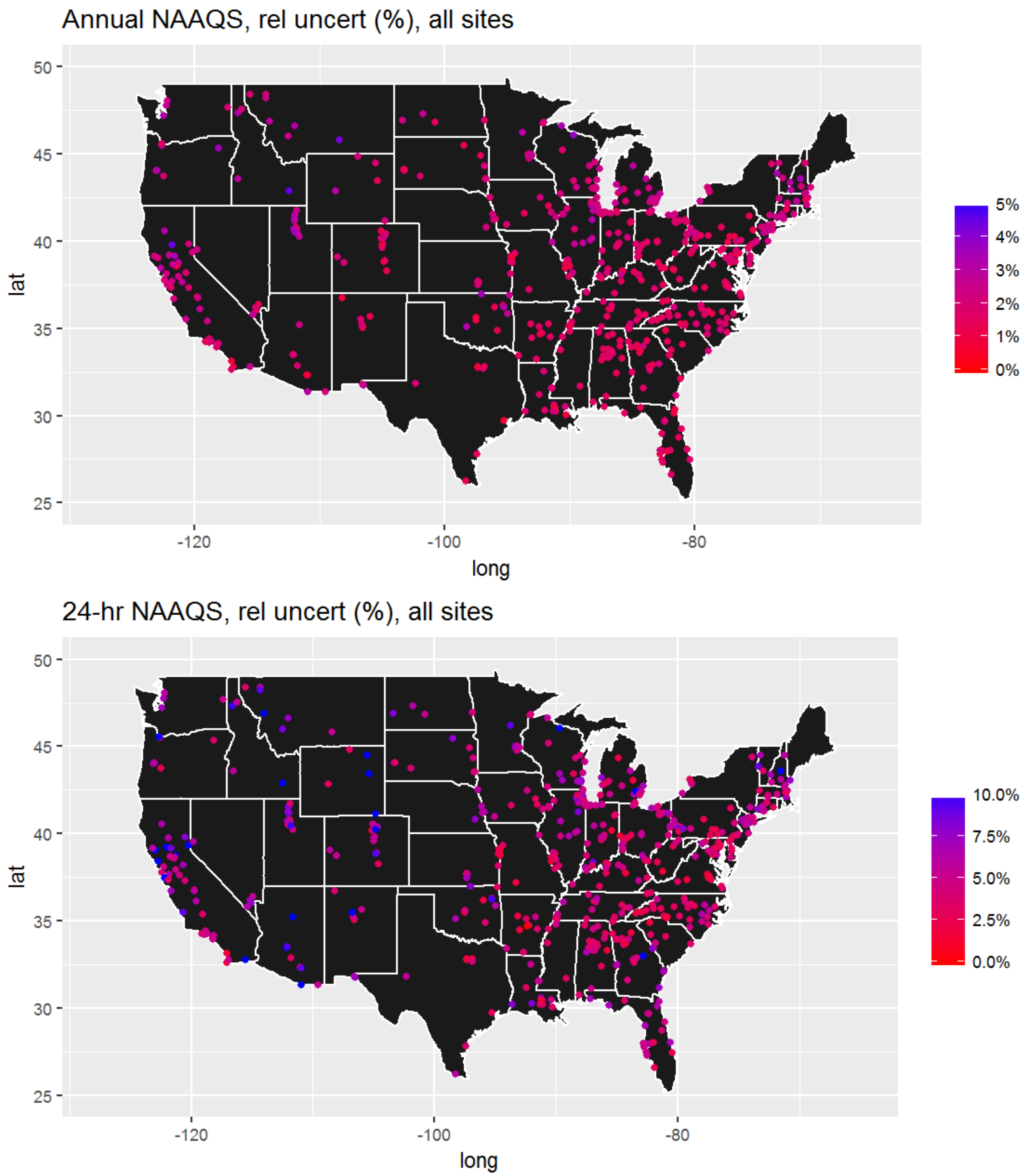


Figure 73: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2006 $PM_{2.5}$ DVs.

3.11 2003-2005 $PM_{2.5}$ bootstrap results

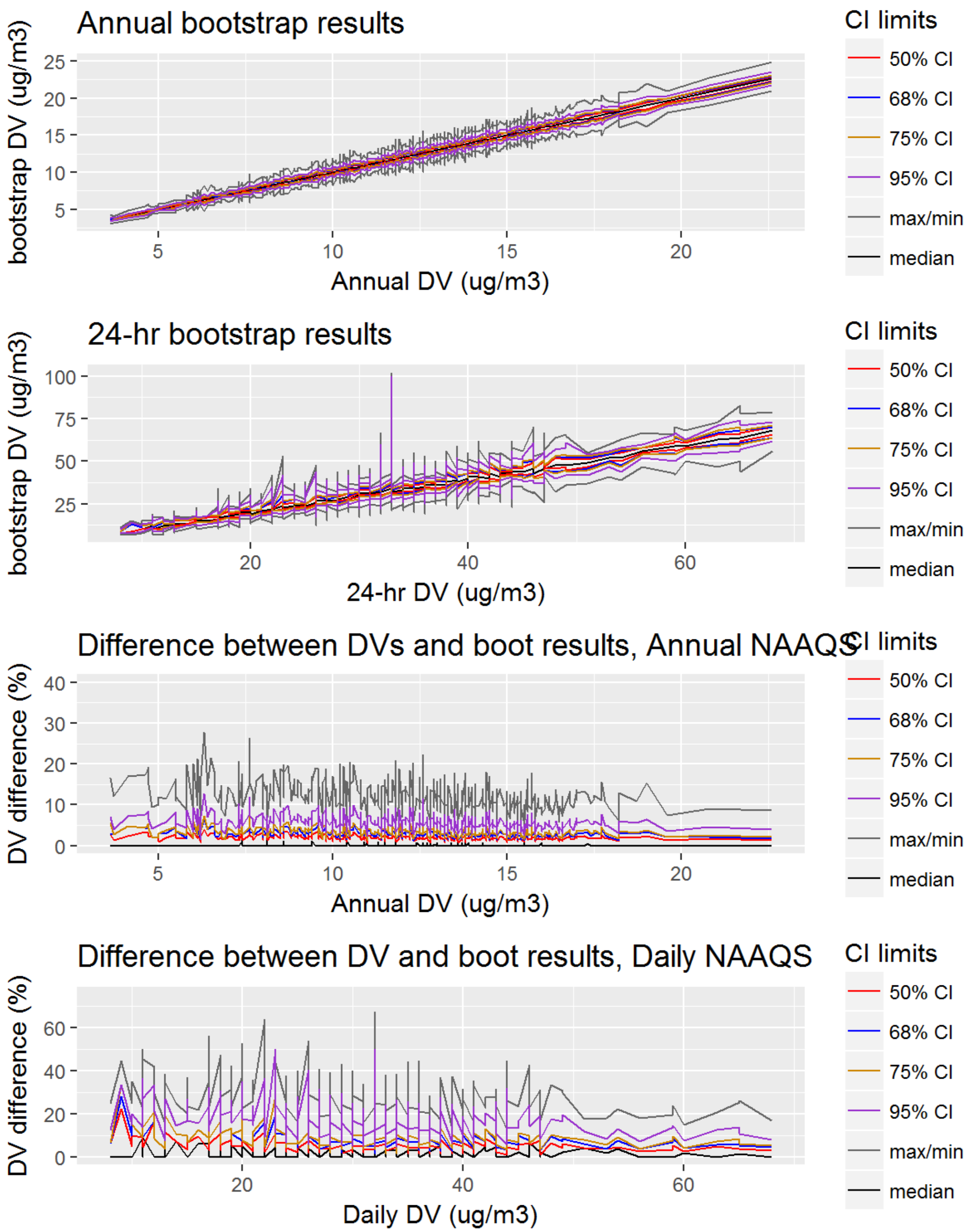
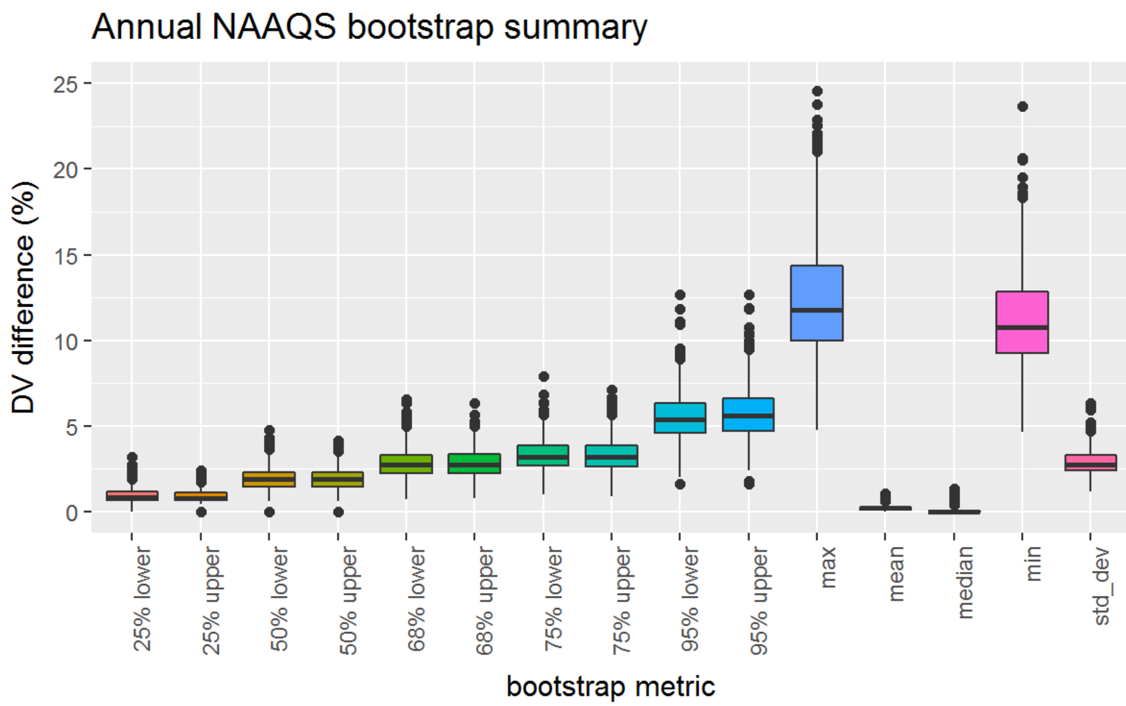


Figure 74: Bootstrap results for the 2005 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.



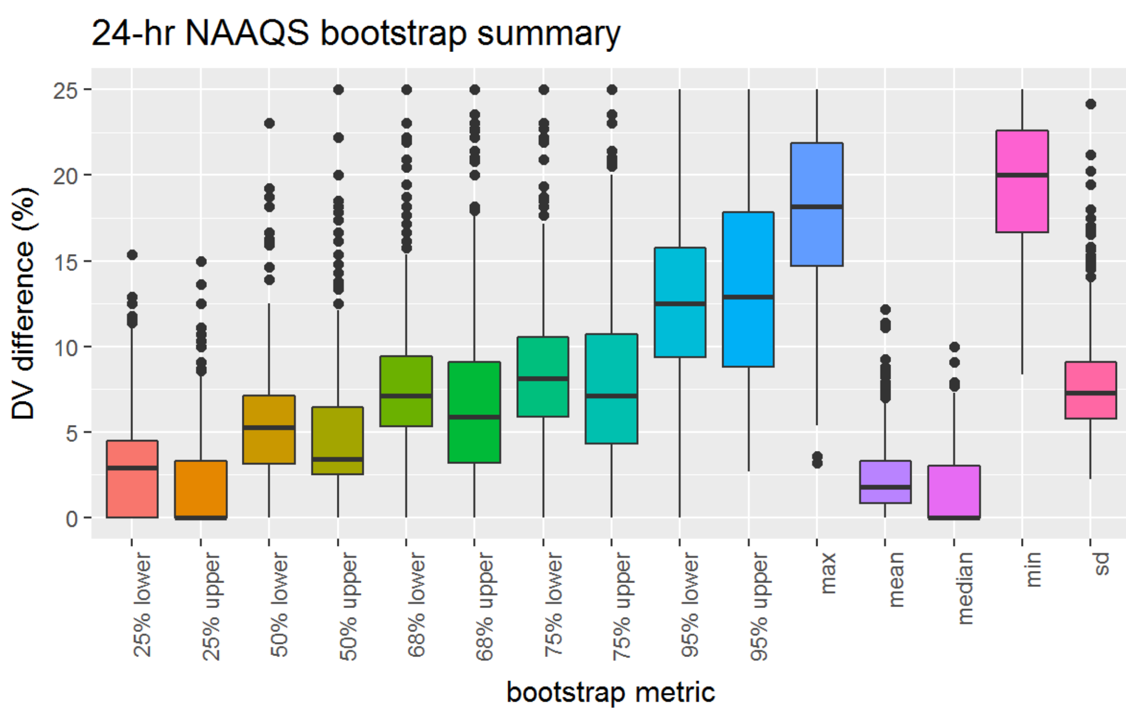


Figure 75: Bootstrap results for the 2005 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

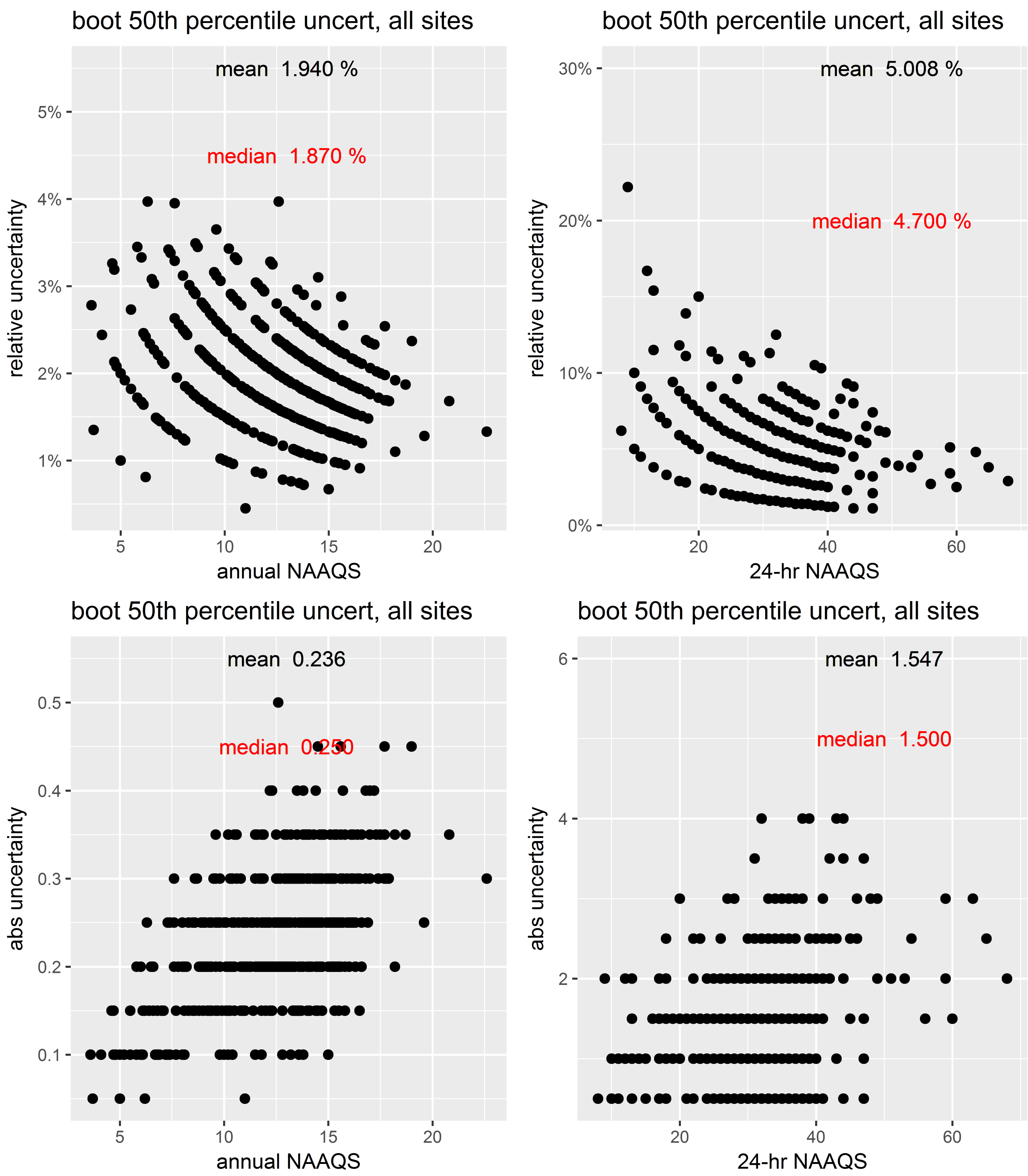
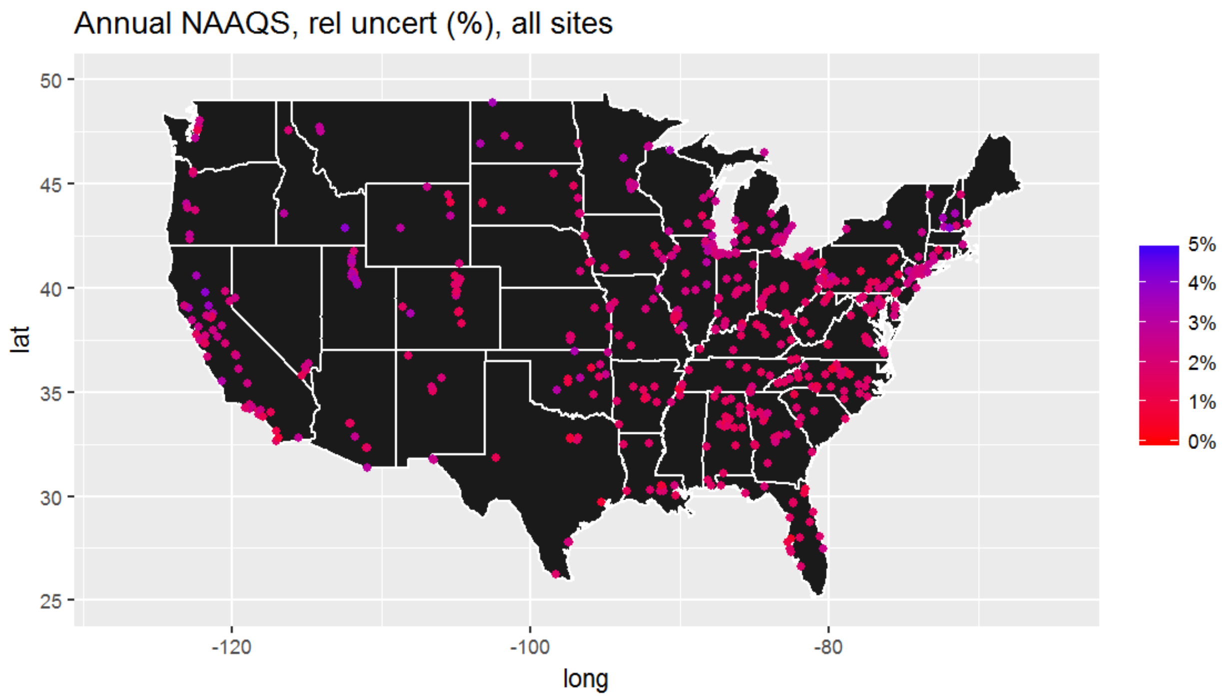


Figure 76: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.



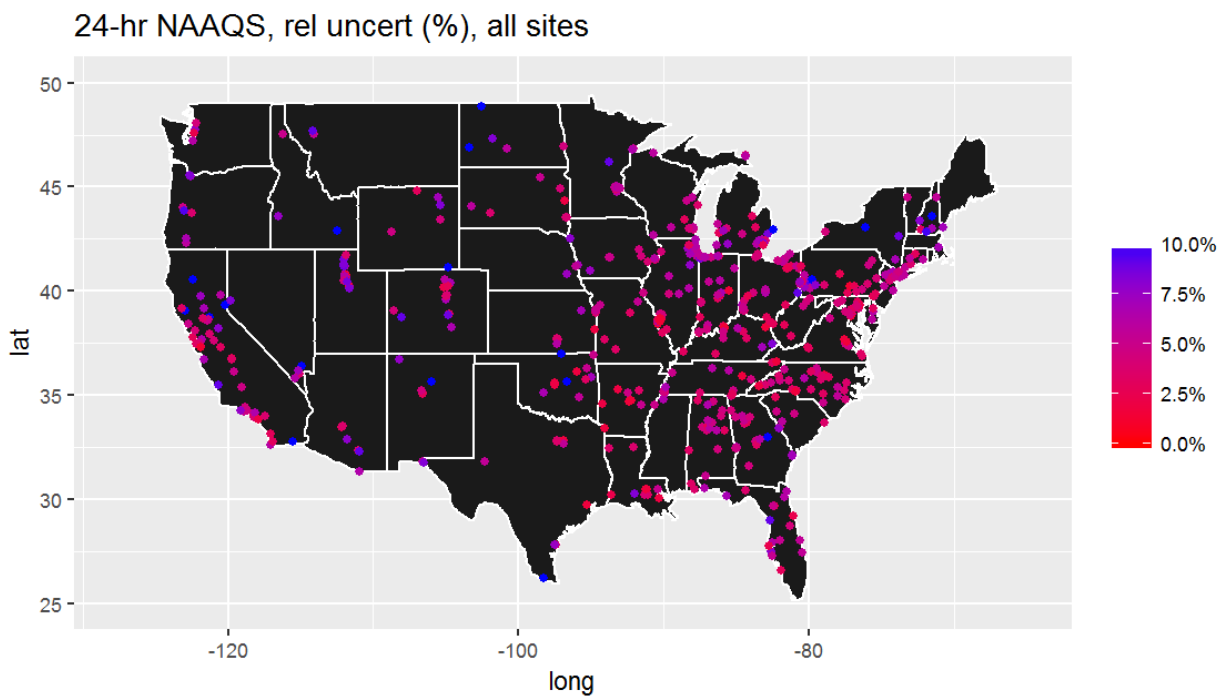


Figure 77: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2005 $PM_{2.5}$ DVs.

3.12 2002-2004 $PM_{2.5}$ bootstrap results

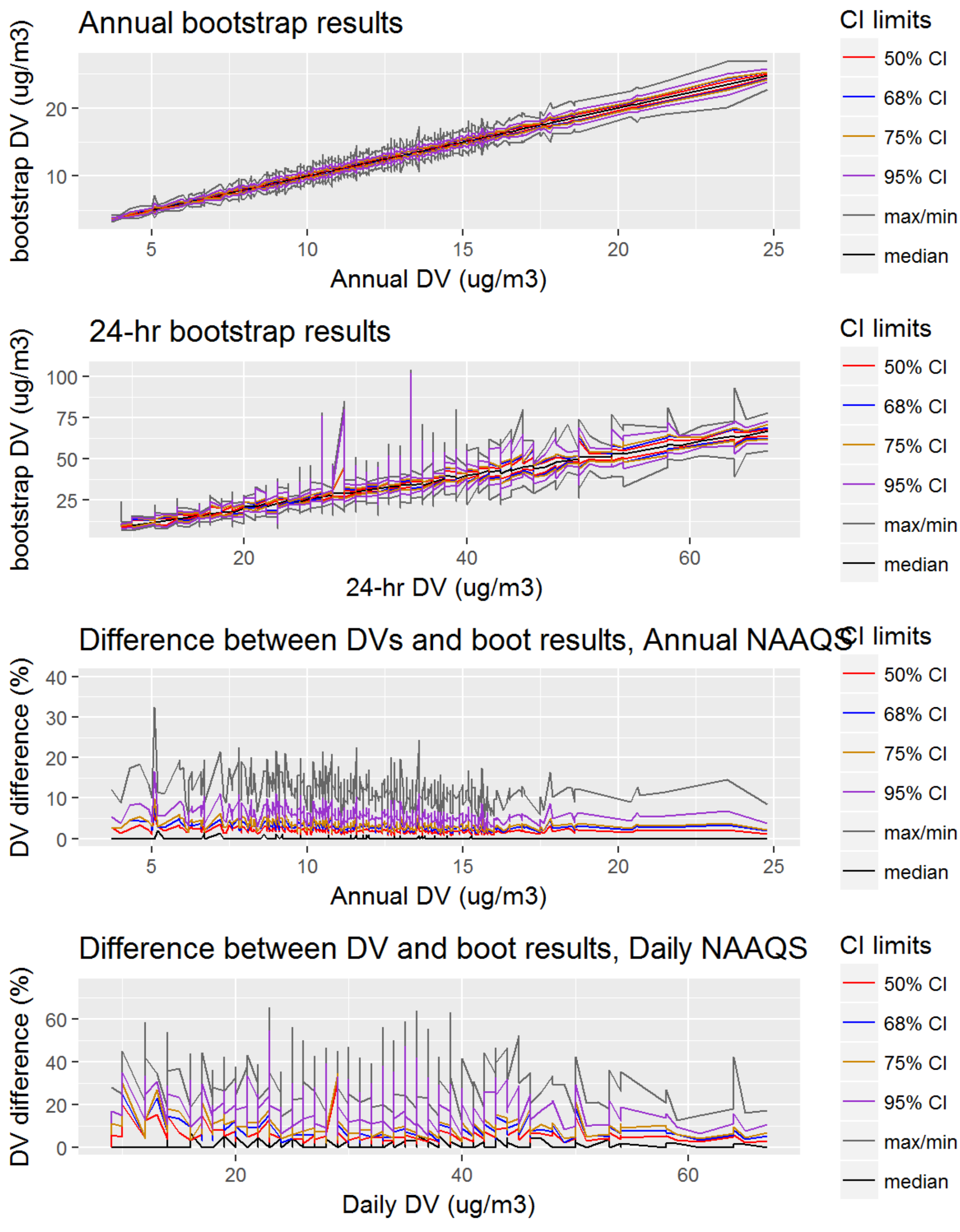


Figure 78: Bootstrap results for the 2004 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

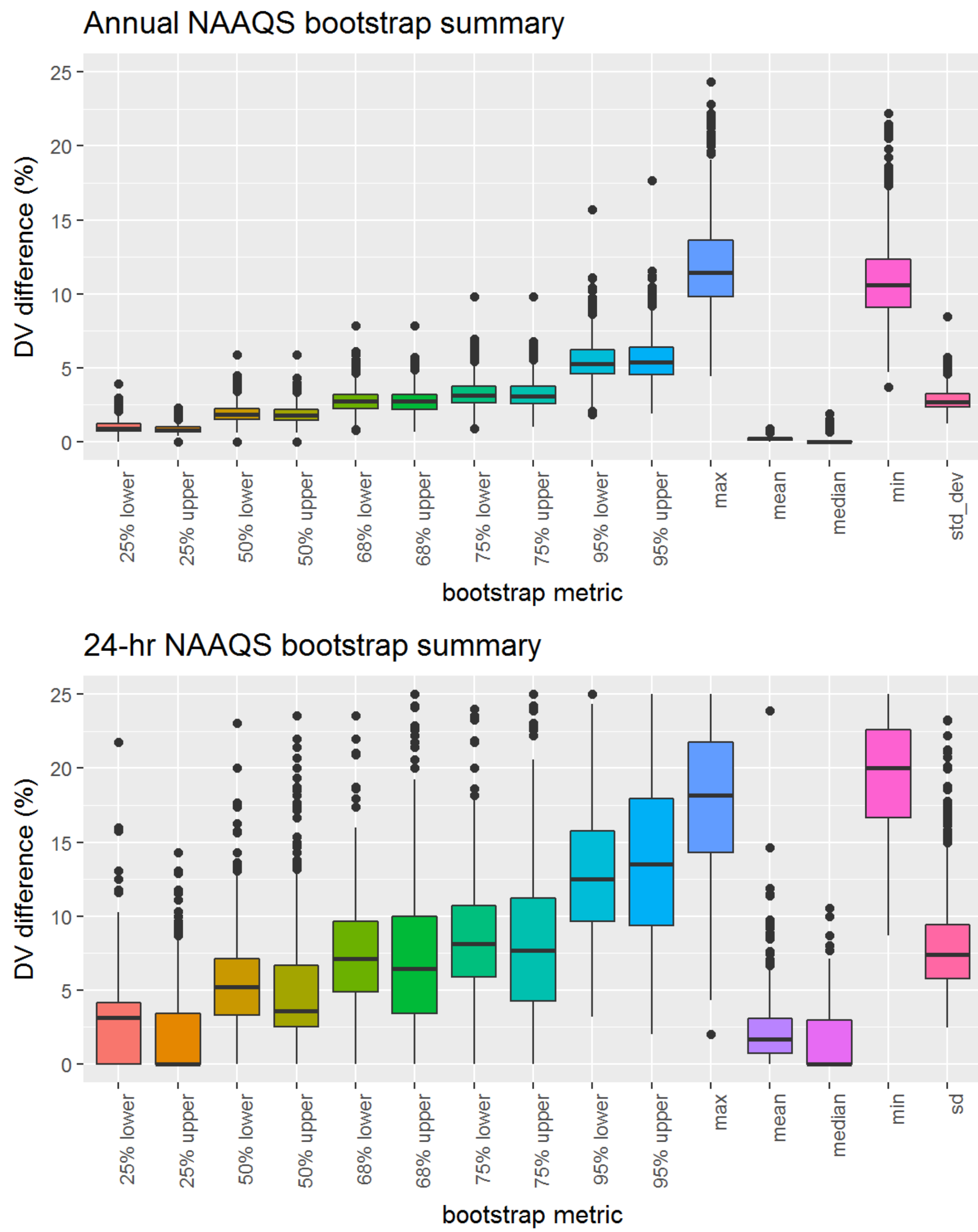


Figure 79: Bootstrap results for the 2004 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

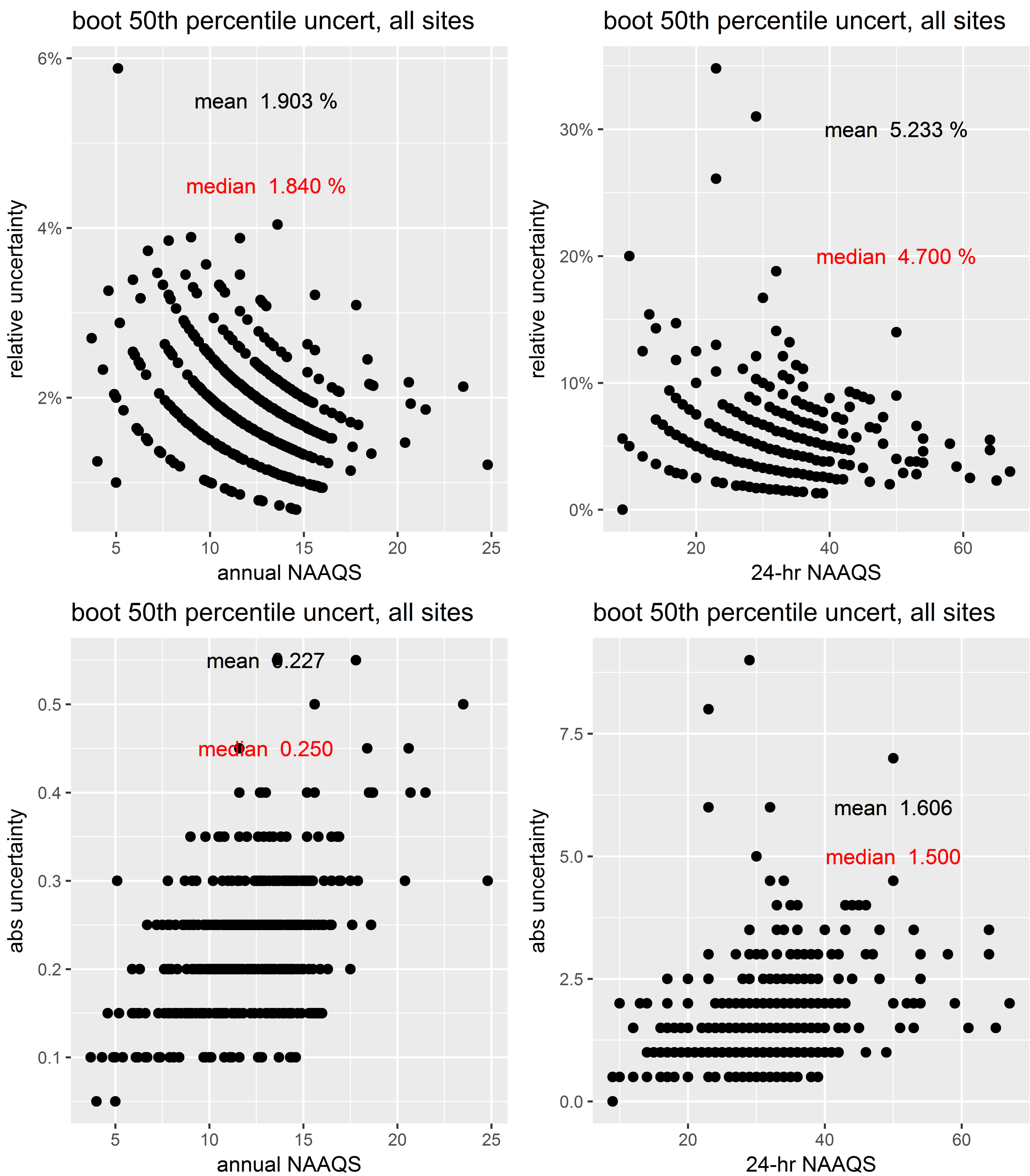


Figure 80: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

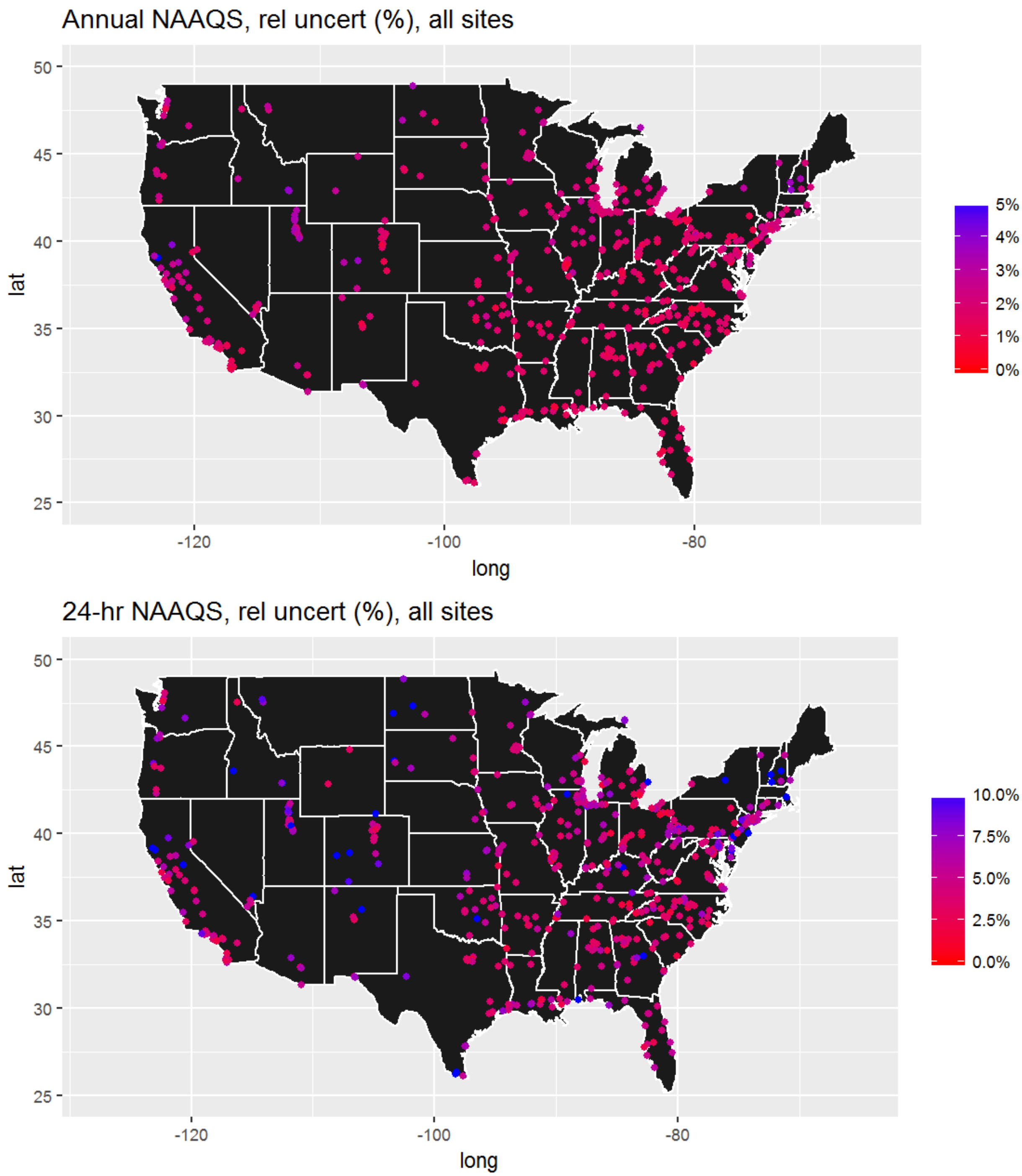


Figure 81: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2004 $PM_{2.5}$ DVs.

3.13 2001-2003 $PM_{2.5}$ bootstrap results

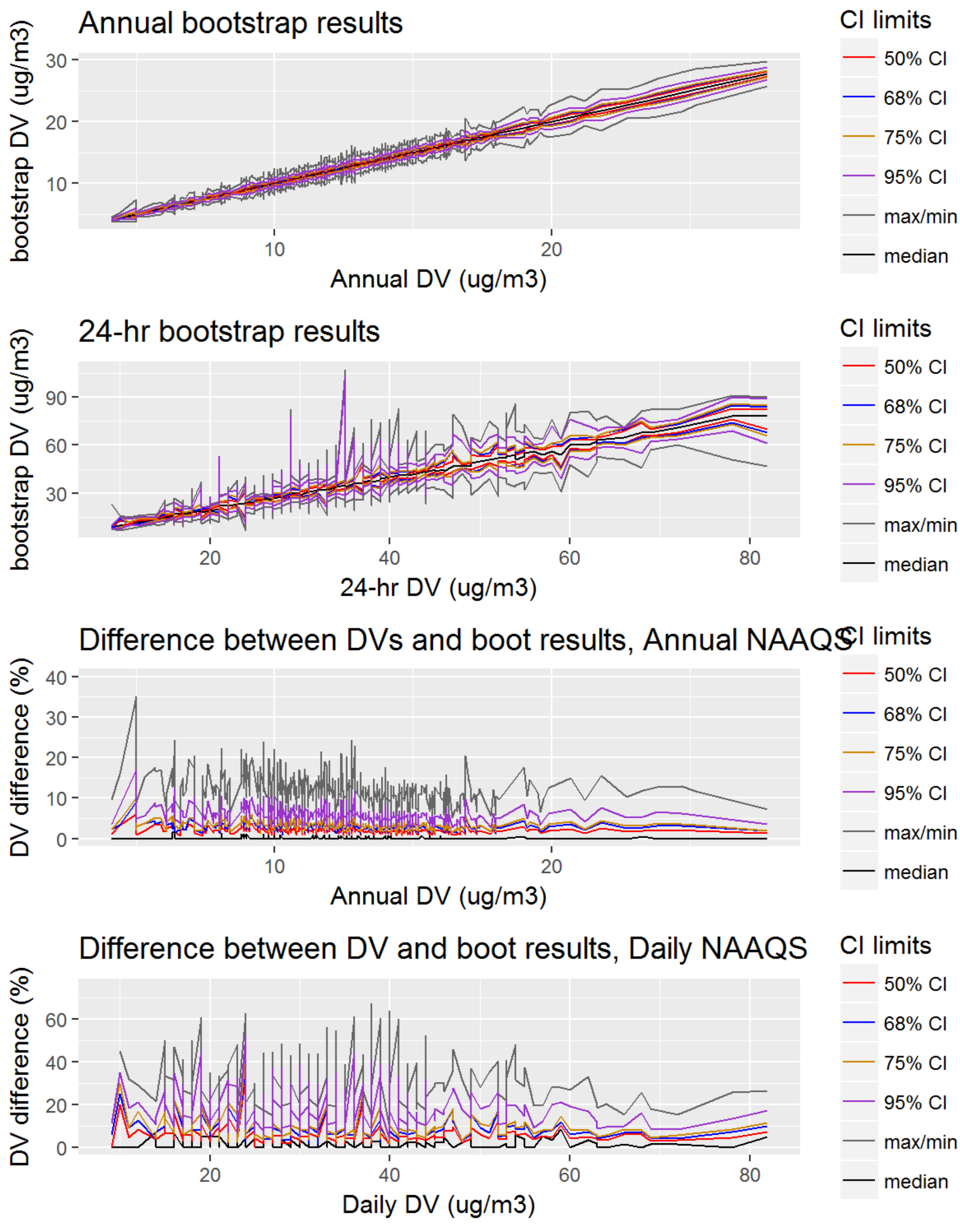


Figure 82: Bootstrap results for the 2003 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

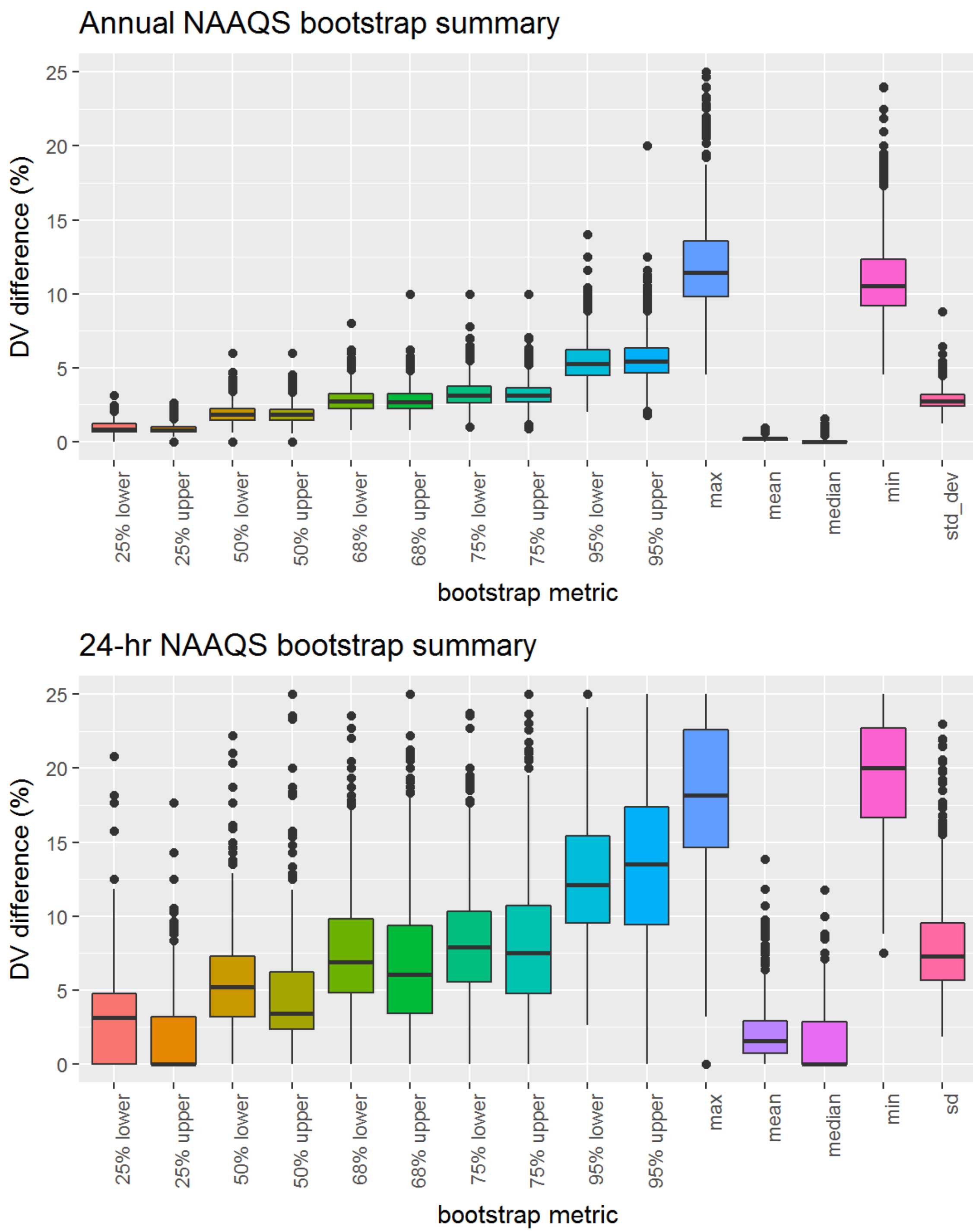


Figure 83: Bootstrap results for the 2003 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

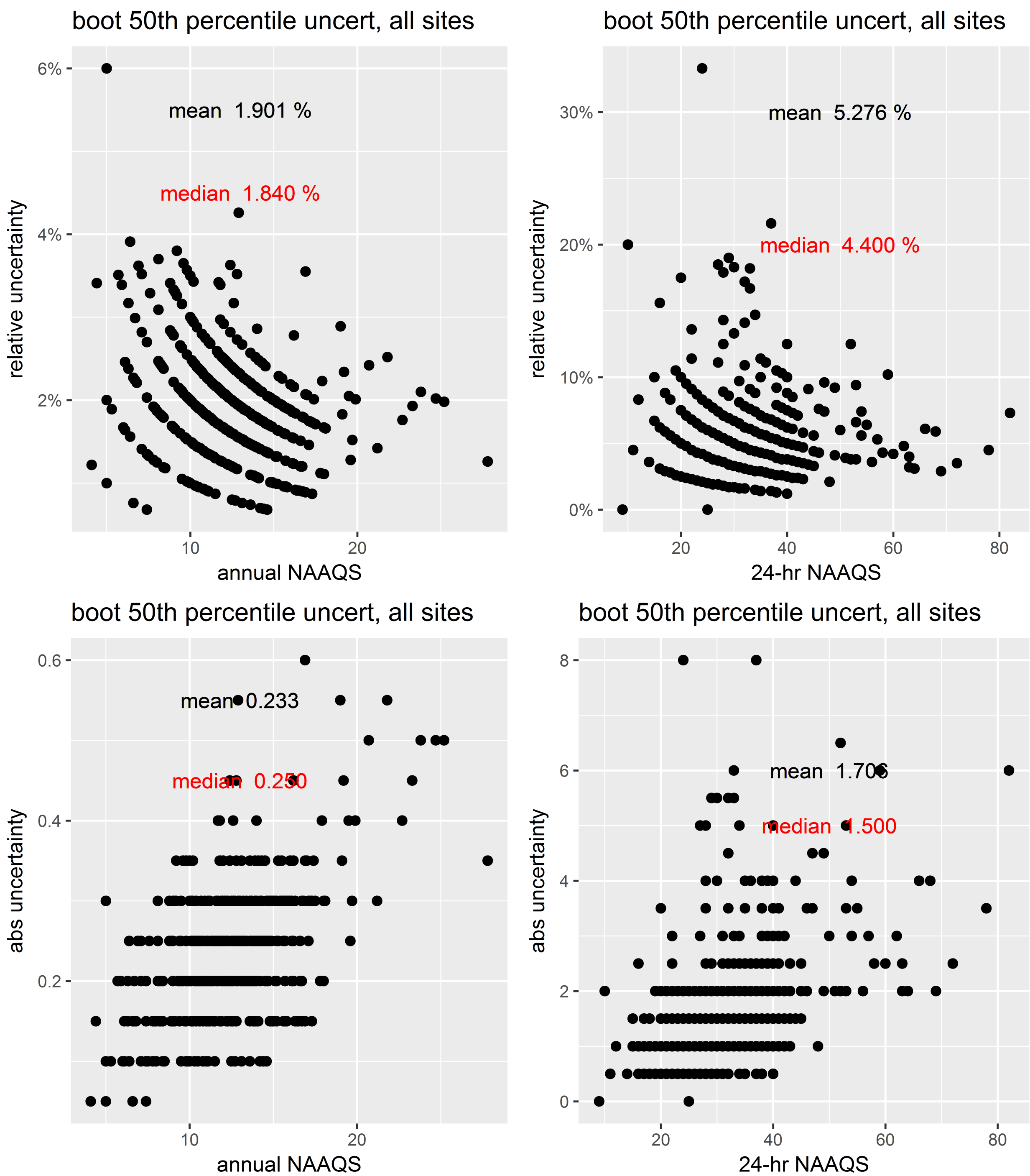


Figure 84: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

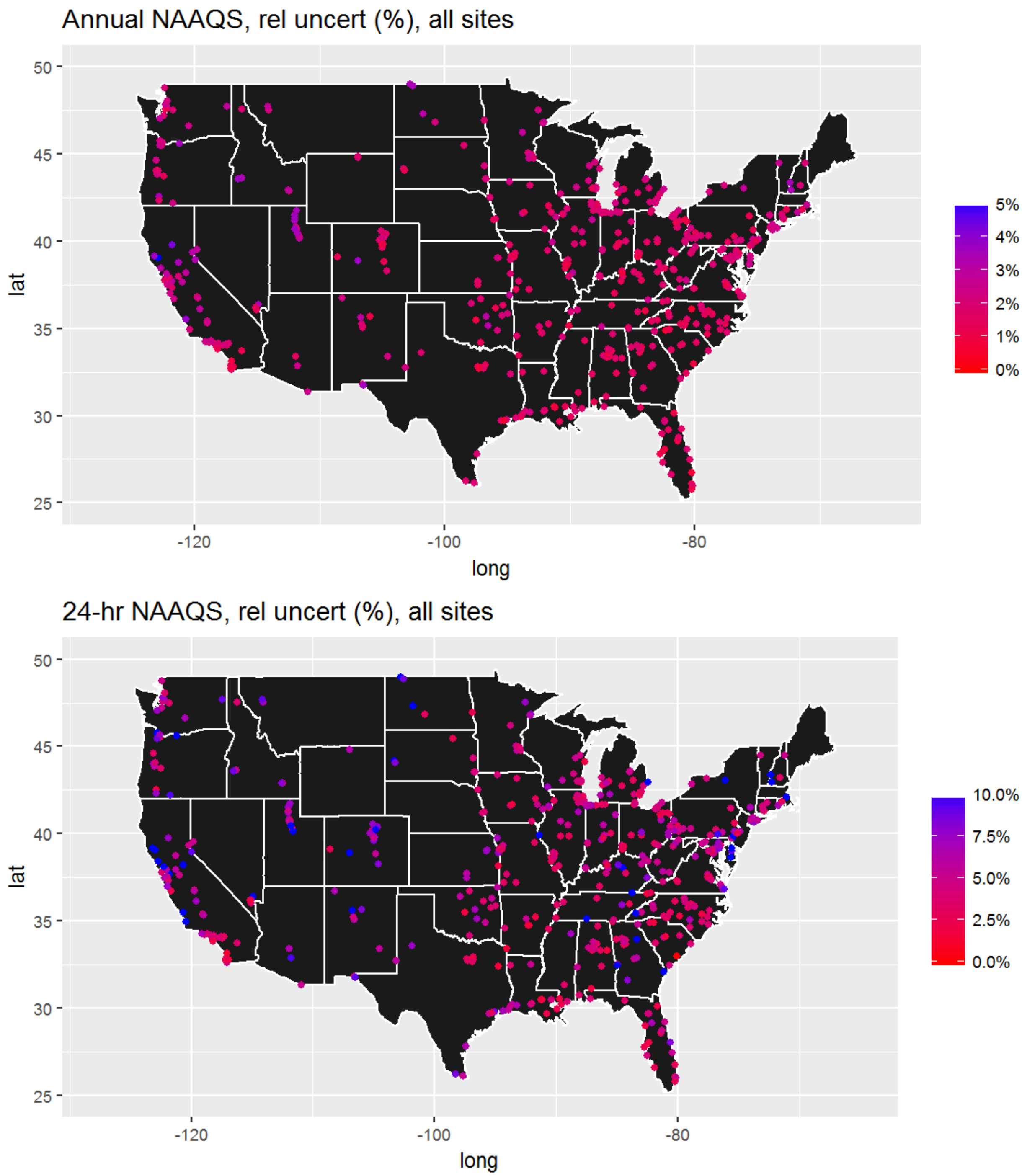


Figure 85: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2003 $PM_{2.5}$ DVs.

3.14 2000-2002 $PM_{2.5}$ bootstrap results

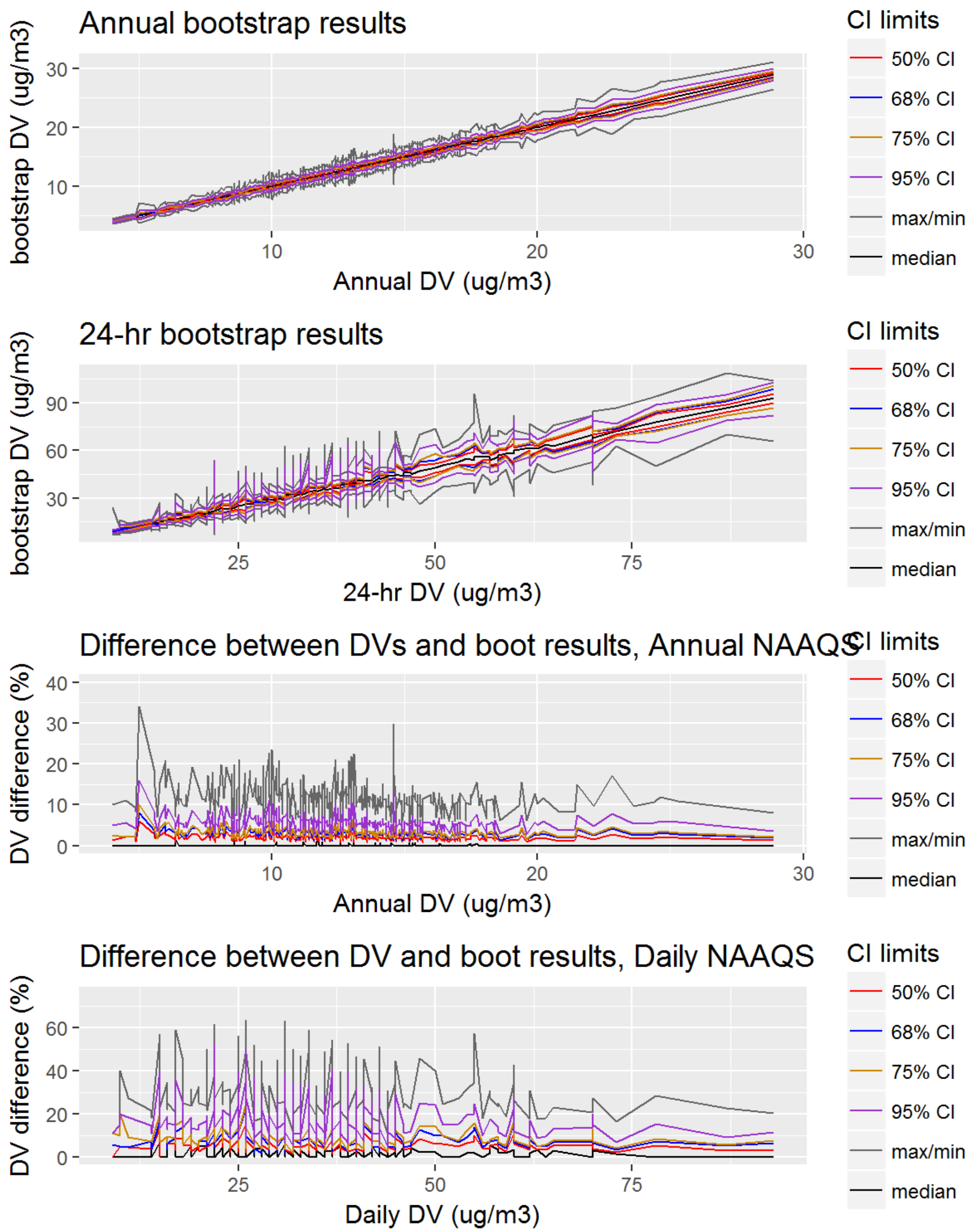


Figure 86: Bootstrap results for the 2002 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

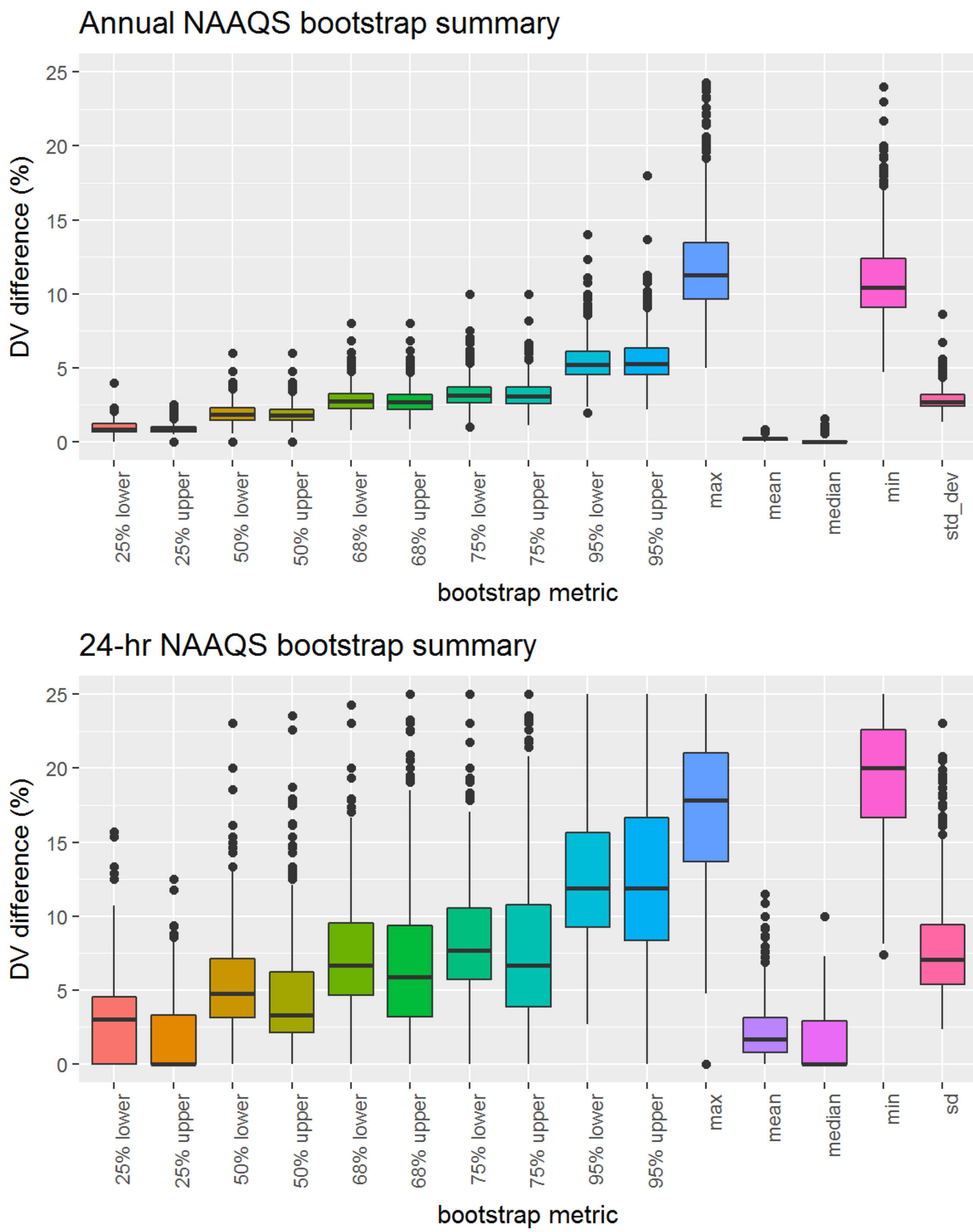


Figure 87: Bootstrap results for the 2002 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

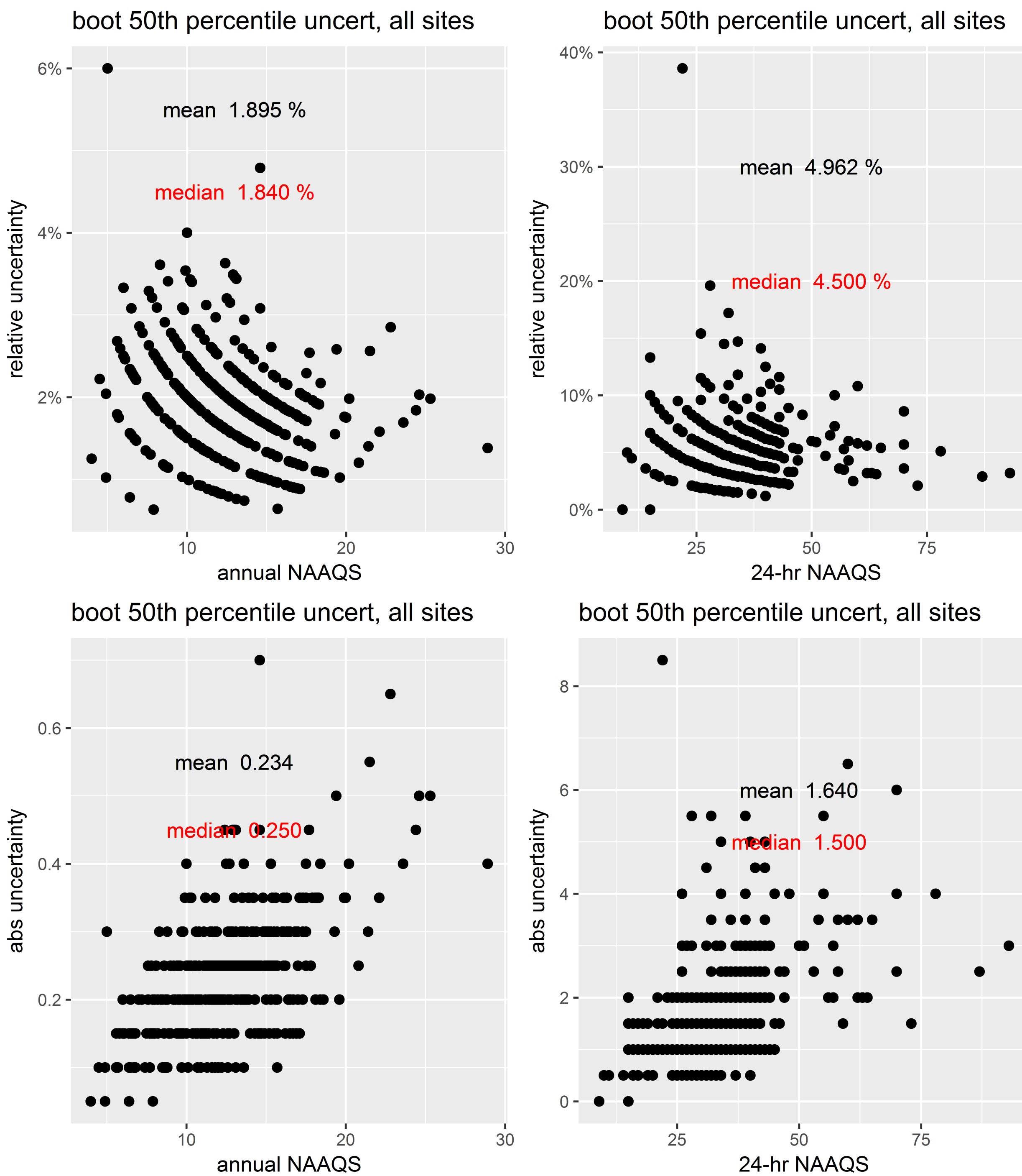


Figure 88: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

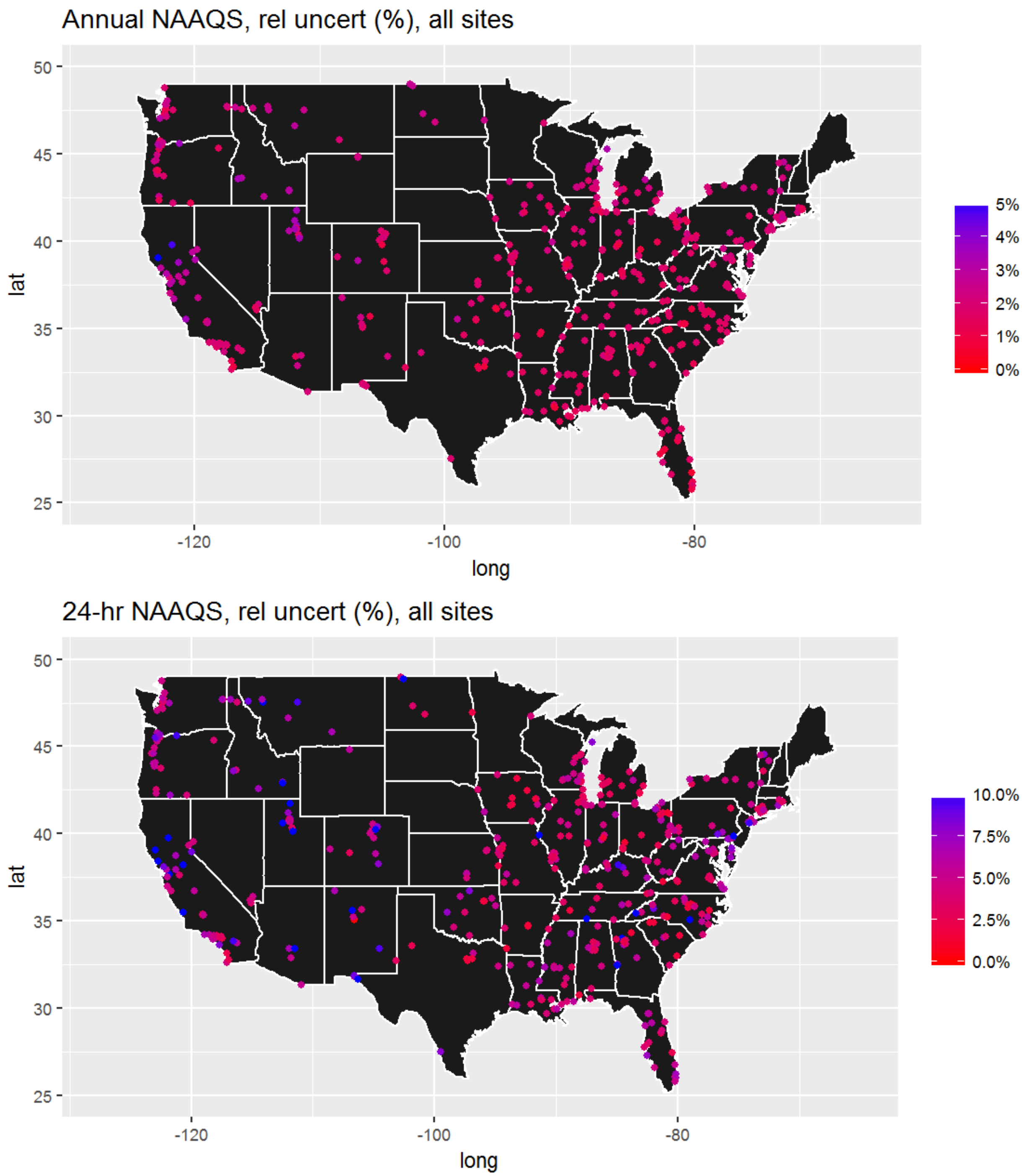
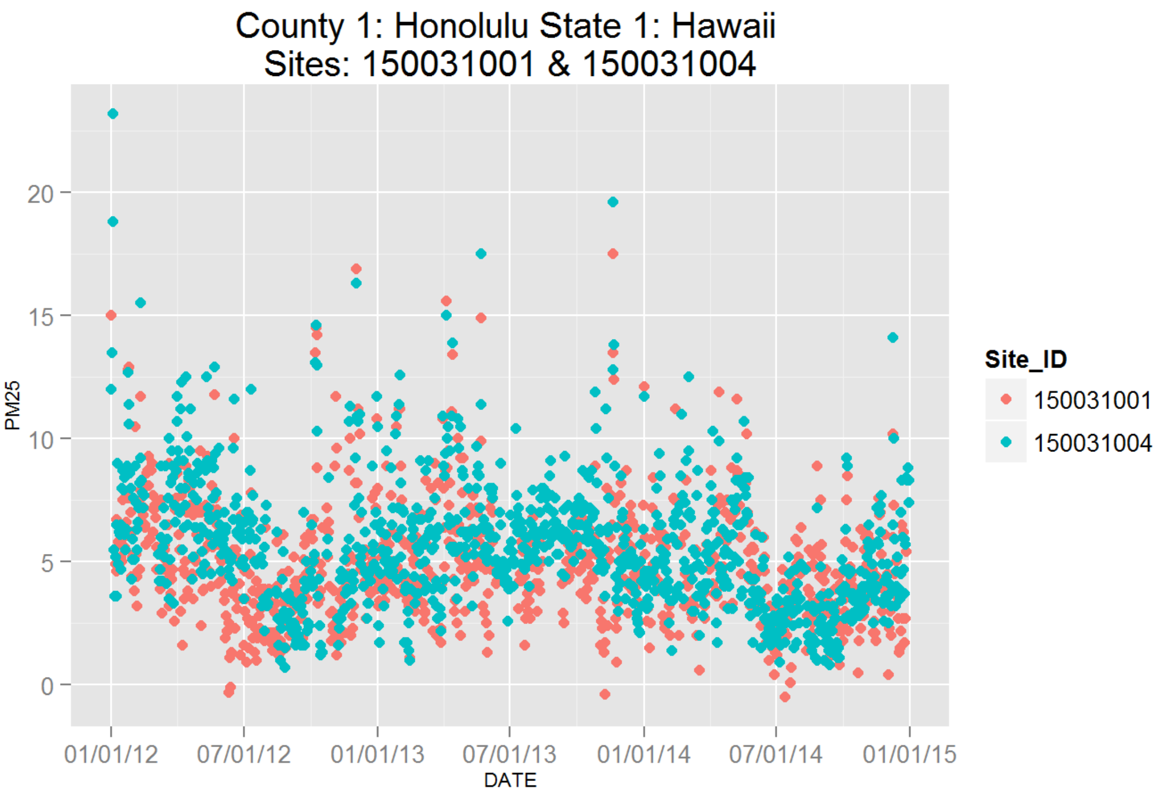


Figure 89: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2002 $PM_{2.5}$ DVs.

4 Comparison plots of nearby sites

Comparison of $PM_{2.5}$ data for paired, nearby sites for the spatial analysis conducted in Section 3.1.2.



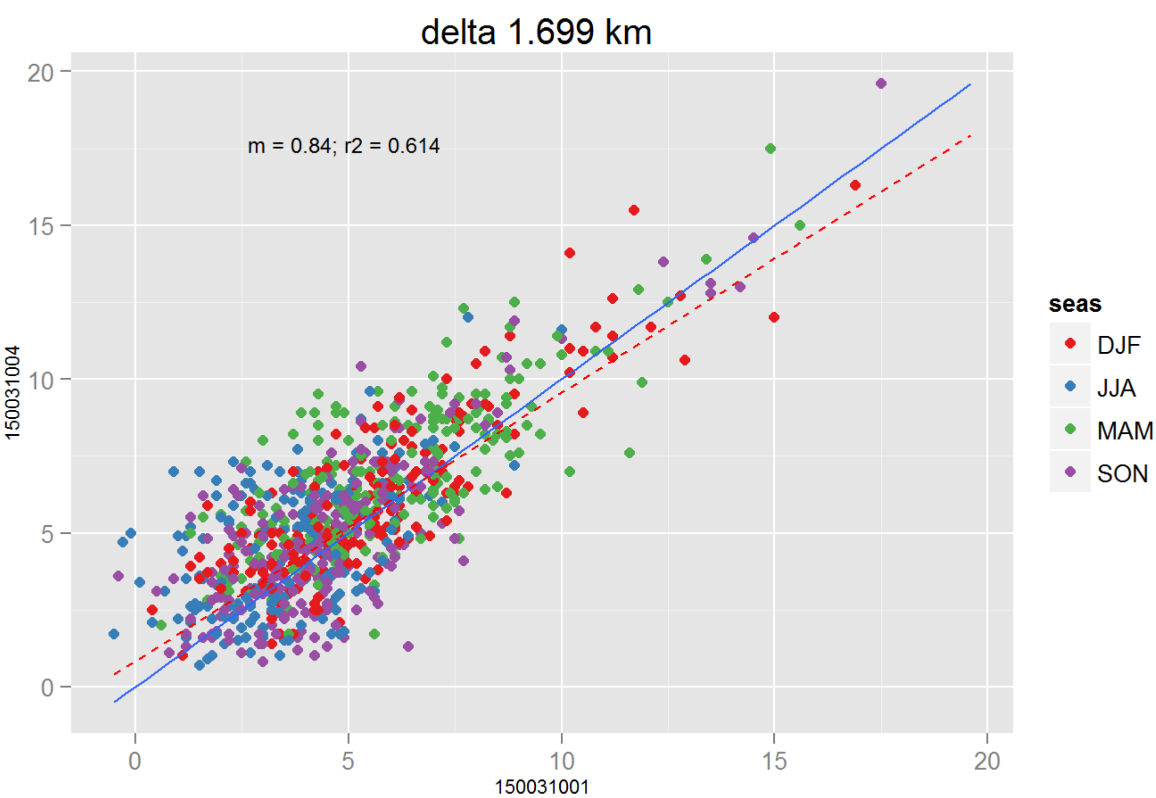


Figure 90: Comparison of $PM_{2.5}$ data for sites 150031001 and 150031001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

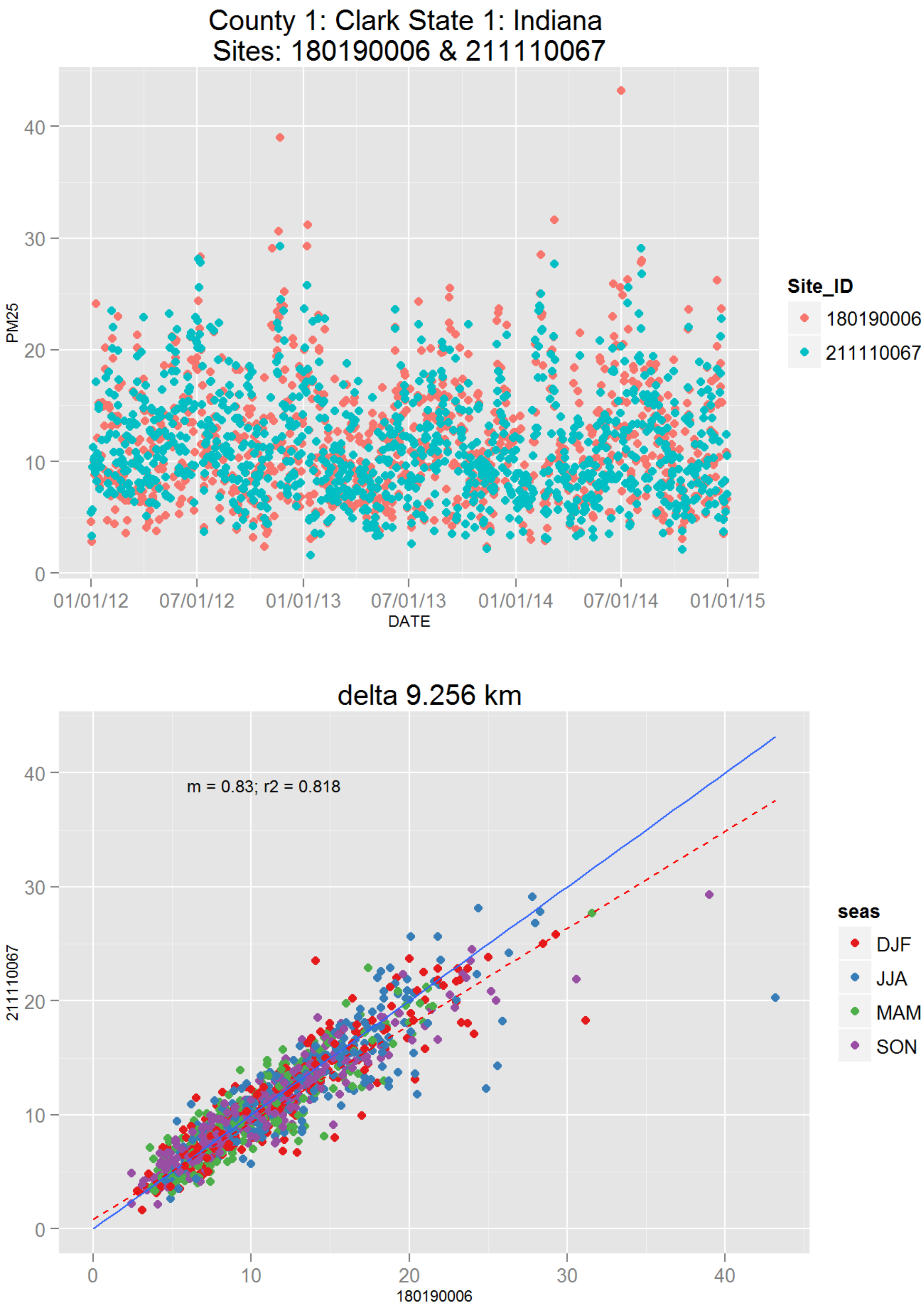


Figure 91: Comparison of $PM_{2.5}$ data for sites 180190006 and 180190006. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

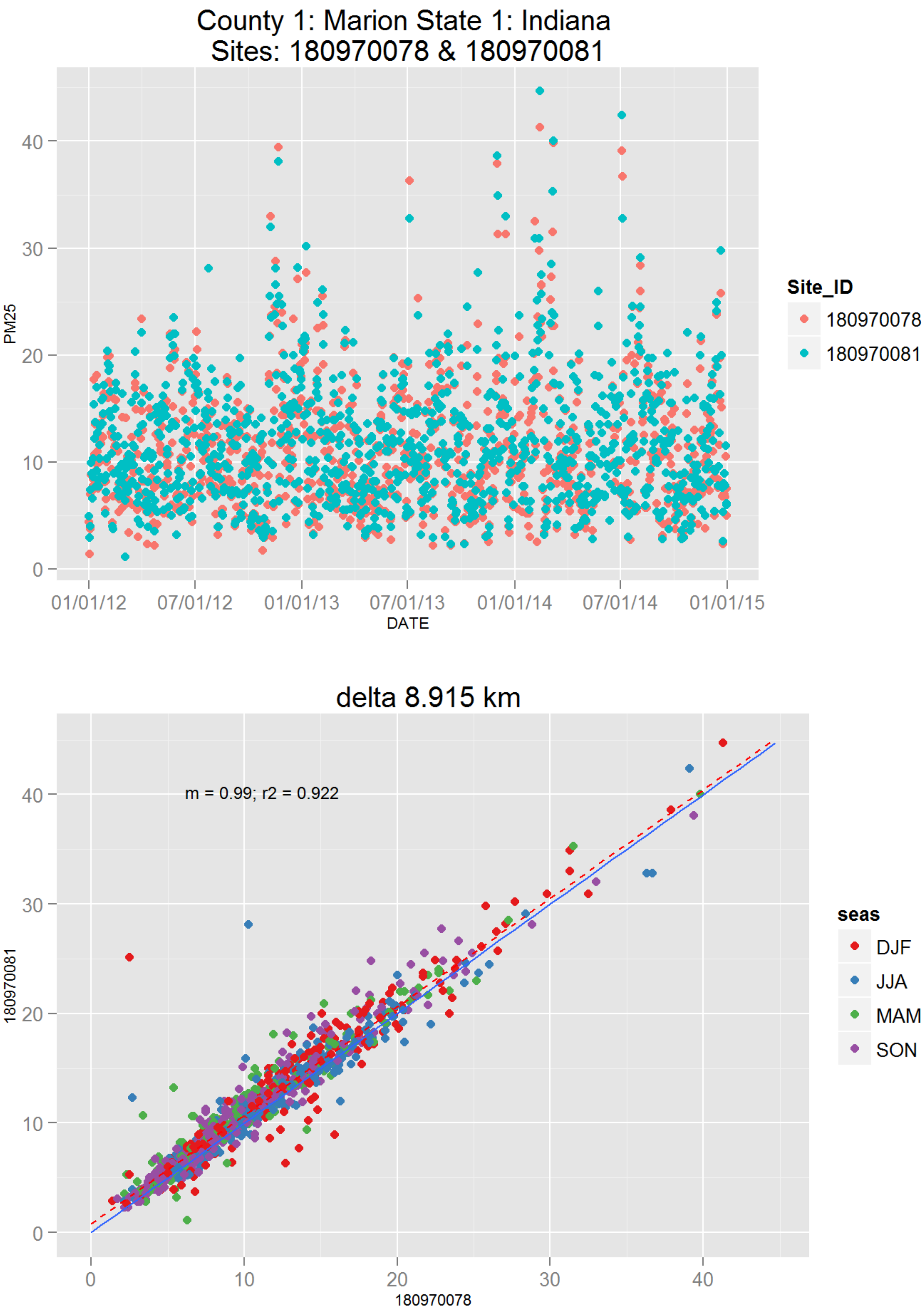
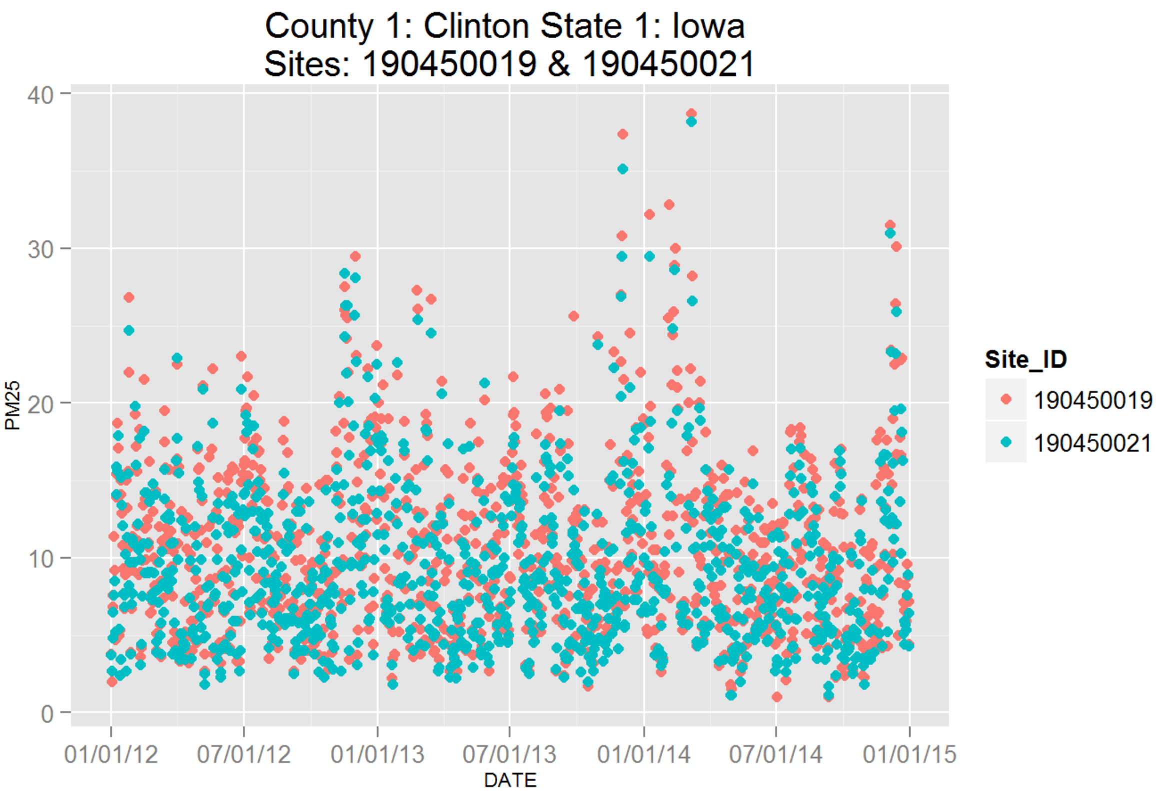


Figure 92: Comparison of $PM_{2.5}$ data for sites 180970078 and 180970078. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



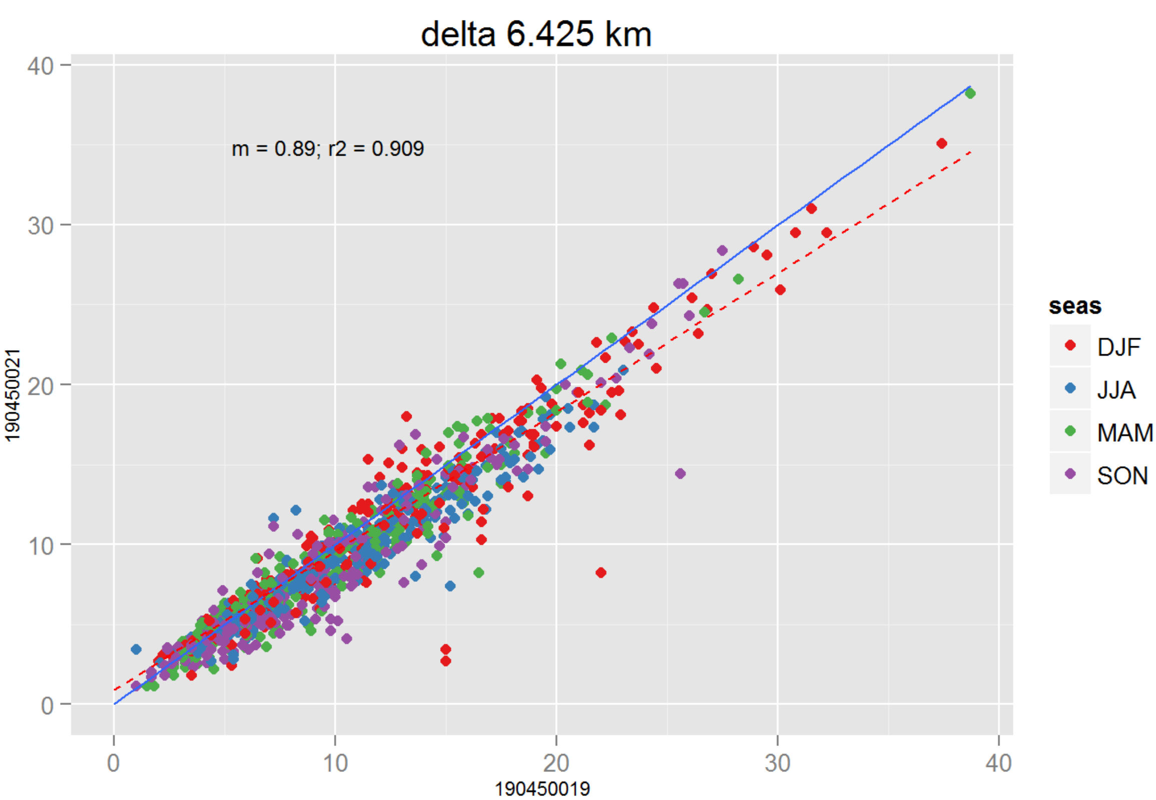


Figure 93: Comparison of $PM_{2.5}$ data for sites 190450019 and 190450019. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 94: Comparison of $PM_{2.5}$ data for sites 220330009 and 220330009. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

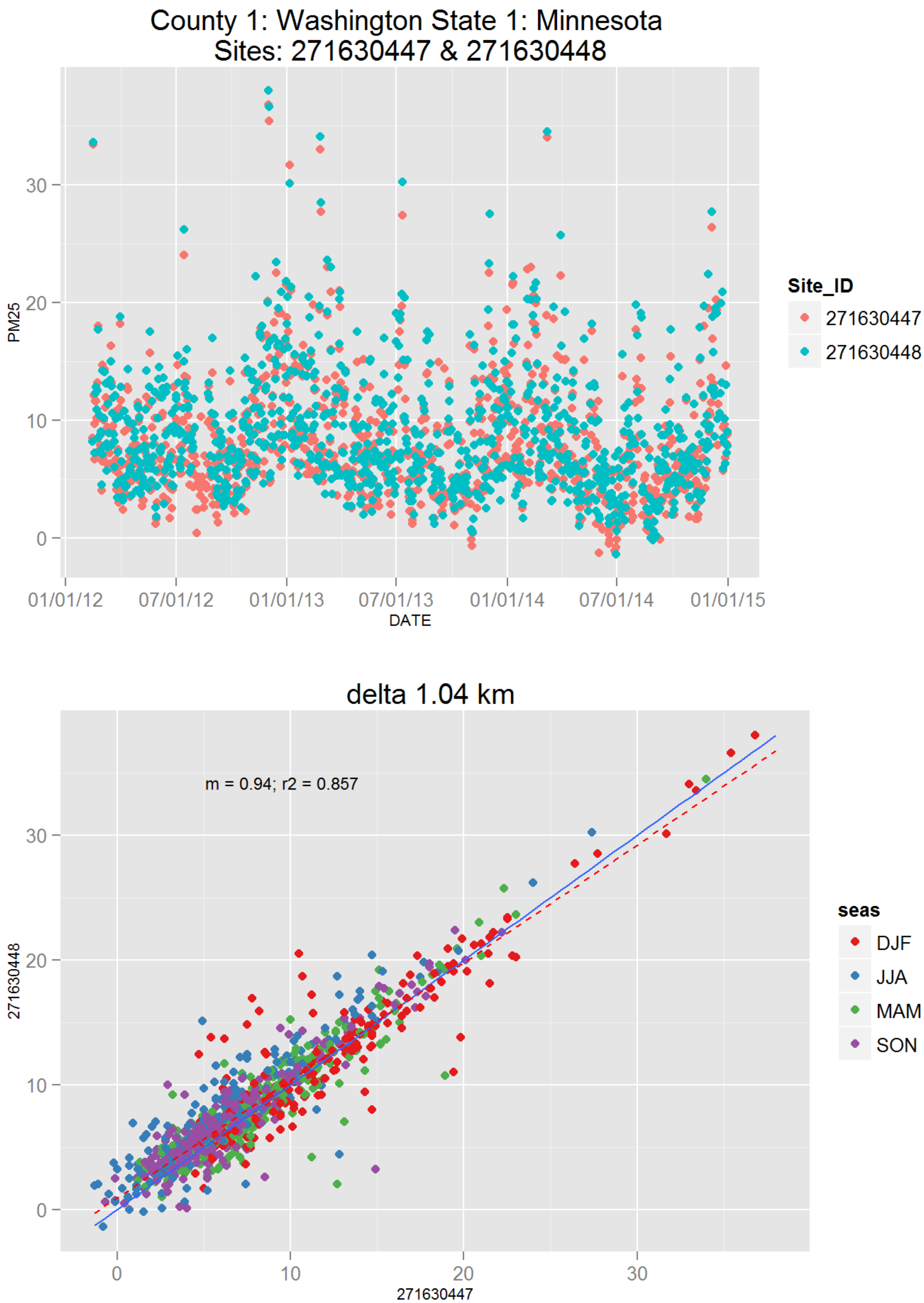


Figure 95: Comparison of $PM_{2.5}$ data for sites 271630447 and 271630447. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

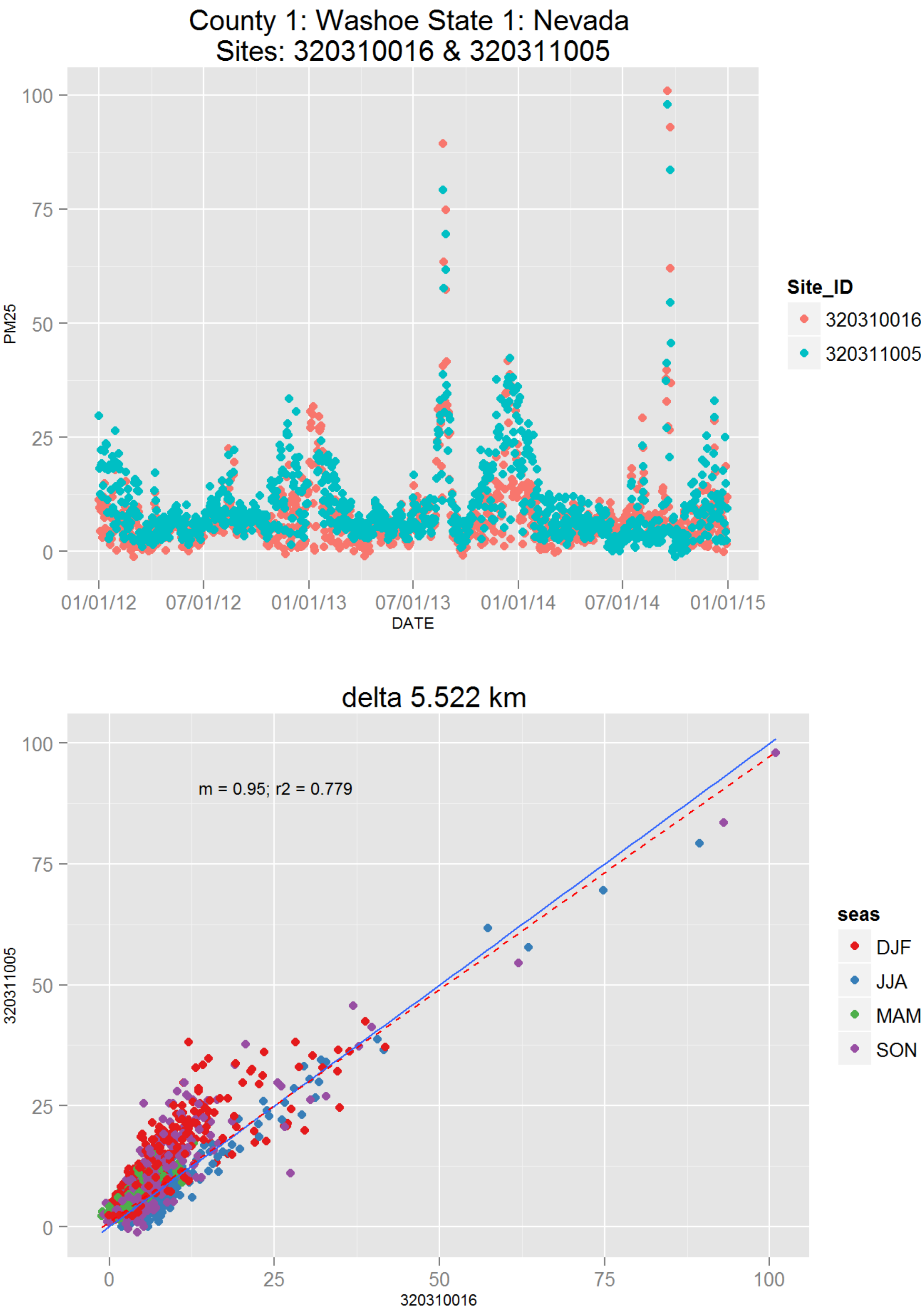


Figure 96: Comparison of $PM_{2.5}$ data for sites 320310016 and 320310016. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

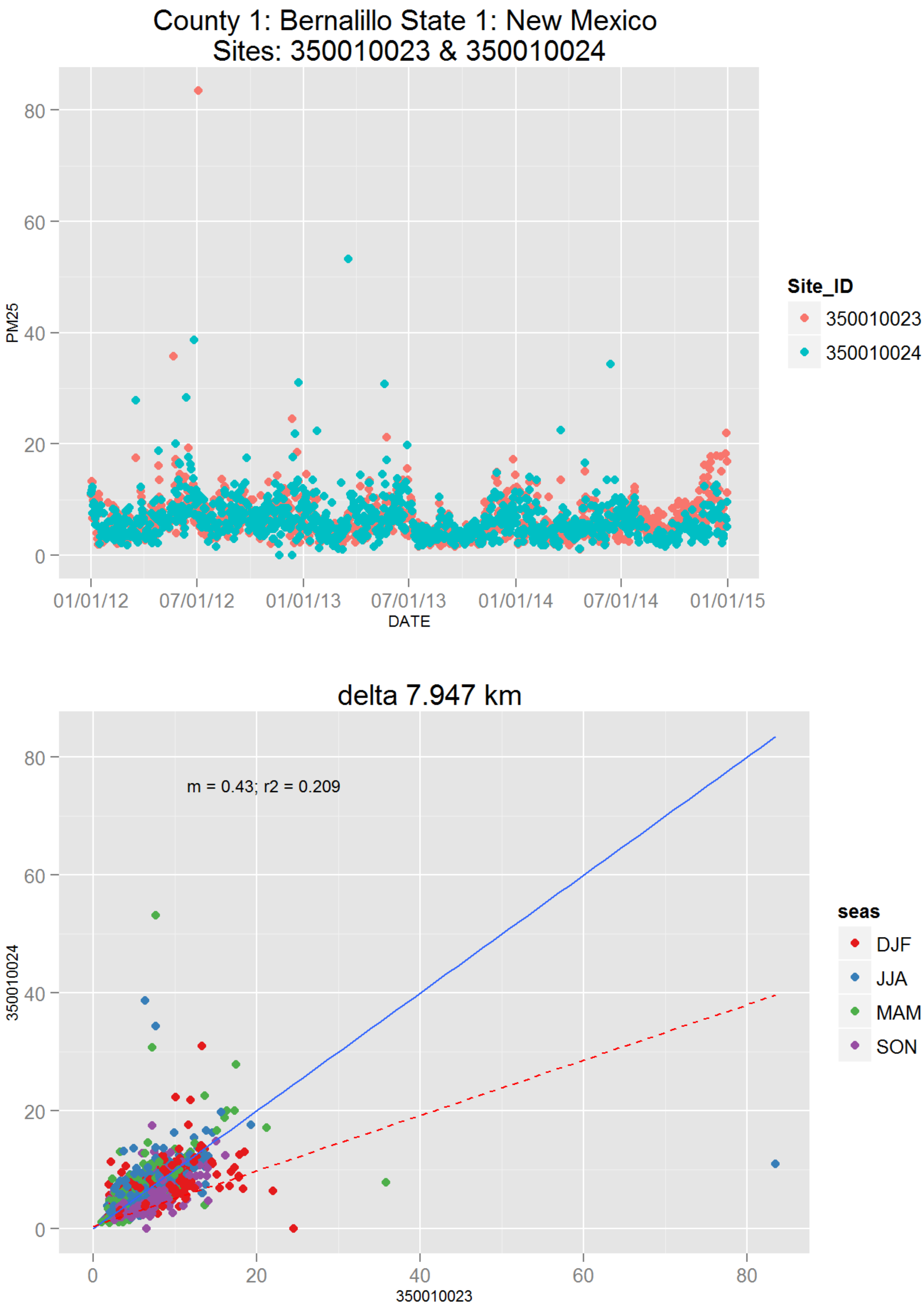


Figure 97: Comparison of $PM_{2.5}$ data for sites 350010023 and 350010023. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

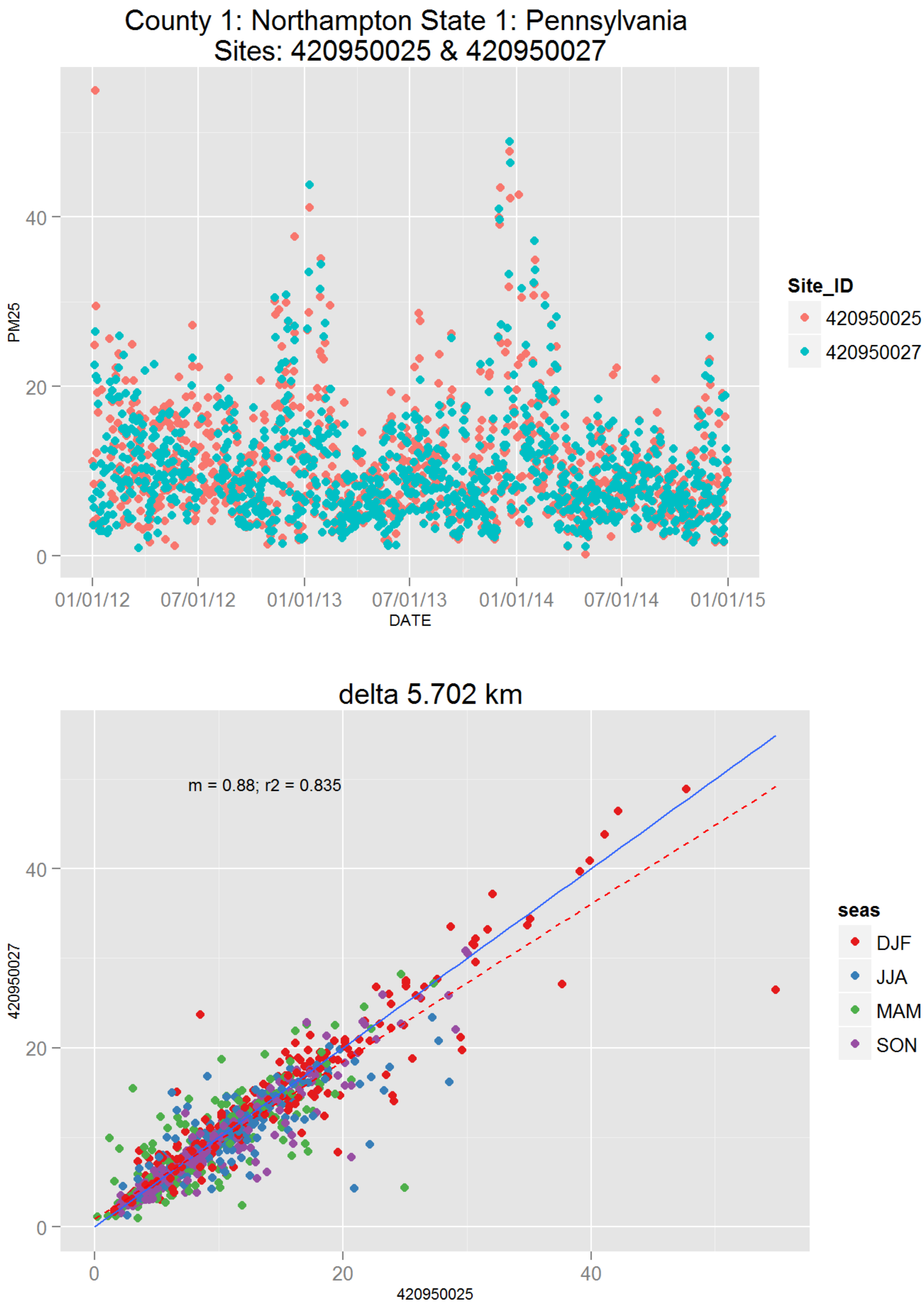


Figure 98: Comparison of $PM_{2.5}$ data for sites 420950025 and 420950025. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

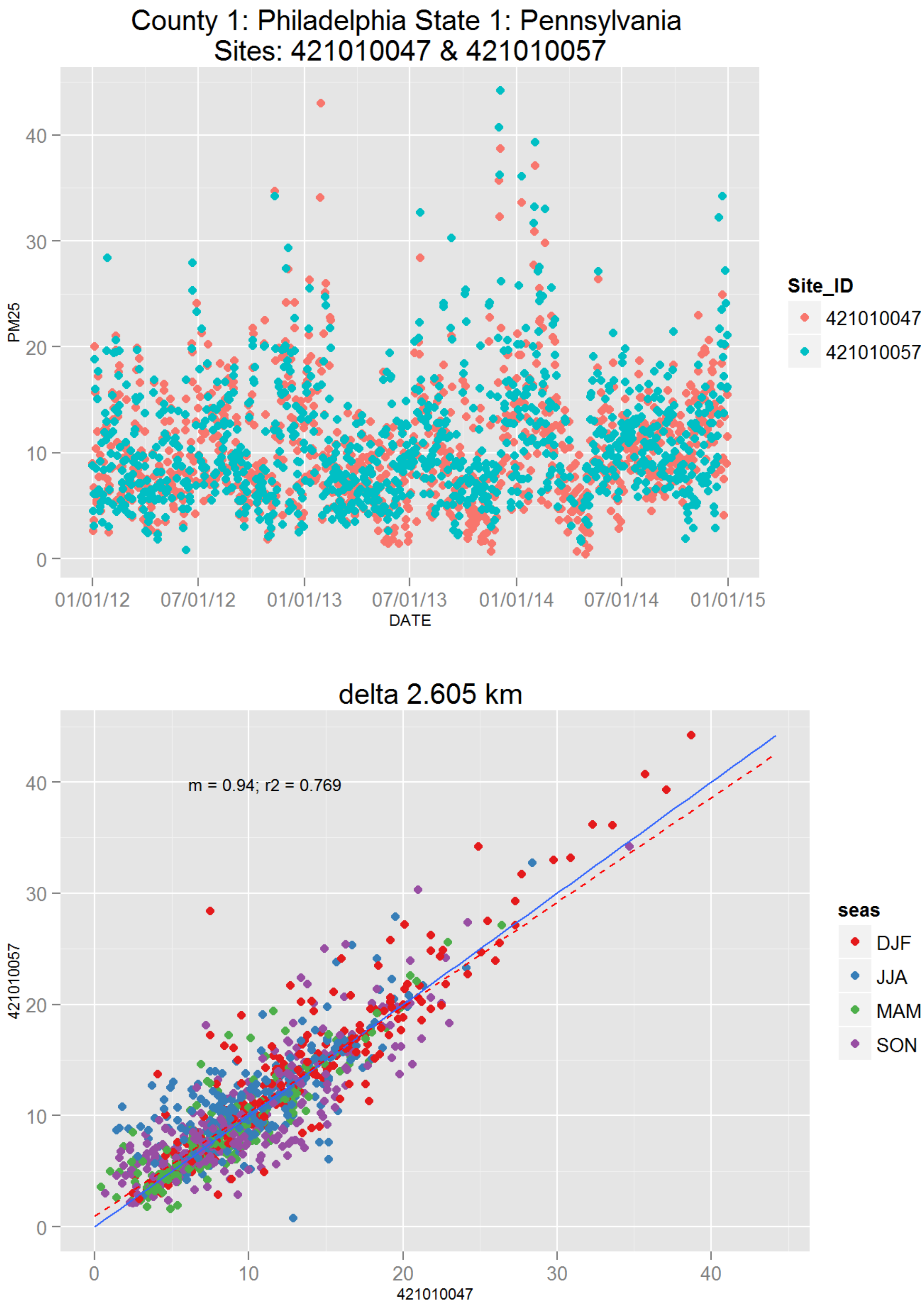
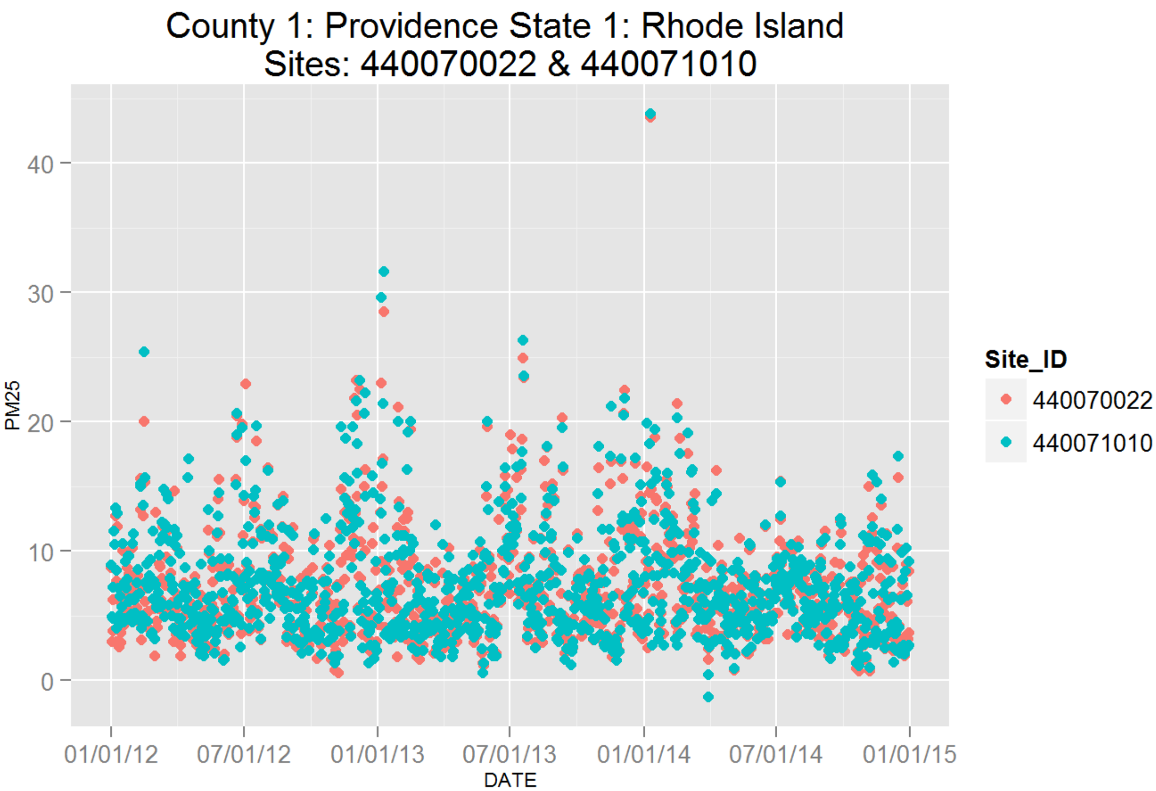


Figure 99: Comparison of $PM_{2.5}$ data for sites 421010047 and 421010047. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 100: Comparison of $PM_{2.5}$ data for sites 421010055 and 421010055. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



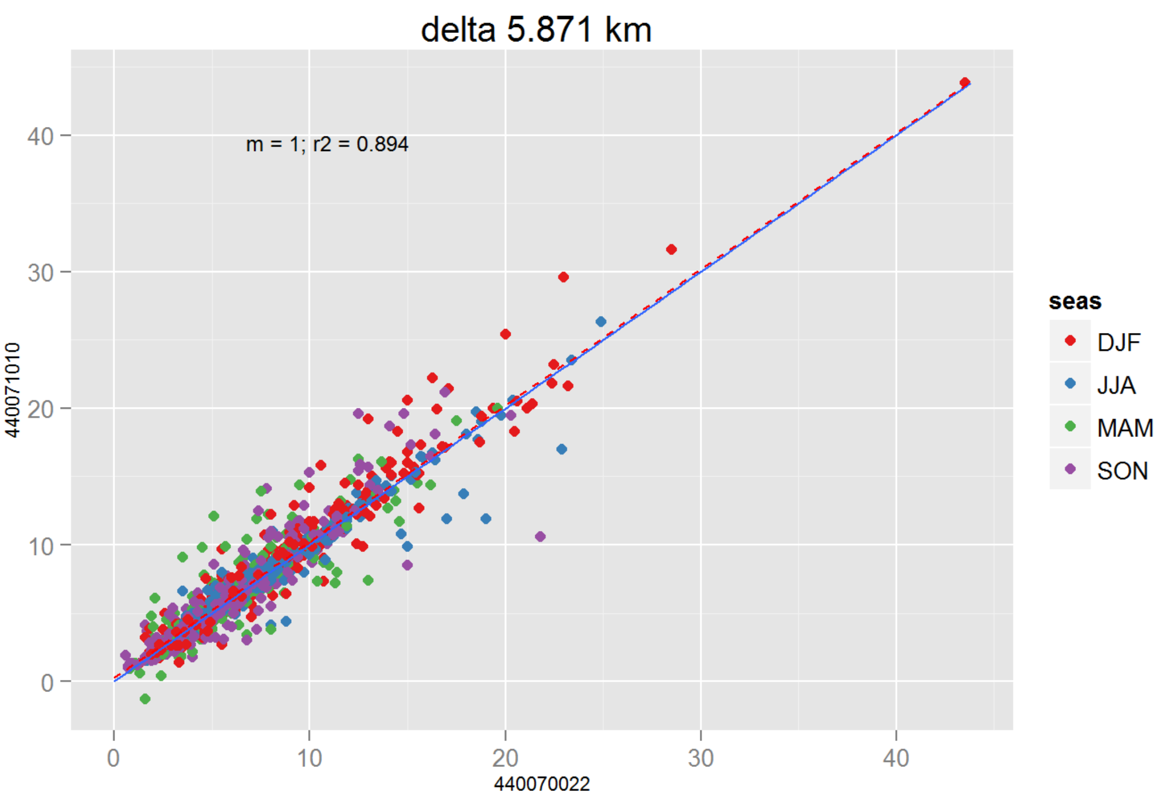


Figure 101: Comparison of $PM_{2.5}$ data for sites 440070022 and 440070022. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

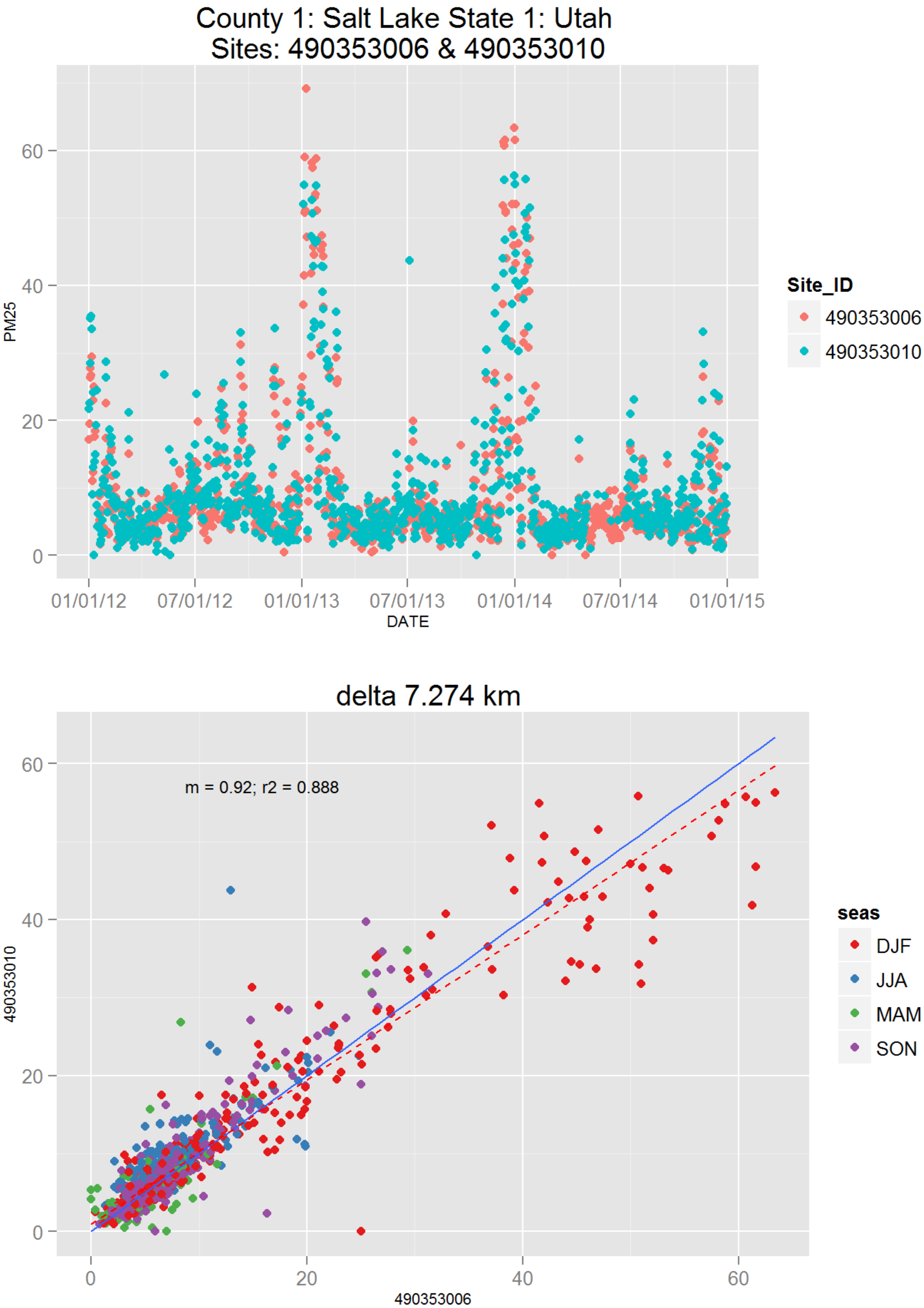


Figure 102: Comparison of $PM_{2.5}$ data for sites 490353006 and 490353006. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

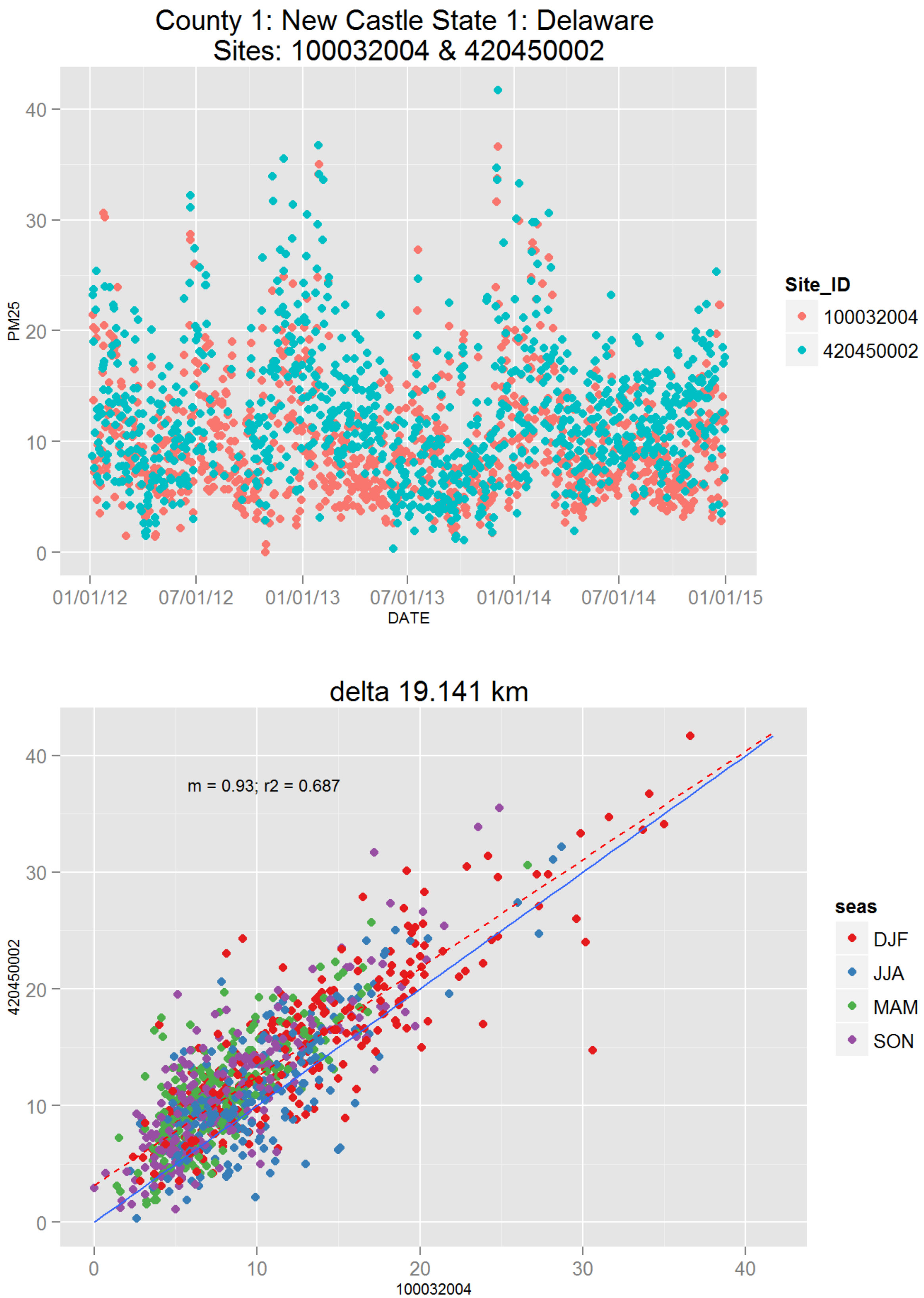


Figure 103: Comparison of $PM_{2.5}$ data for sites 100032004 and 100032004. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

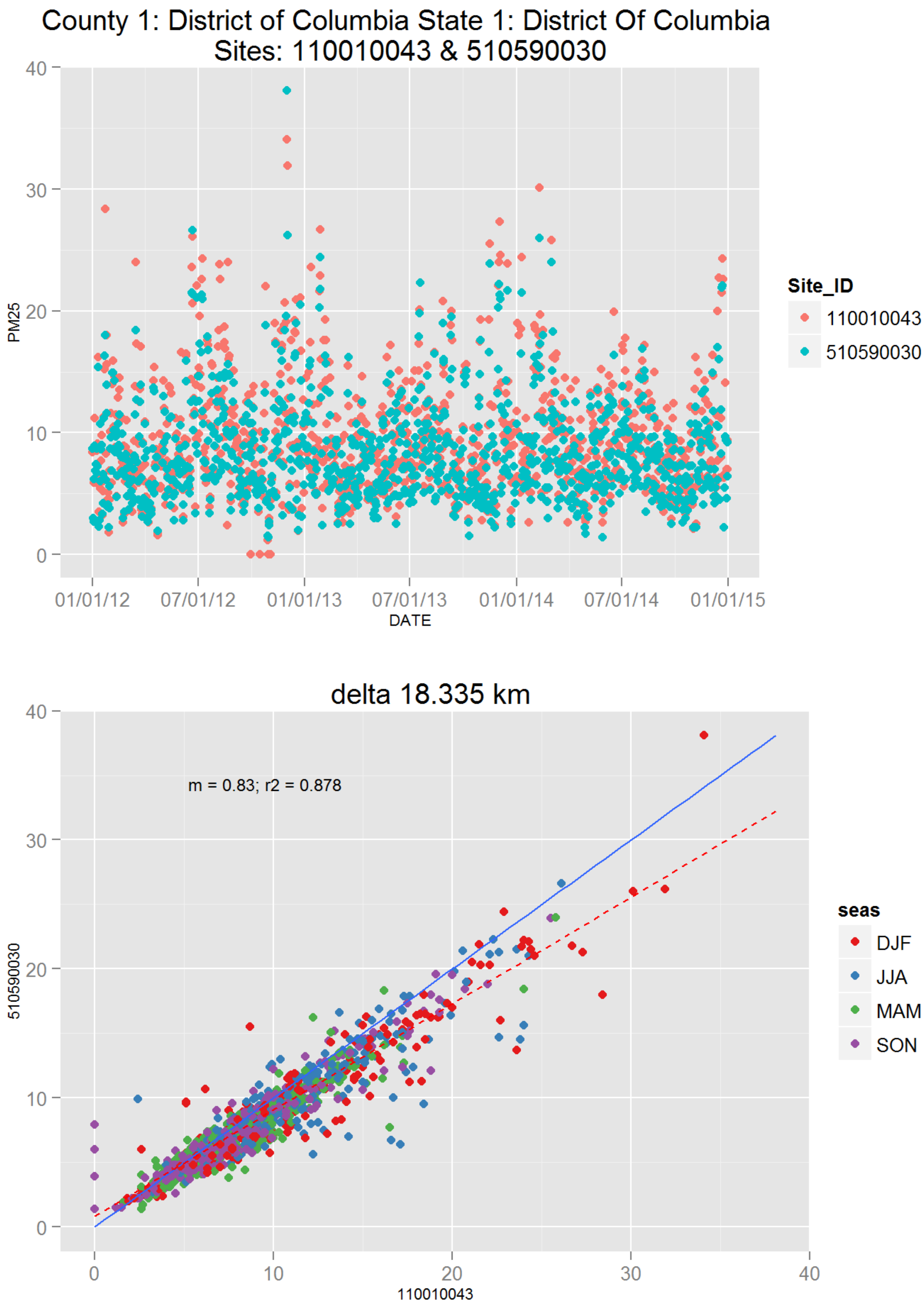


Figure 104: Comparison of $PM_{2.5}$ data for sites 110010043 and 110010043. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 105: Comparison of $PM_{2.5}$ data for sites 130670003 and 130670003. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 106: Comparison of $PM_{2.5}$ data for sites 150011006 and 150011006. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

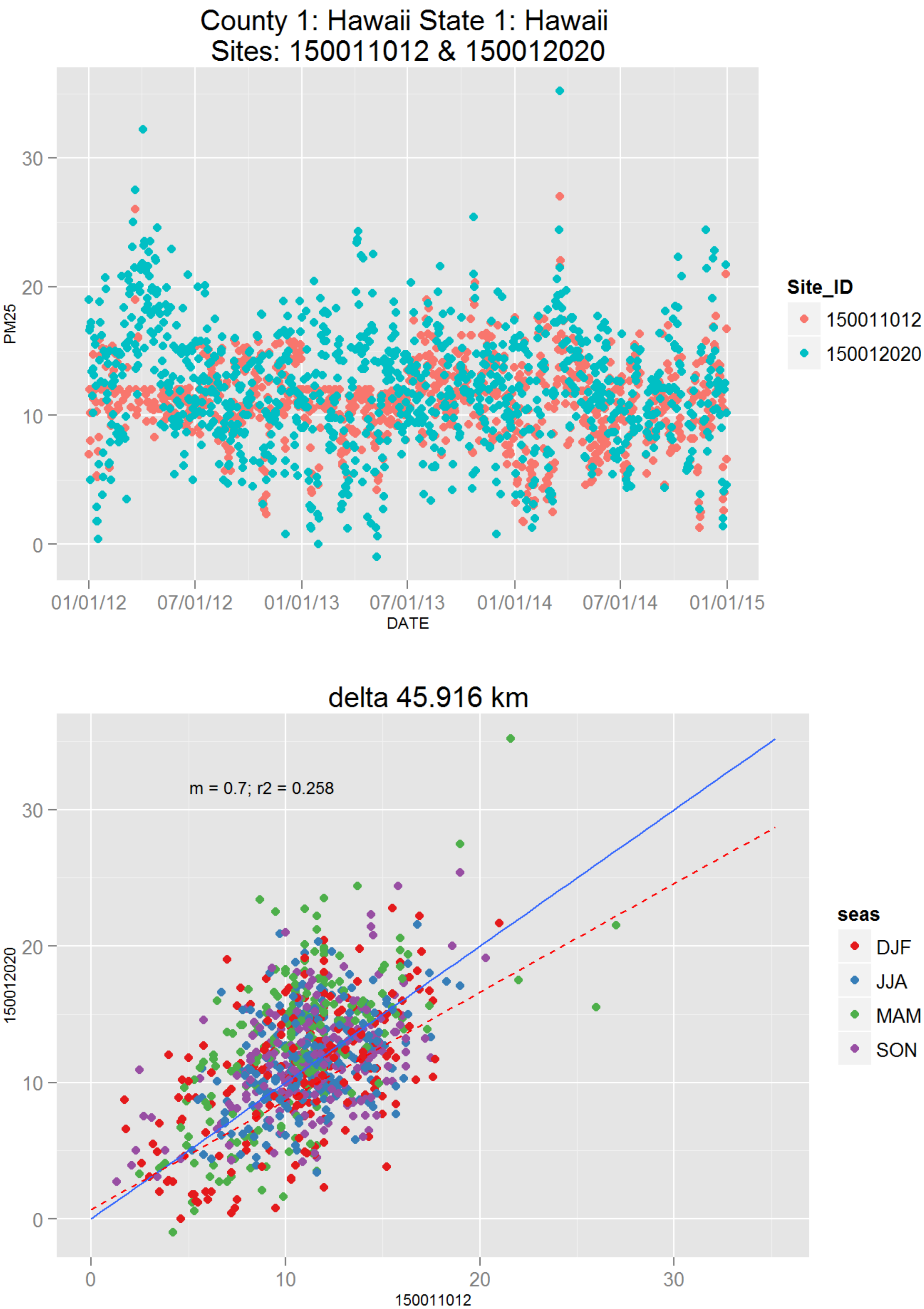


Figure 107: Comparison of $PM_{2.5}$ data for sites 150011012 and 150011012. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

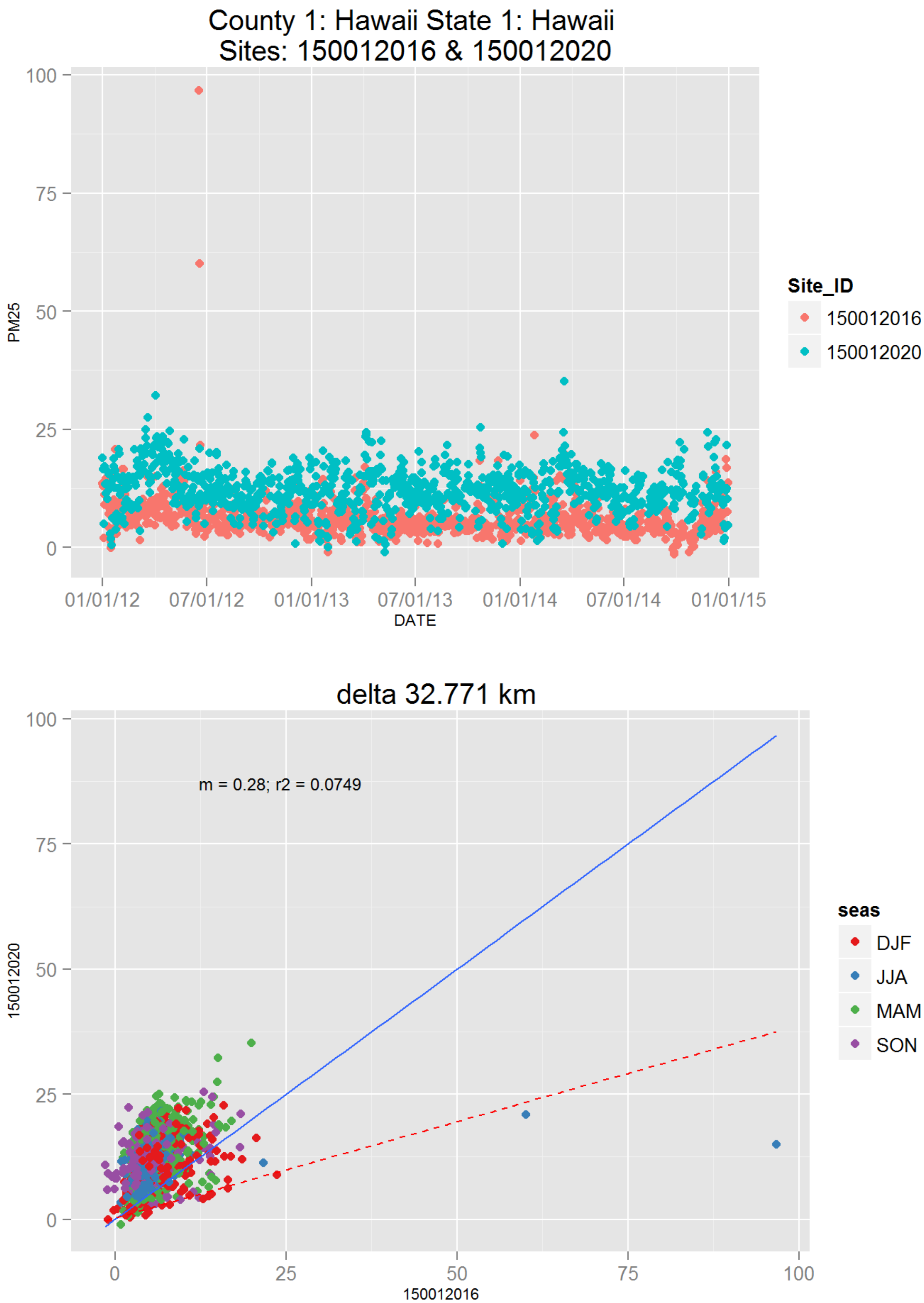
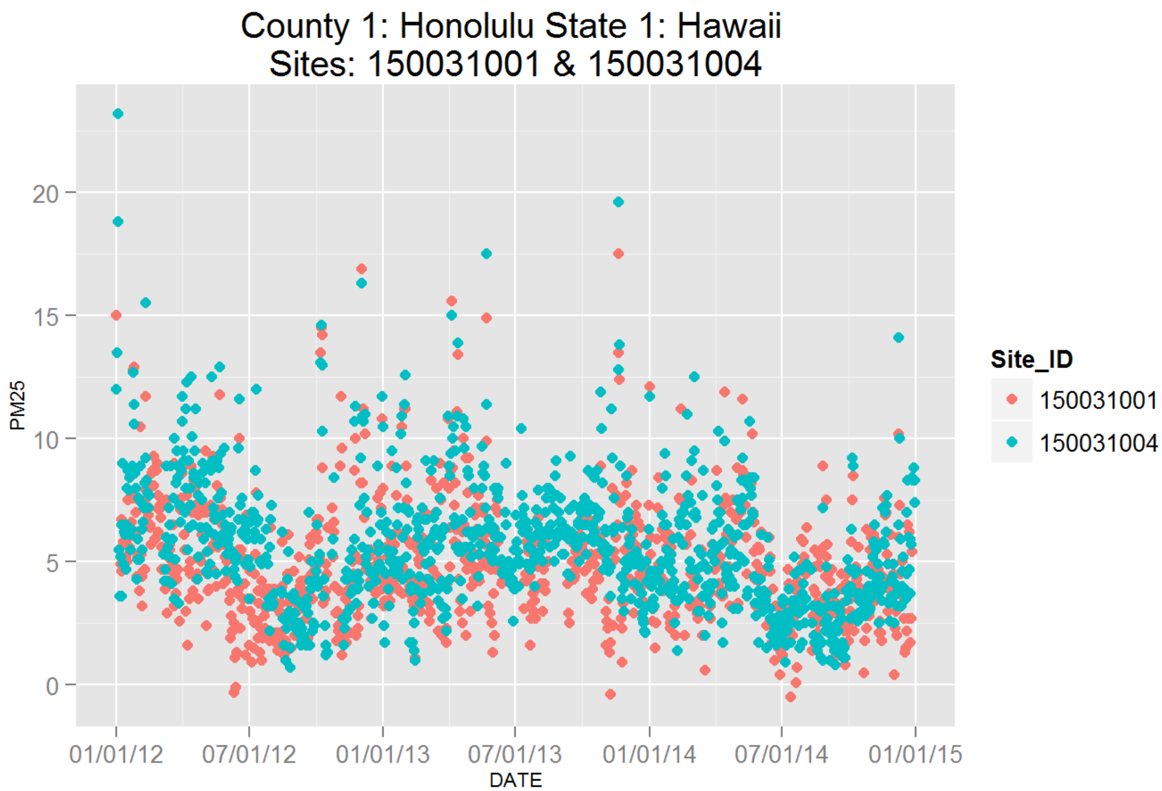


Figure 108: Comparison of $PM_{2.5}$ data for sites 150012016 and 150012014. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



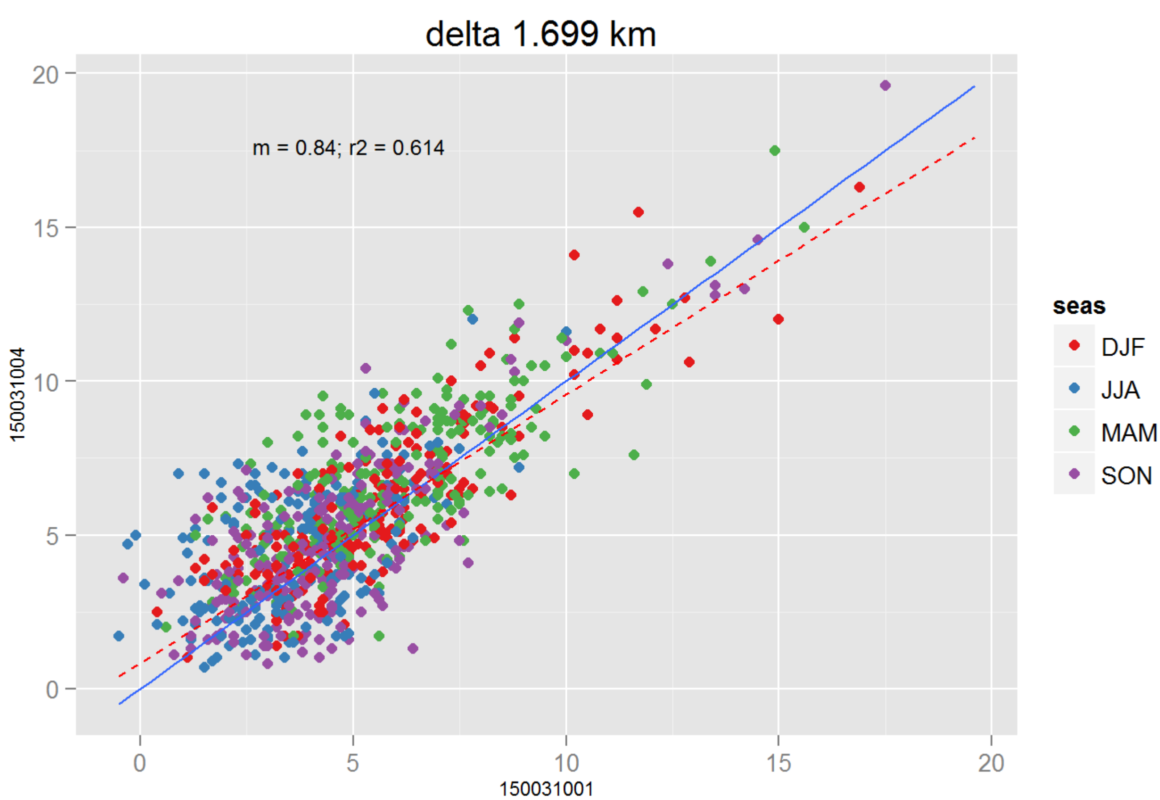


Figure 109: Comparison of $PM_{2.5}$ data for sites 150031001 and 150031001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

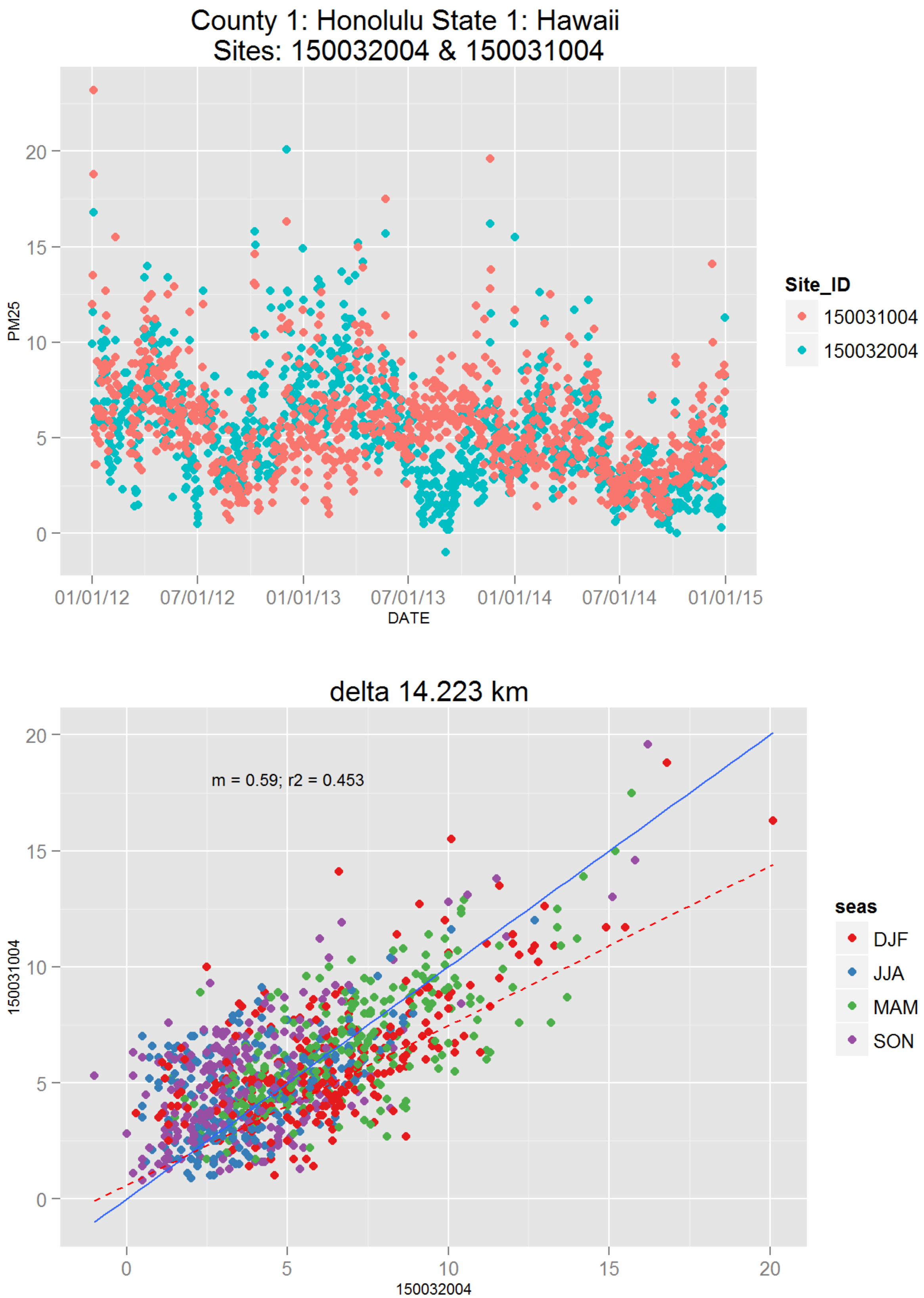


Figure 110: Comparison of $PM_{2.5}$ data for sites 150032004 and 150032004. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

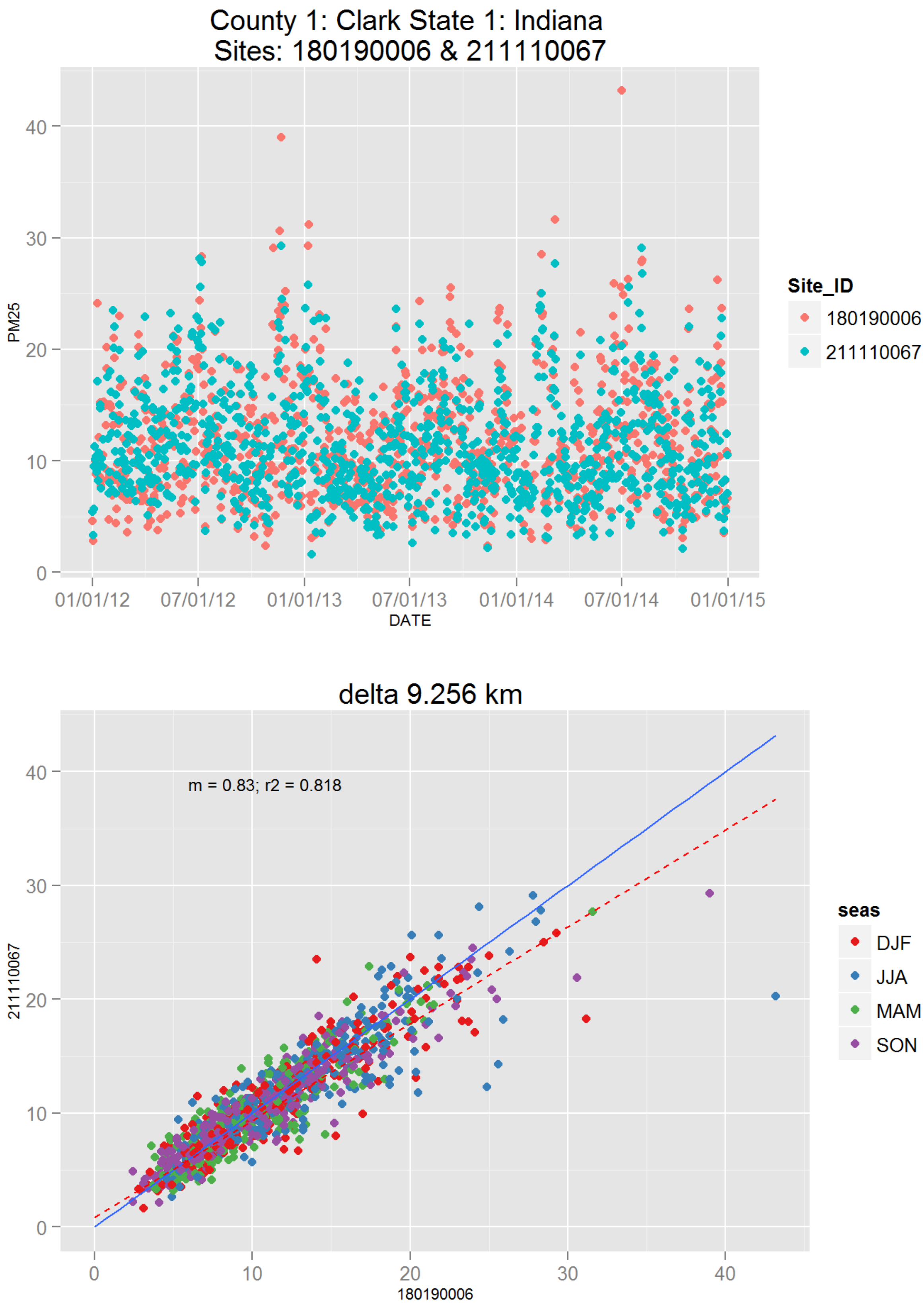


Figure 111: Comparison of $PM_{2.5}$ data for sites 180190006 and 180190006. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

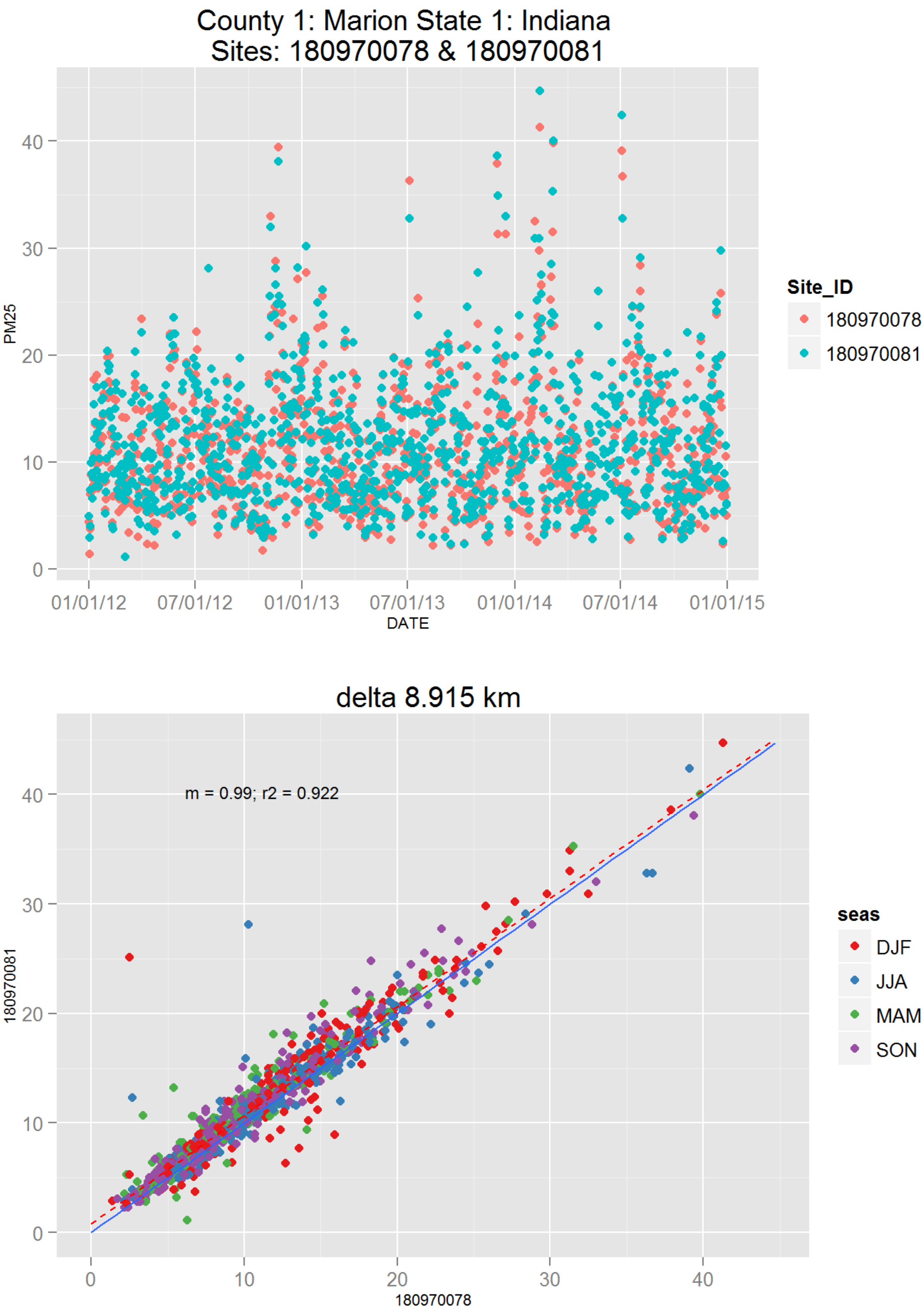


Figure 112: Comparison of $PM_{2.5}$ data for sites 180970078 and 180970078. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

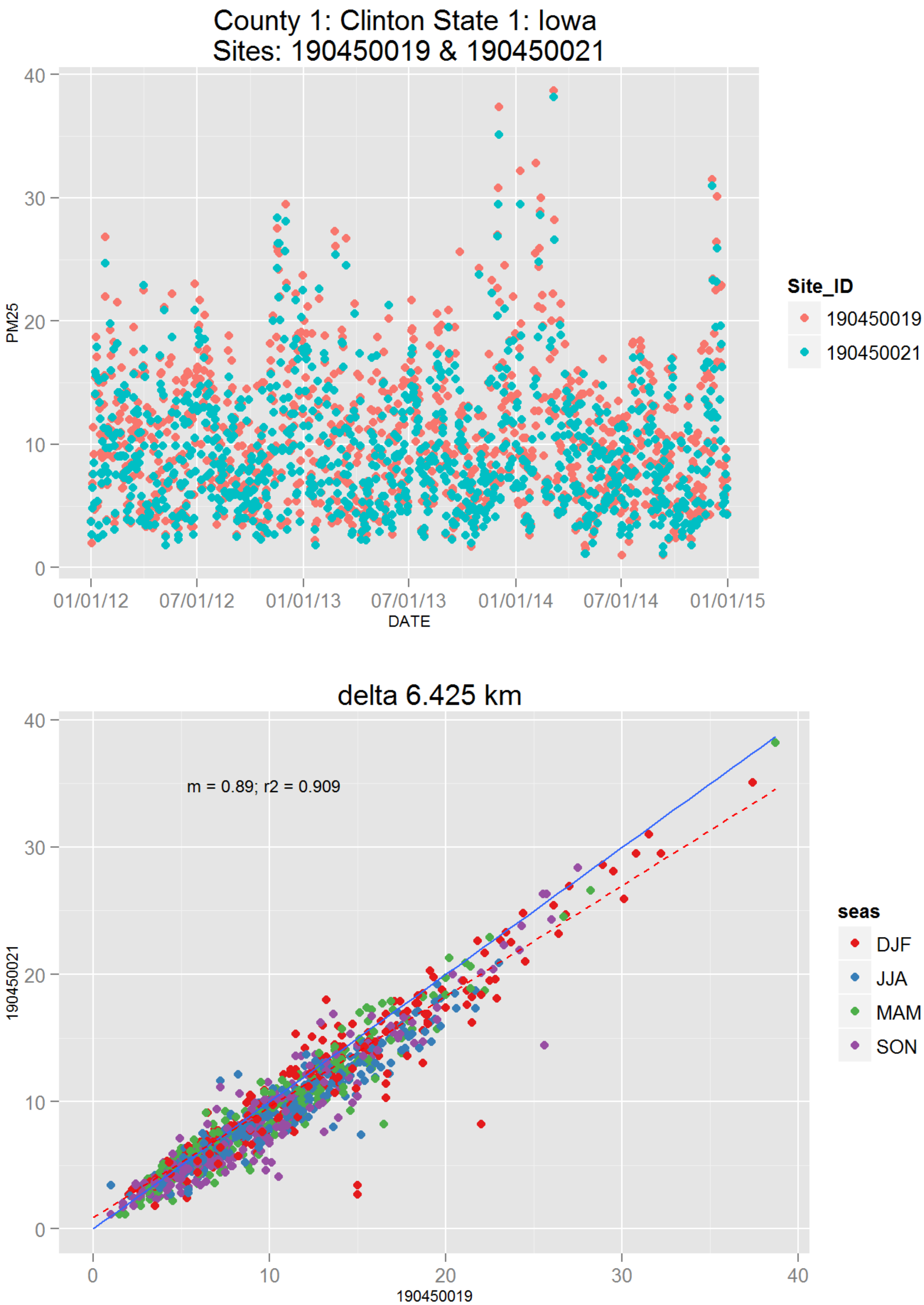


Figure 113: Comparison of $PM_{2.5}$ data for sites 190450019 and 190450019. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

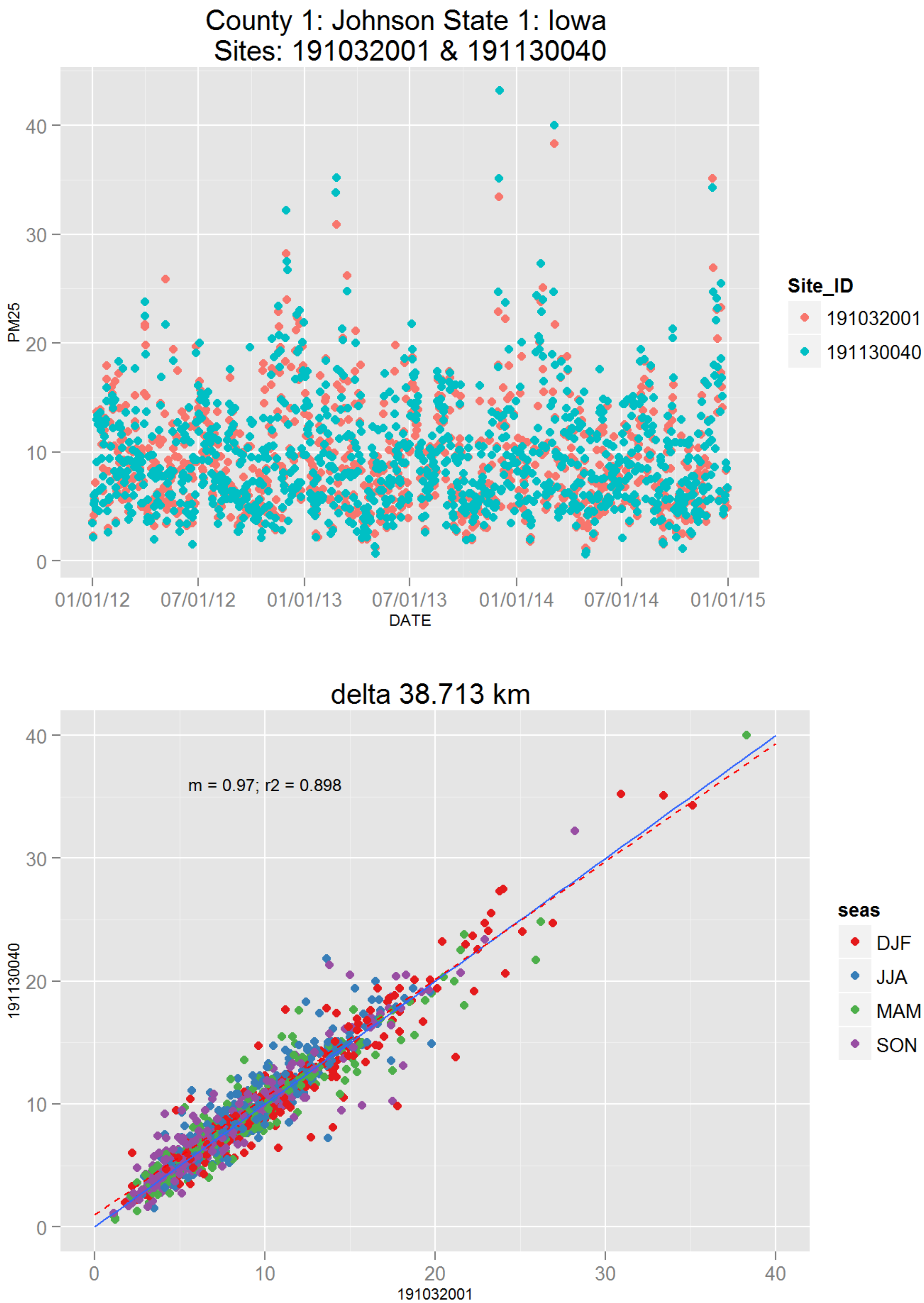


Figure 114: Comparison of $PM_{2.5}$ data for sites 191032001 and 191032001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

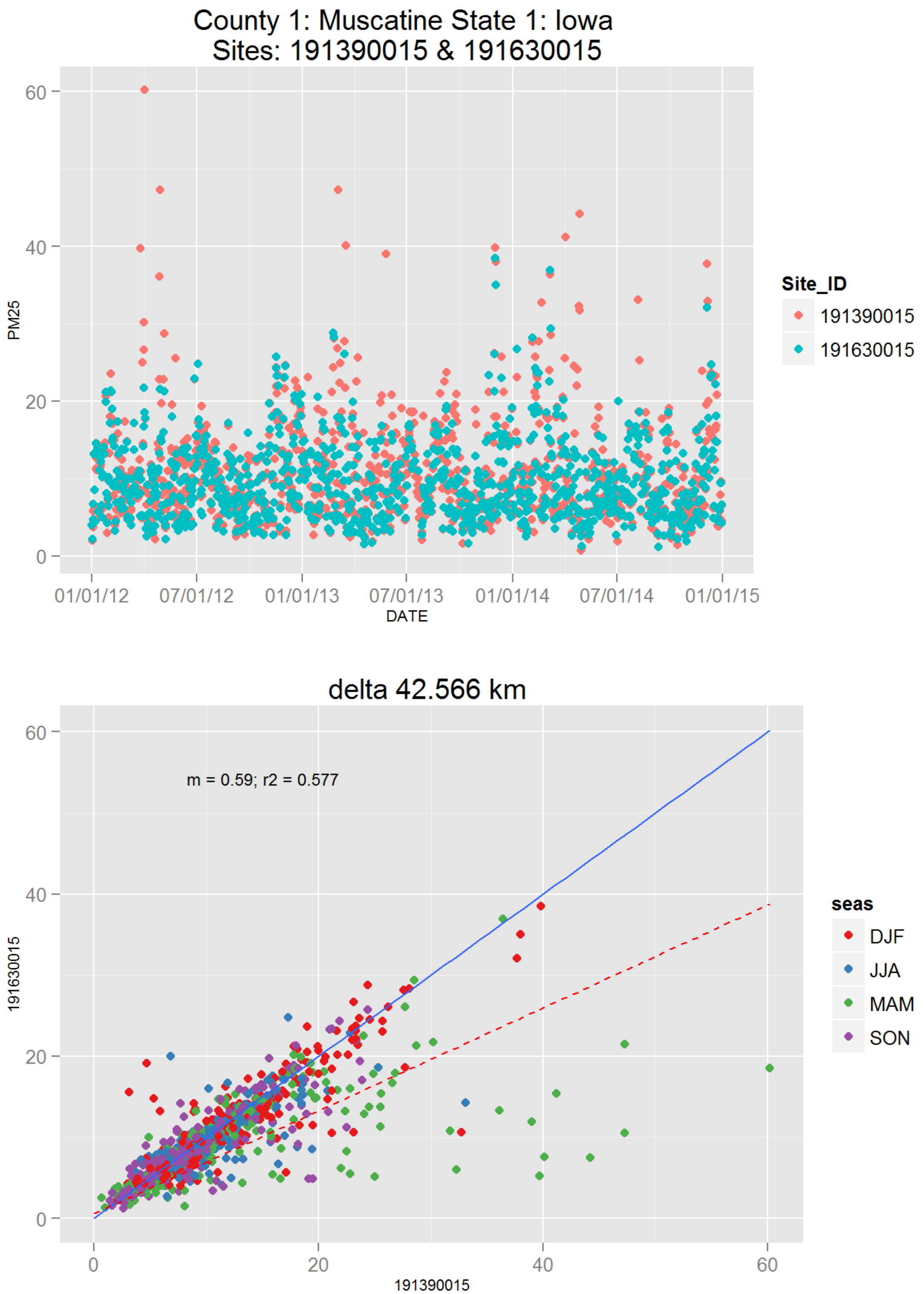


Figure 115: Comparison of $PM_{2.5}$ data for sites 191390015 and 191390015. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 116: Comparison of $PM_{2.5}$ data for sites 211110051 and 211110051. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 117: Comparison of $PM_{2.5}$ data for sites 220330009 and 220330009. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

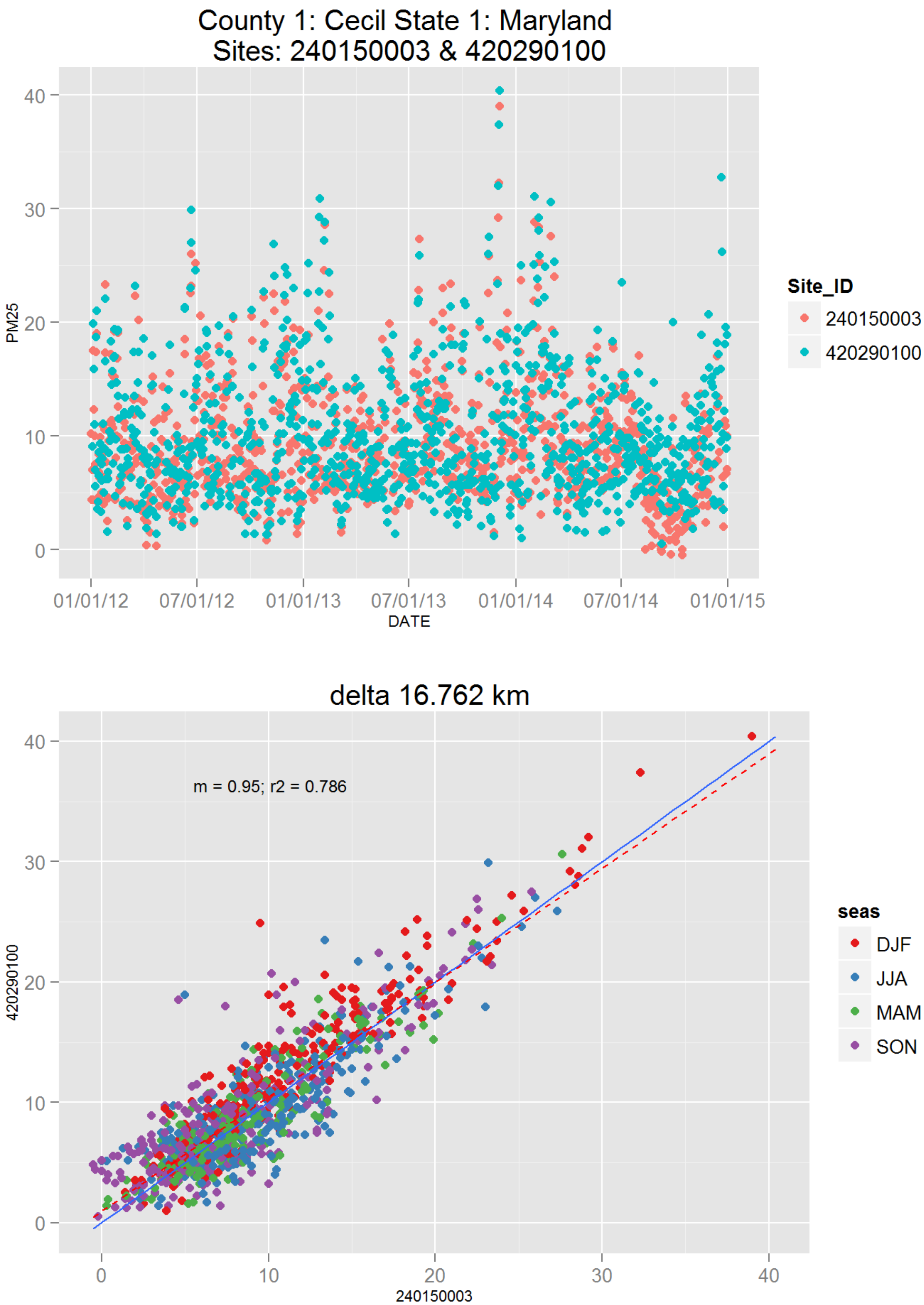
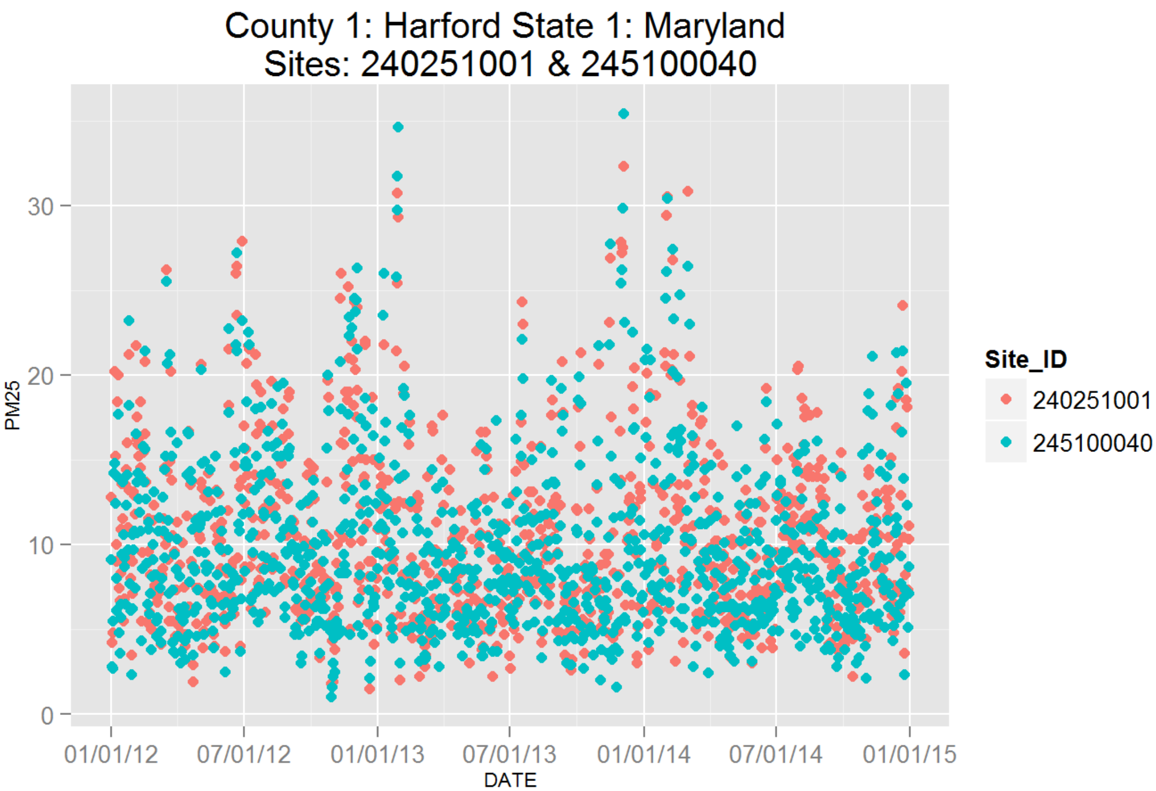


Figure 118: Comparison of $PM_{2.5}$ data for sites 240150003 and 240150003. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



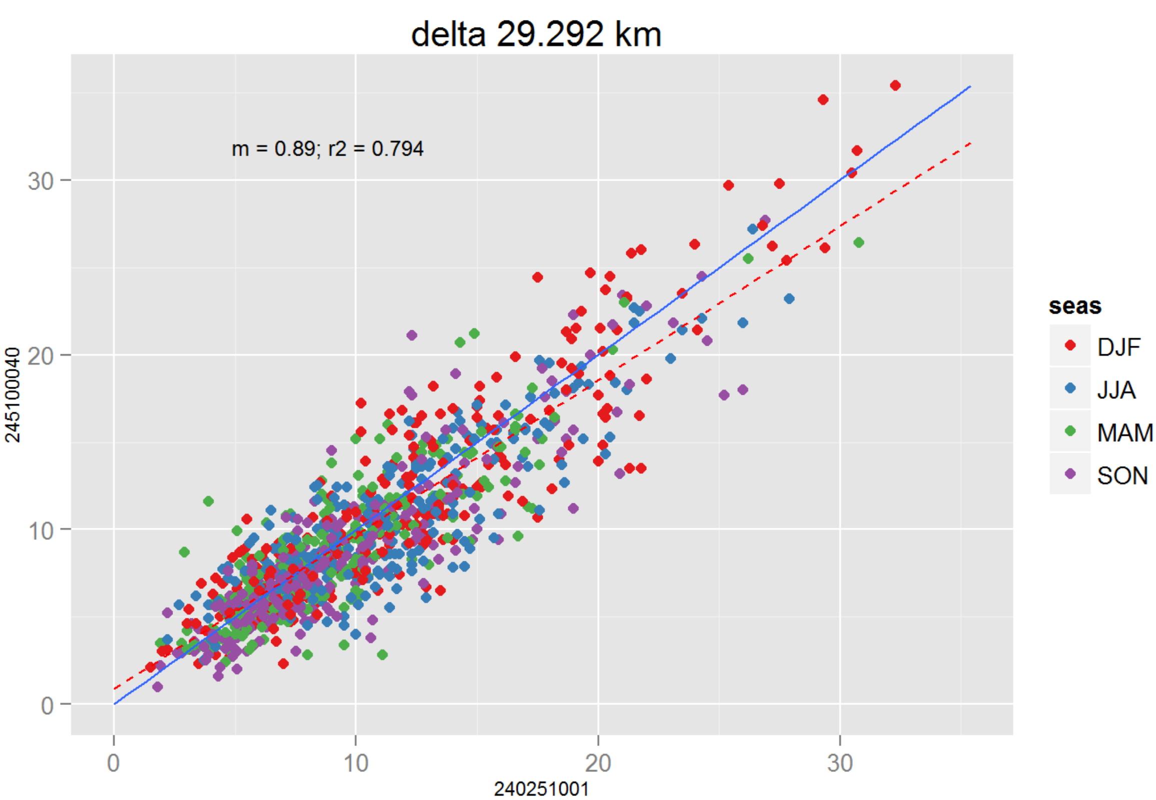


Figure 119: Comparison of $PM_{2.5}$ data for sites 240251001 and 240251001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 120: Comparison of $PM_{2.5}$ data for sites 240290002 and 240290002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

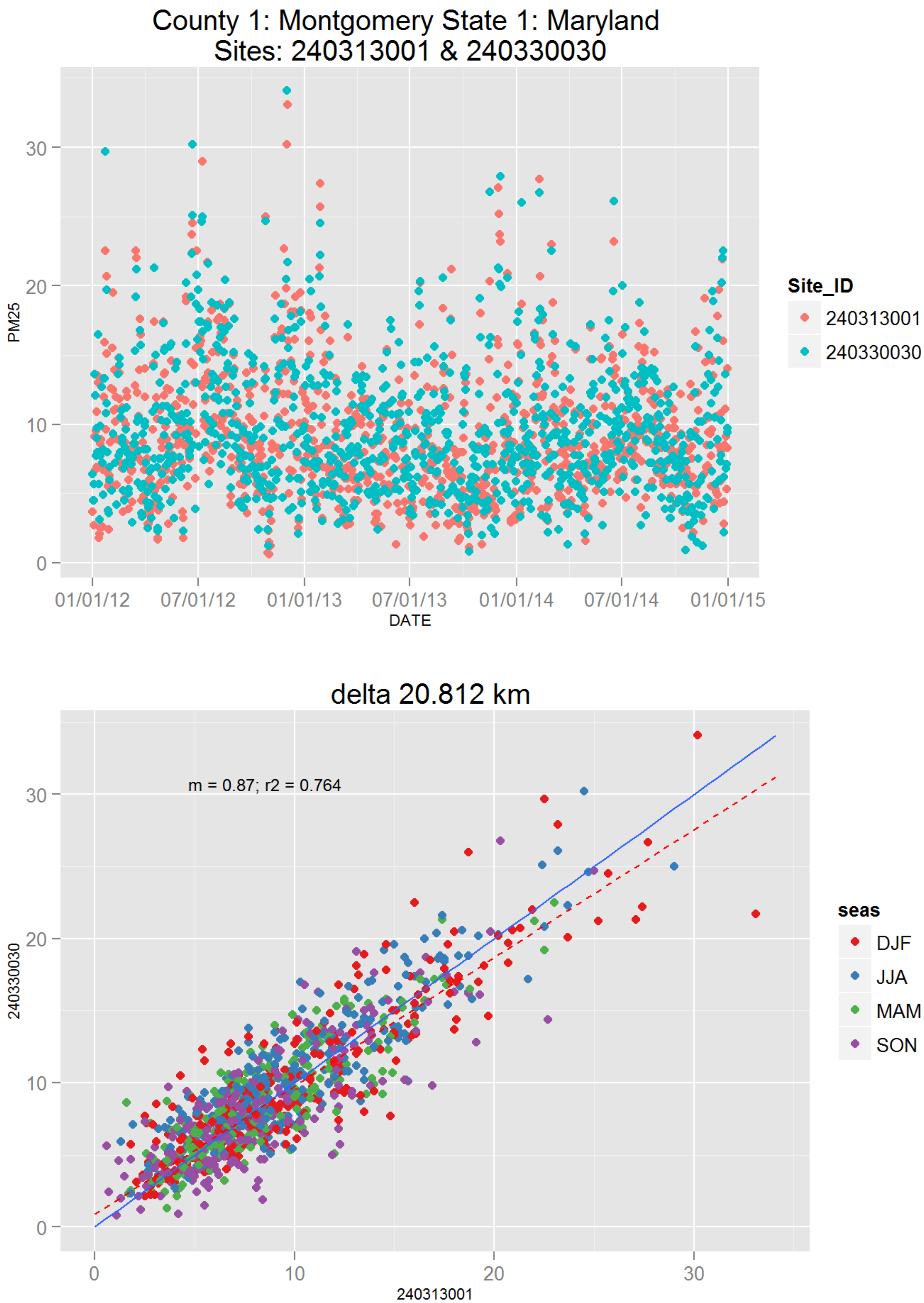
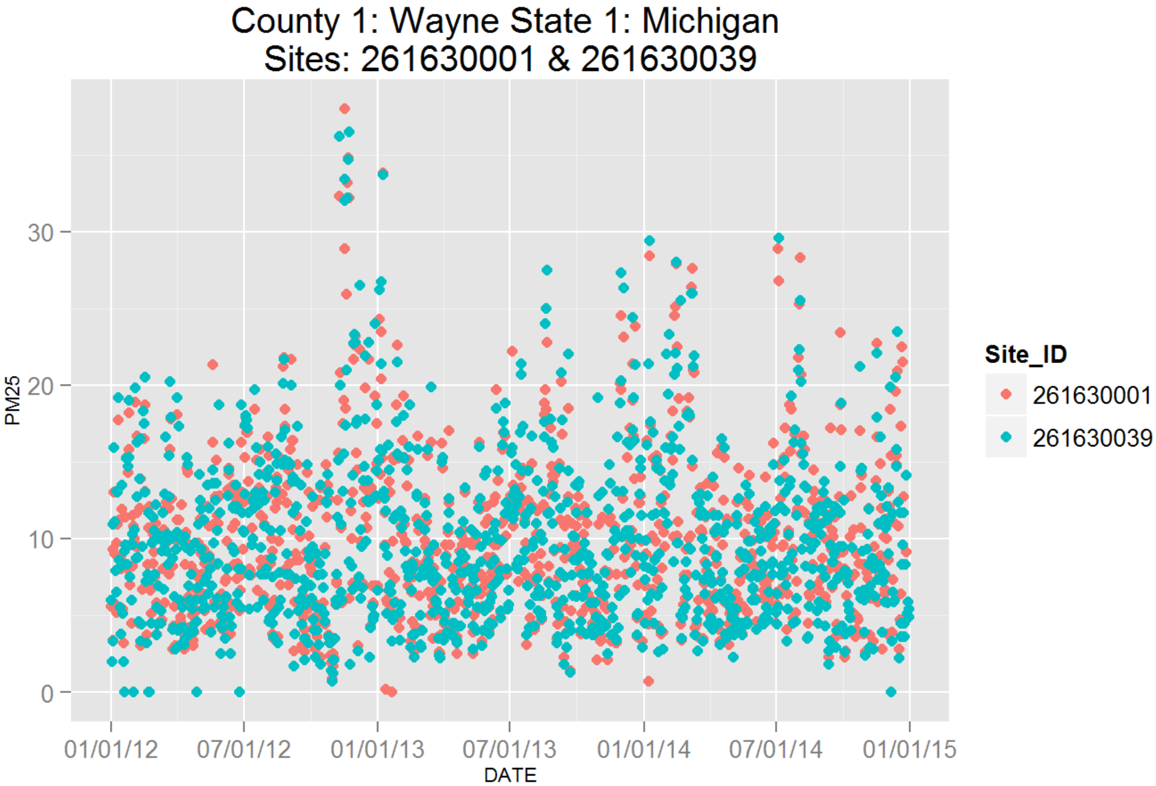


Figure 121: Comparison of $PM_{2.5}$ data for sites 240313001 and 240313001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 122: Comparison of $PM_{2.5}$ data for sites 240330030 and 240330030. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



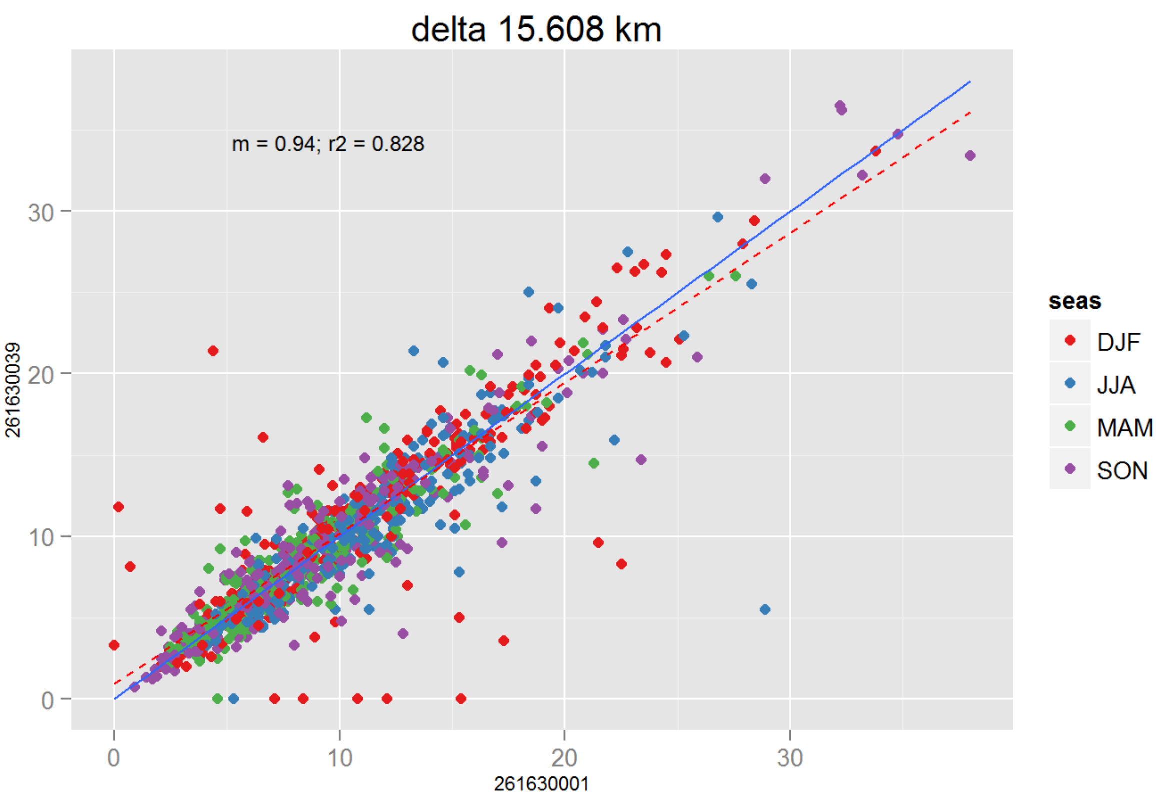


Figure 123: Comparison of $PM_{2.5}$ data for sites 261630001 and 261630001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 124: Comparison of $PM_{2.5}$ data for sites 270031002 and 270031002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 125: Comparison of $PM_{2.5}$ data for sites 270530963 and 270530963. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

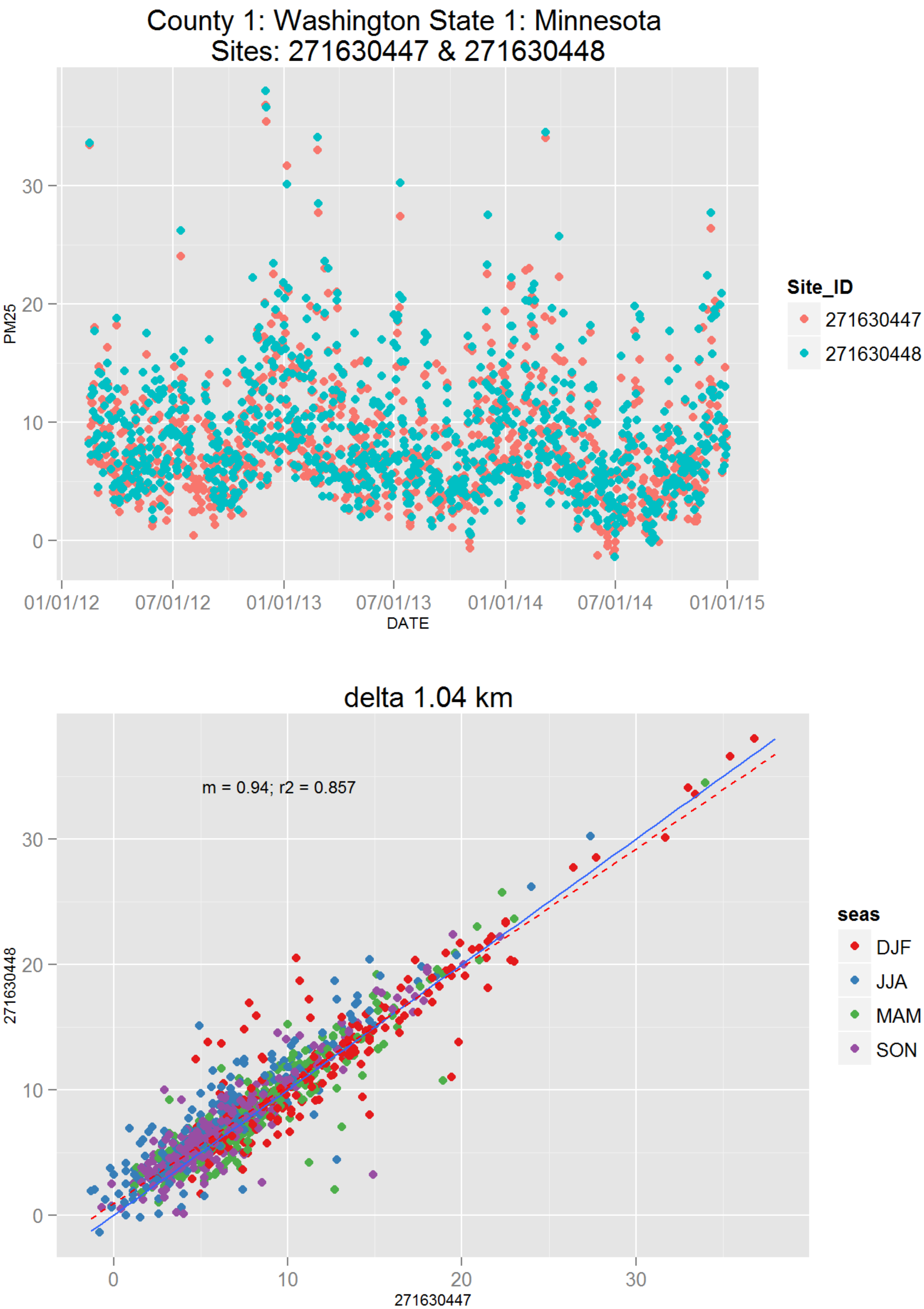


Figure 126: Comparison of $PM_{2.5}$ data for sites 271630447 and 271630447. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 127: Comparison of $PM_{2.5}$ data for sites 290370003 and 290370003. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

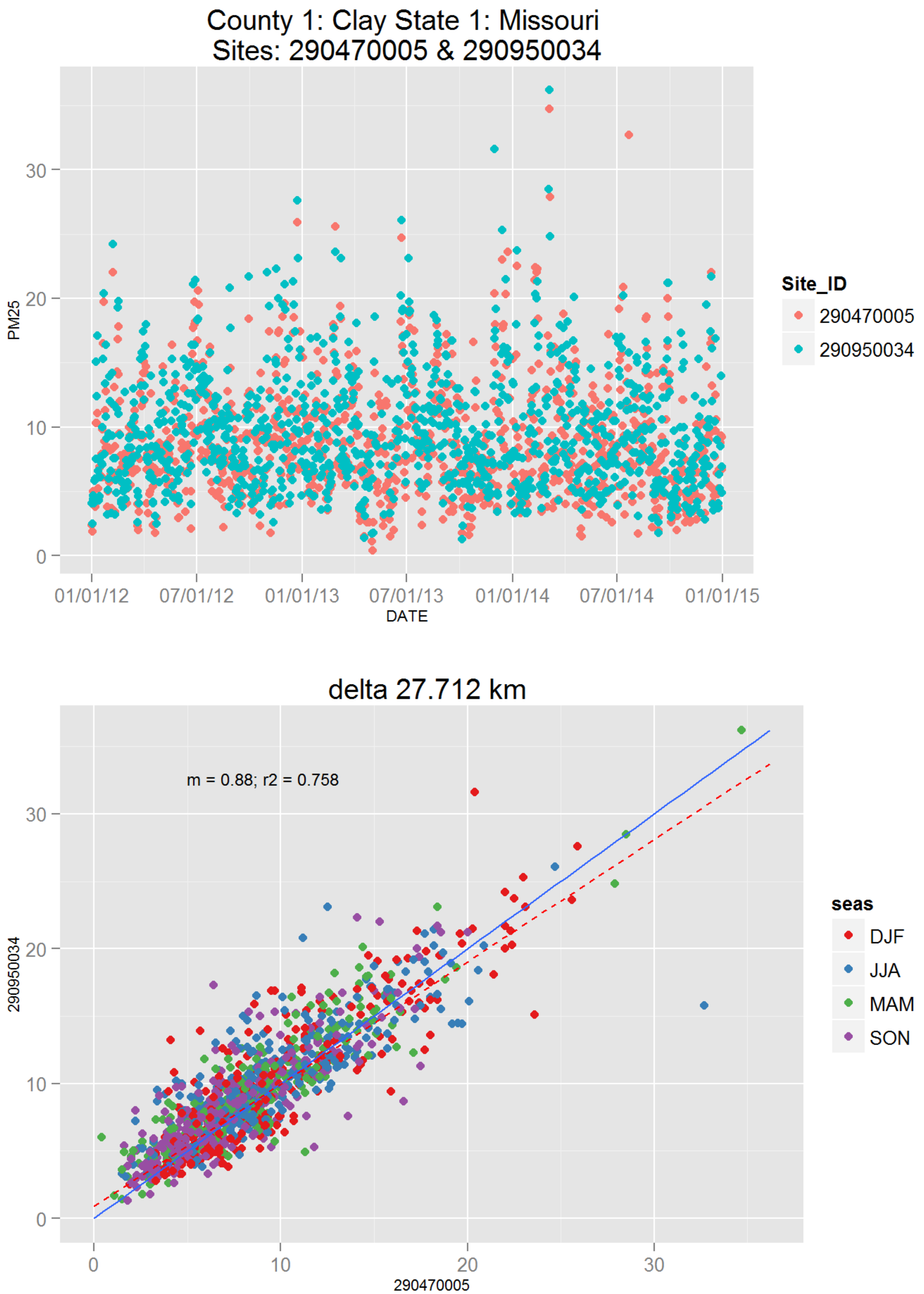
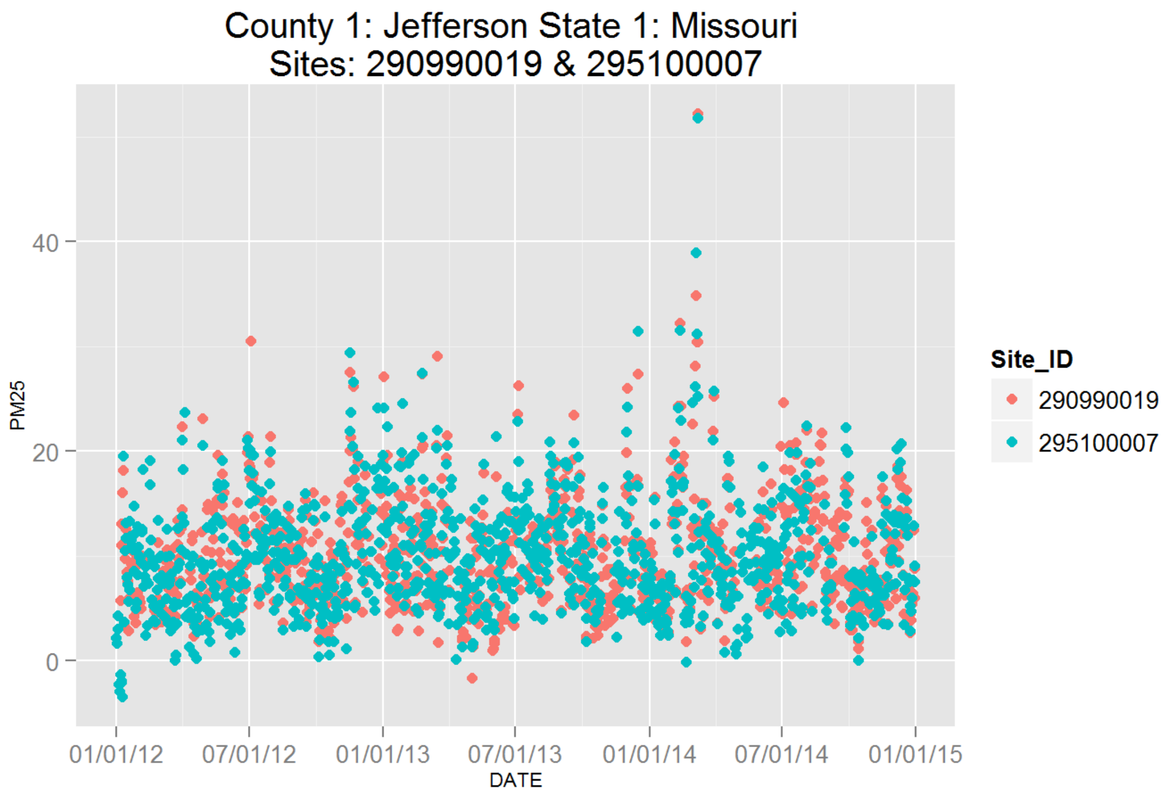


Figure 128: Comparison of $PM_{2.5}$ data for sites 290470005 and 290470005. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



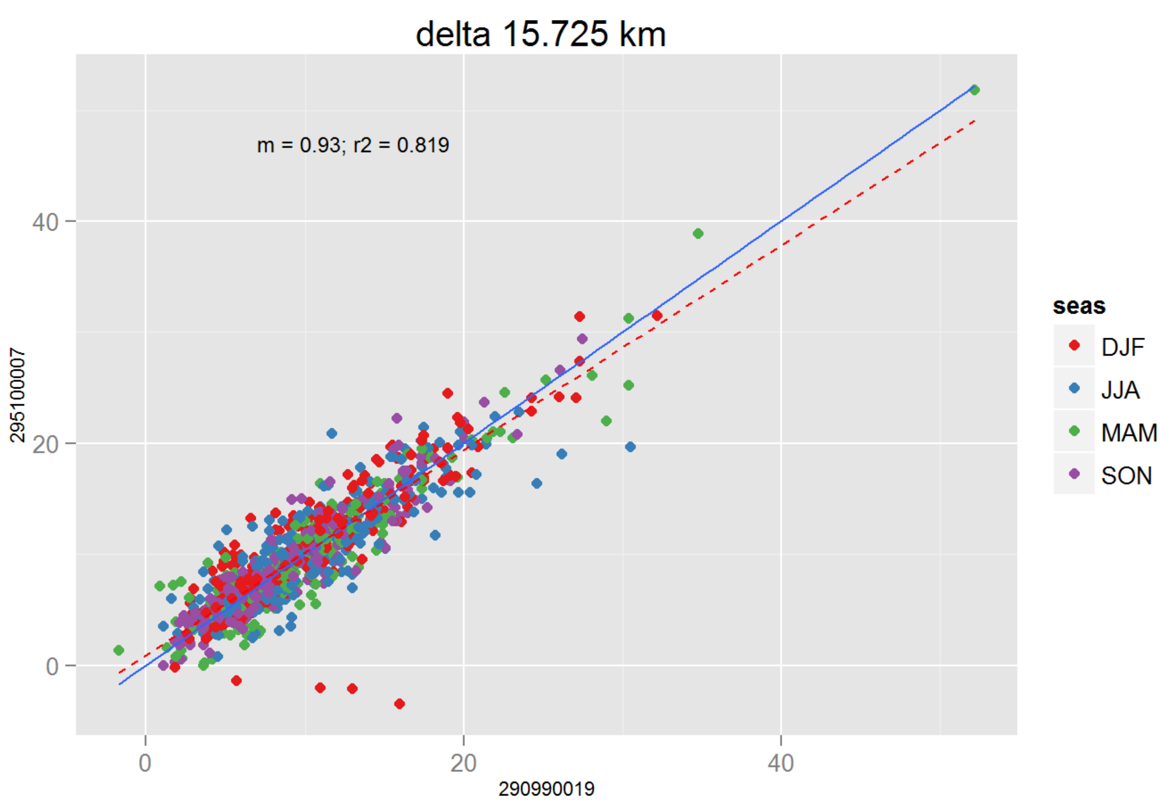
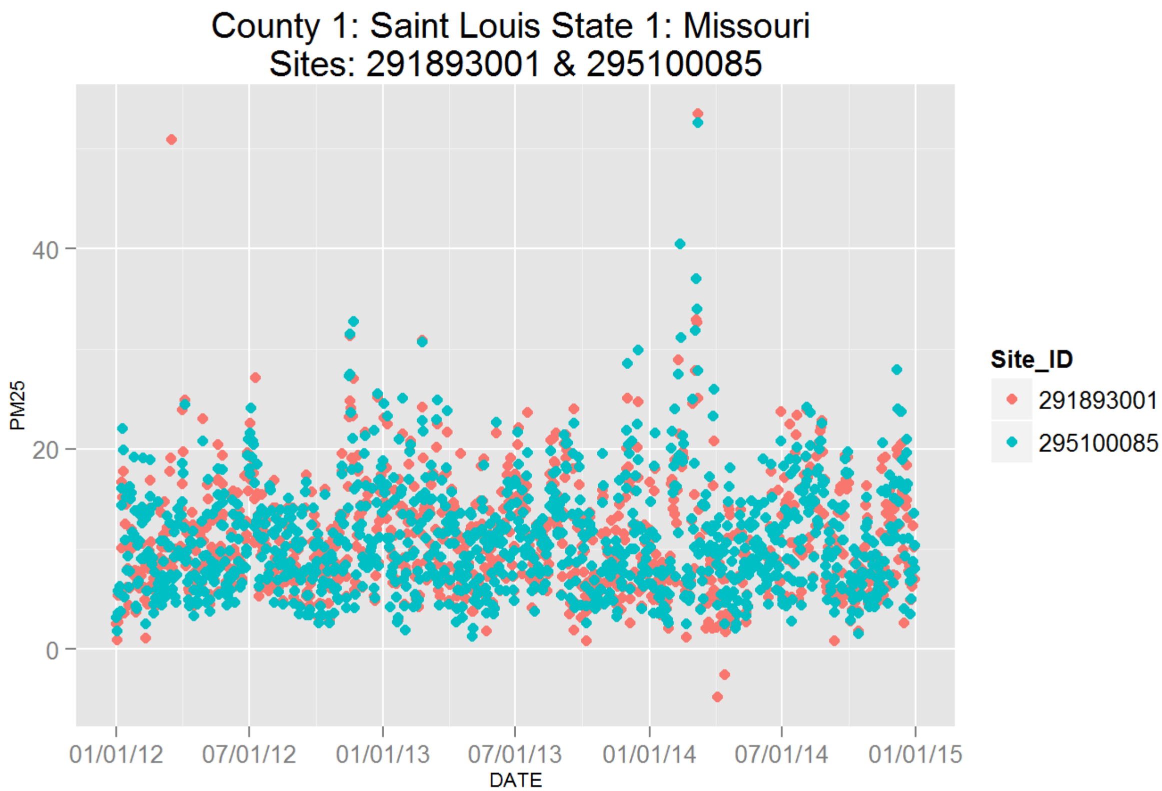


Figure 129: Comparison of $PM_{2.5}$ data for sites 290990019 and 290990019. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



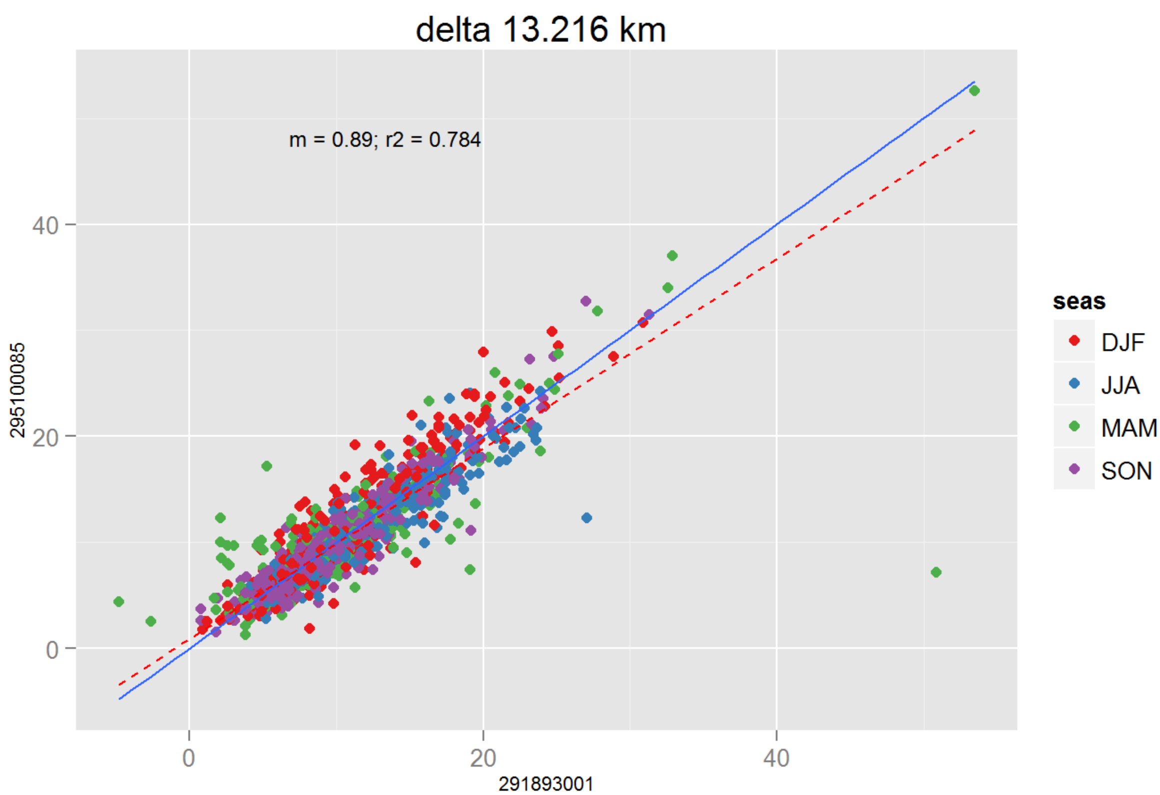


Figure 130: Comparison of $PM_{2.5}$ data for sites 291893001 and 291893001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

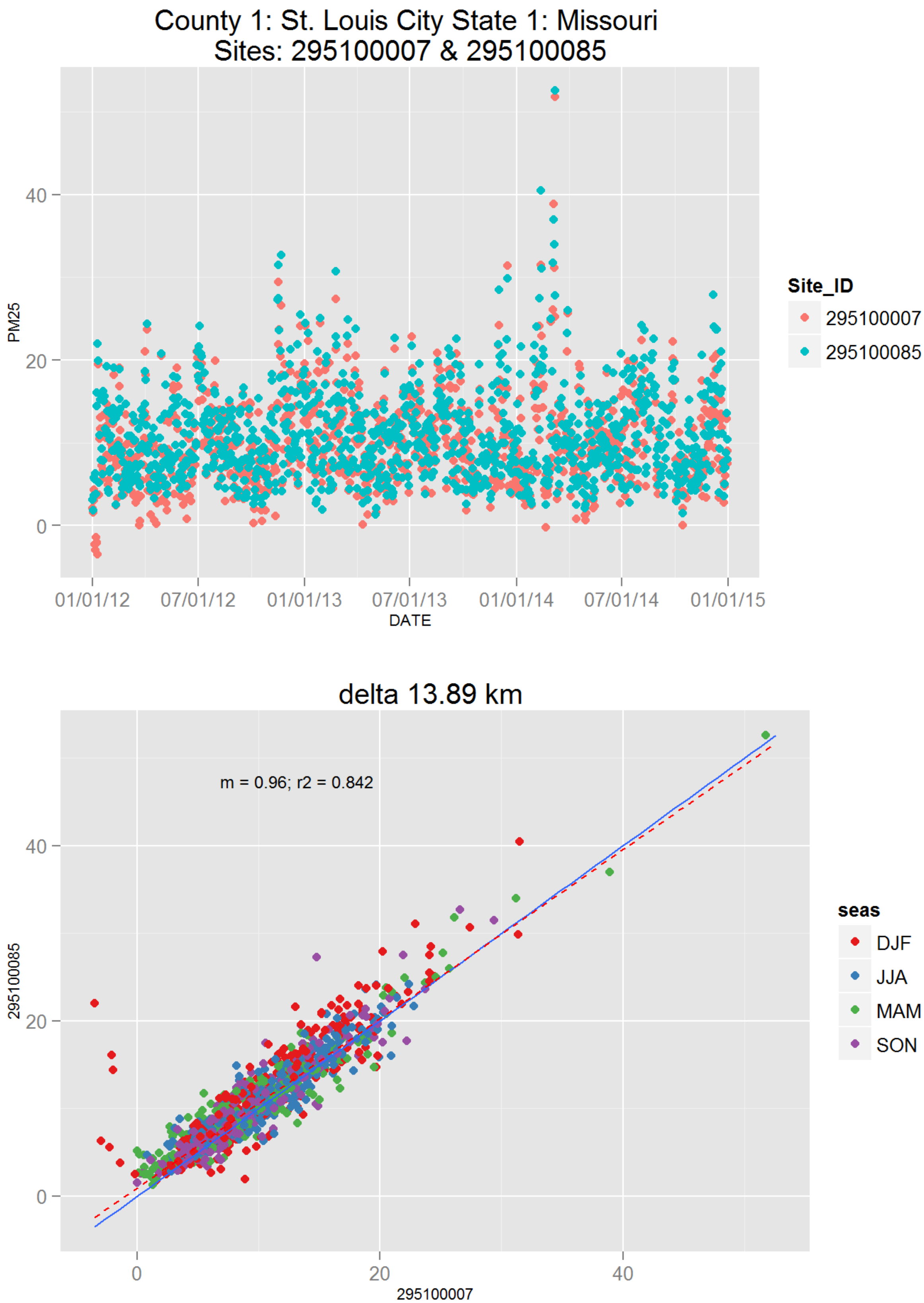
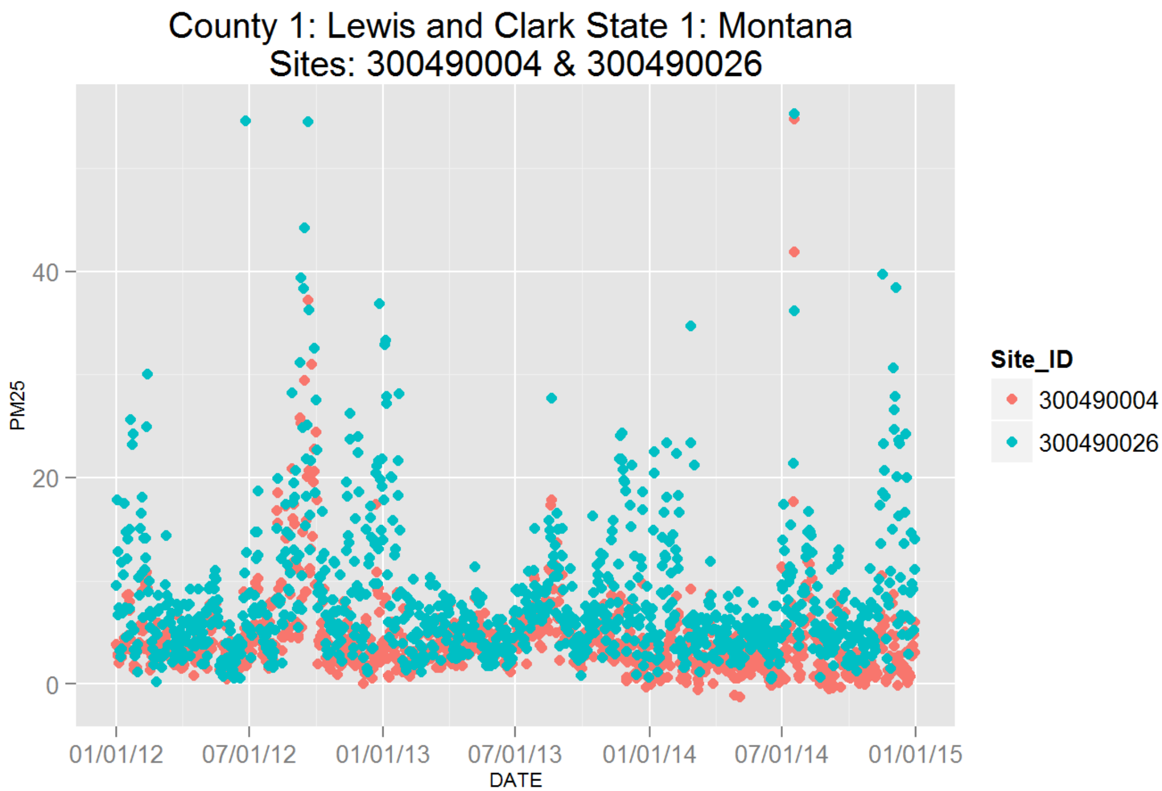


Figure 131: Comparison of $PM_{2.5}$ data for sites 295100007 and 295100007. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



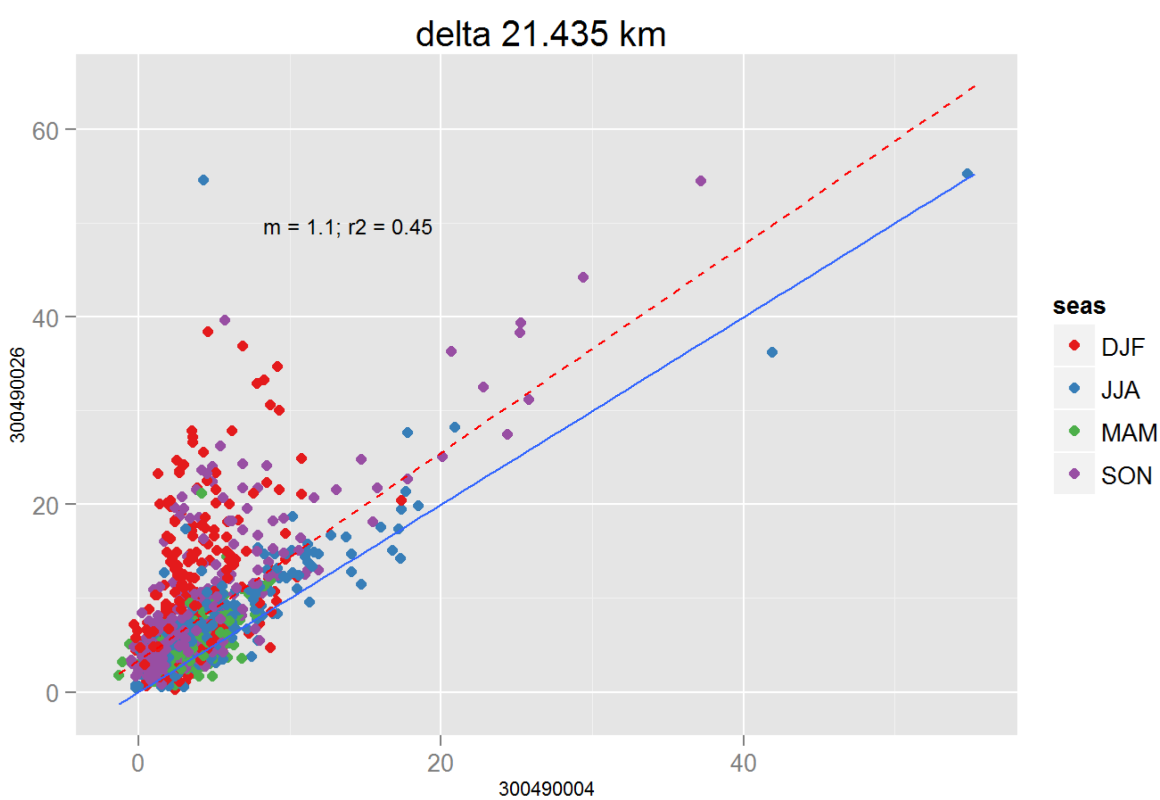


Figure 132: Comparison of $PM_{2.5}$ data for sites 300490004 and 300490004. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

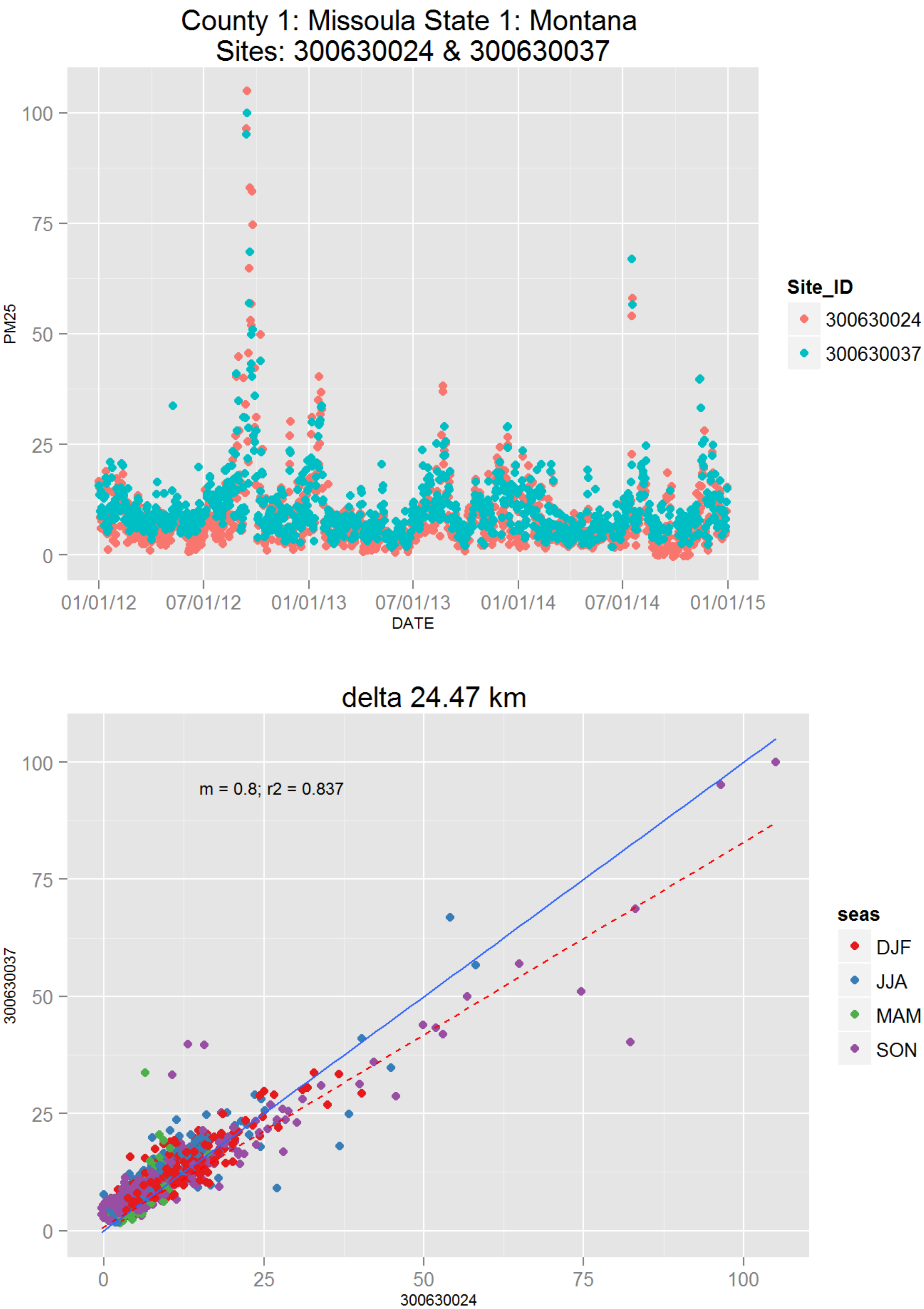


Figure 133: Comparison of $PM_{2.5}$ data for sites 300630024 and 300630024. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

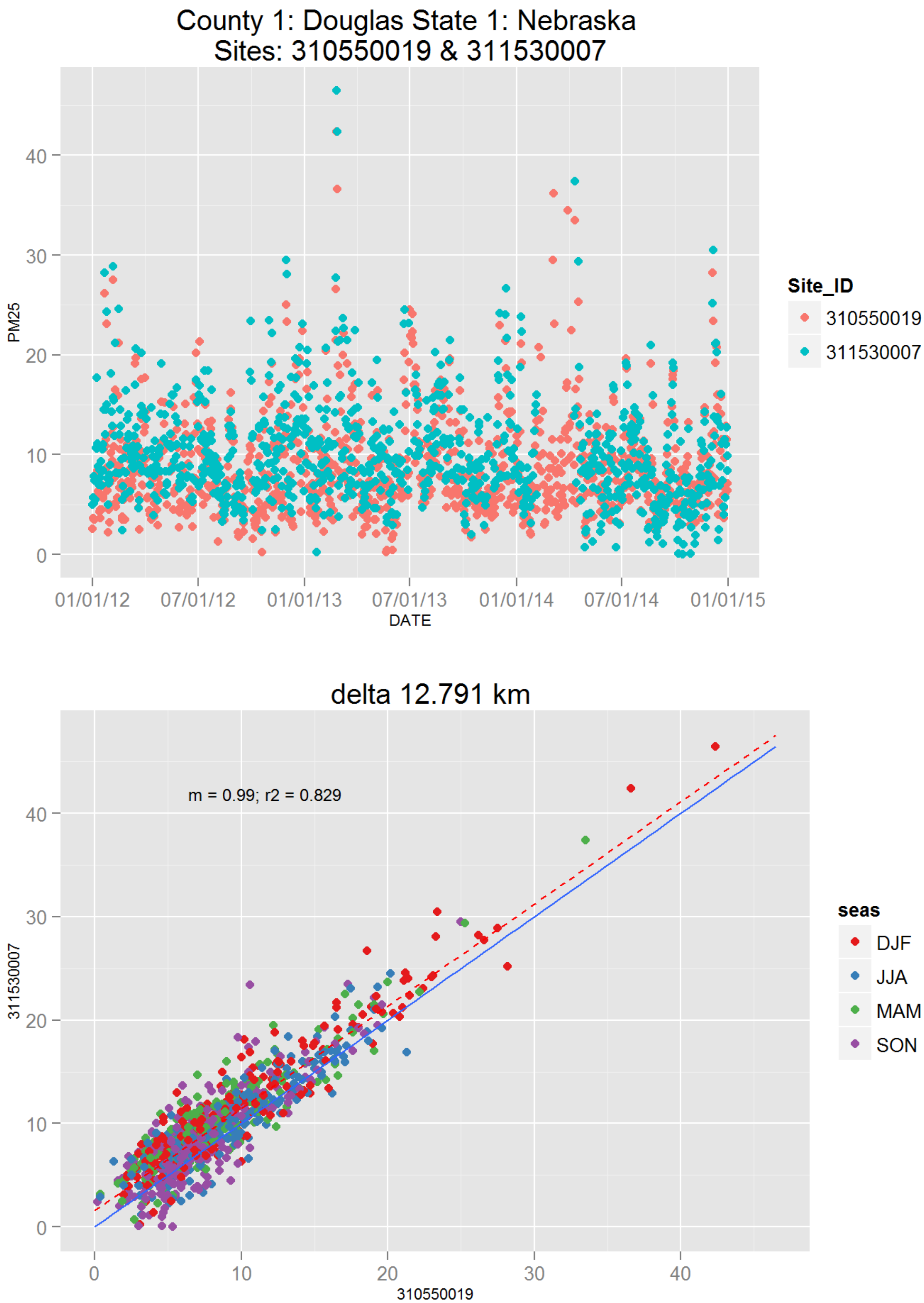


Figure 134: Comparison of $PM_{2.5}$ data for sites 310550019 and 310550019. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

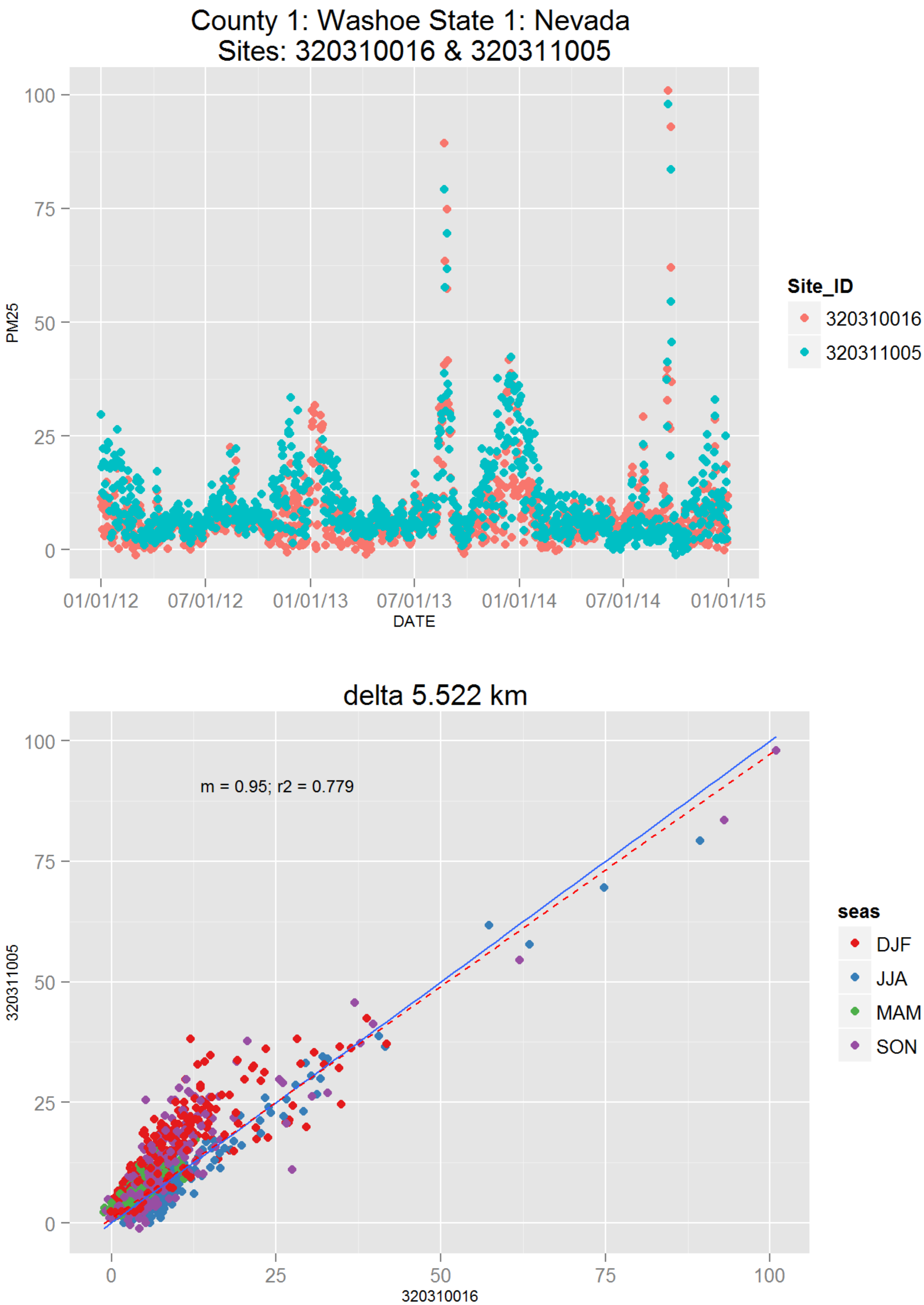


Figure 135: Comparison of $PM_{2.5}$ data for sites 320310016 and 320310016. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 136: Comparison of $PM_{2.5}$ data for sites 330050007 and 330050007. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 137: Comparison of $PM_{2.5}$ data for sites 330150018 and 330150018. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

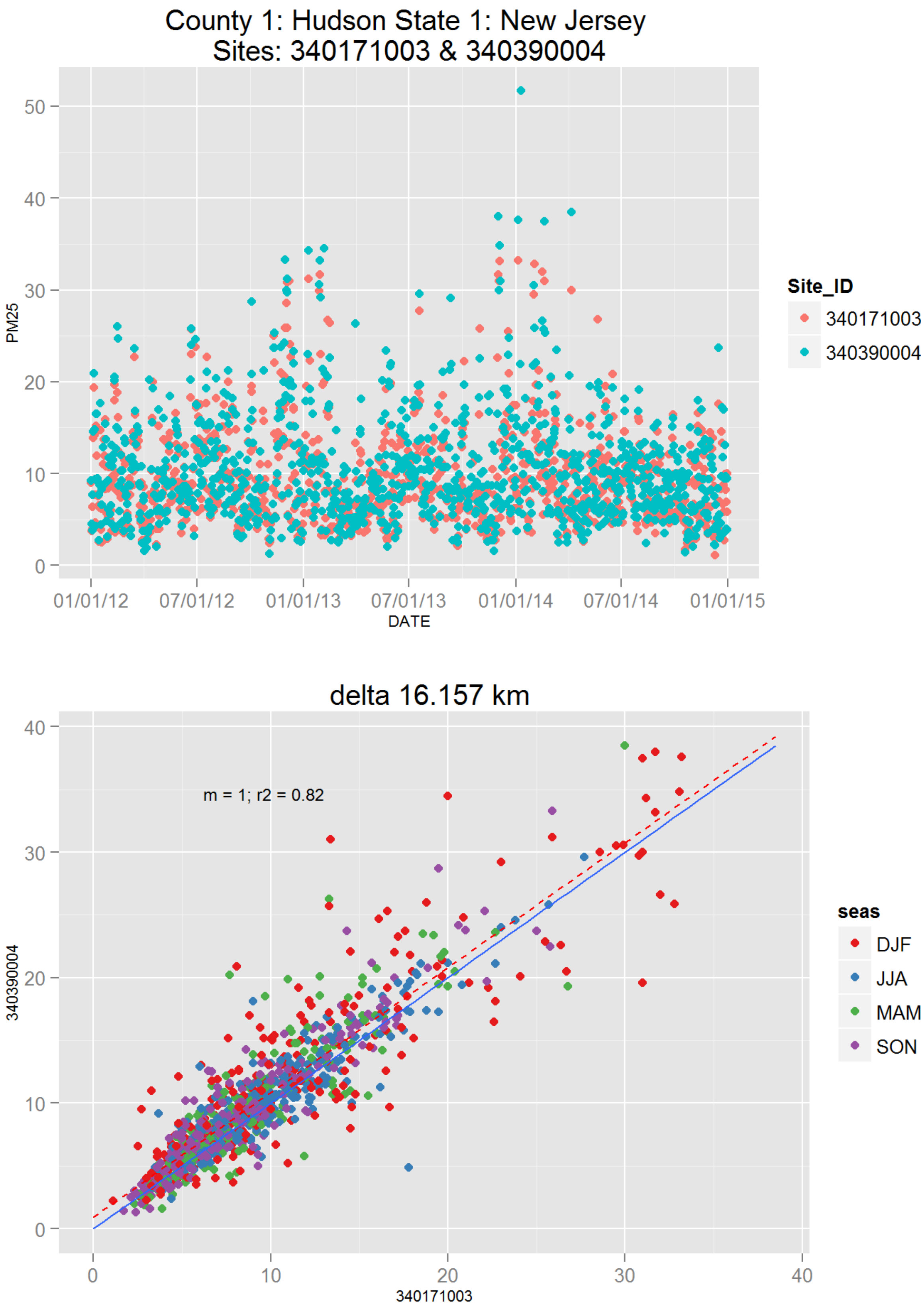
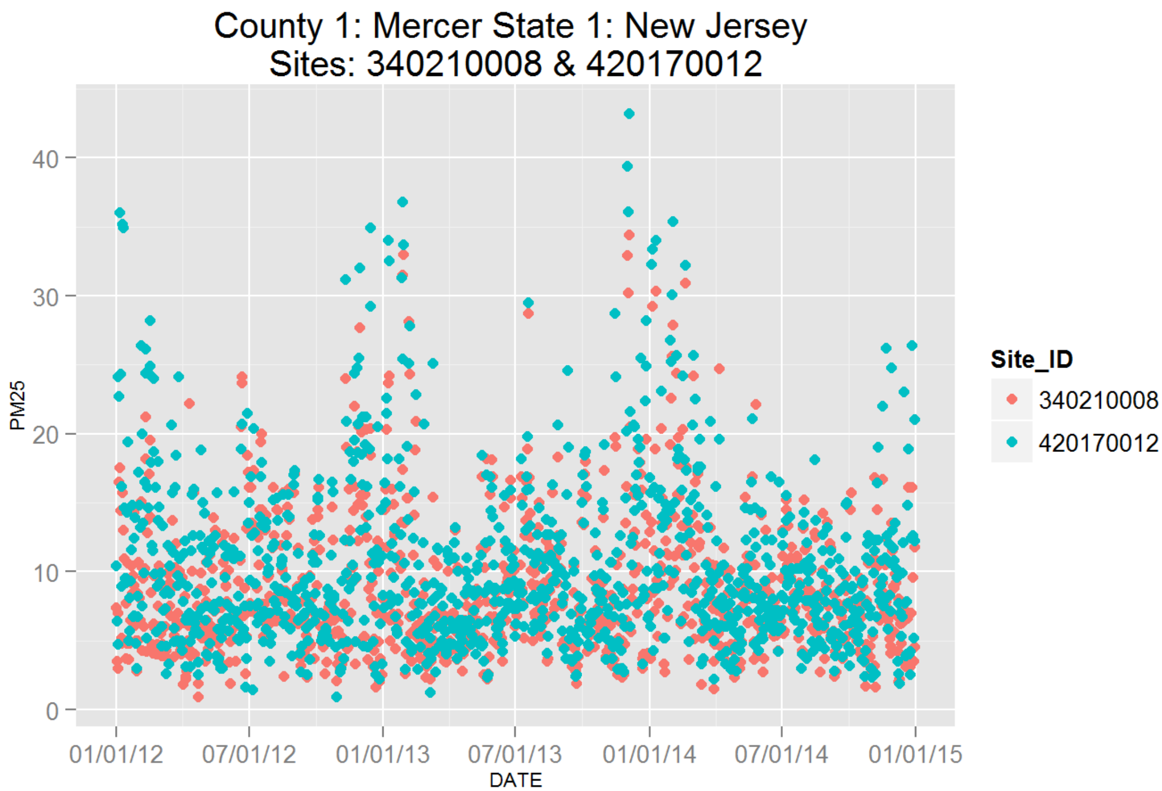


Figure 138: Comparison of $PM_{2.5}$ data for sites 340171003 and 340171003. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



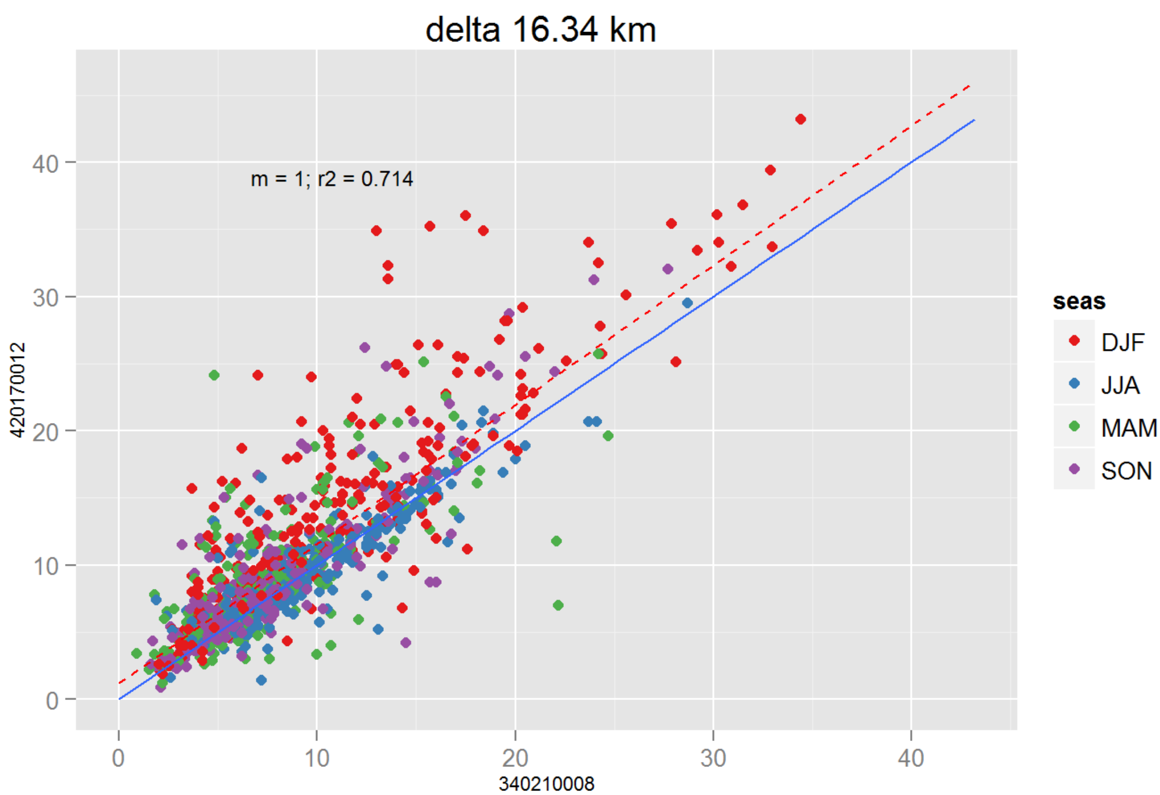


Figure 139: Comparison of $PM_{2.5}$ data for sites 340210008 and 340210008. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

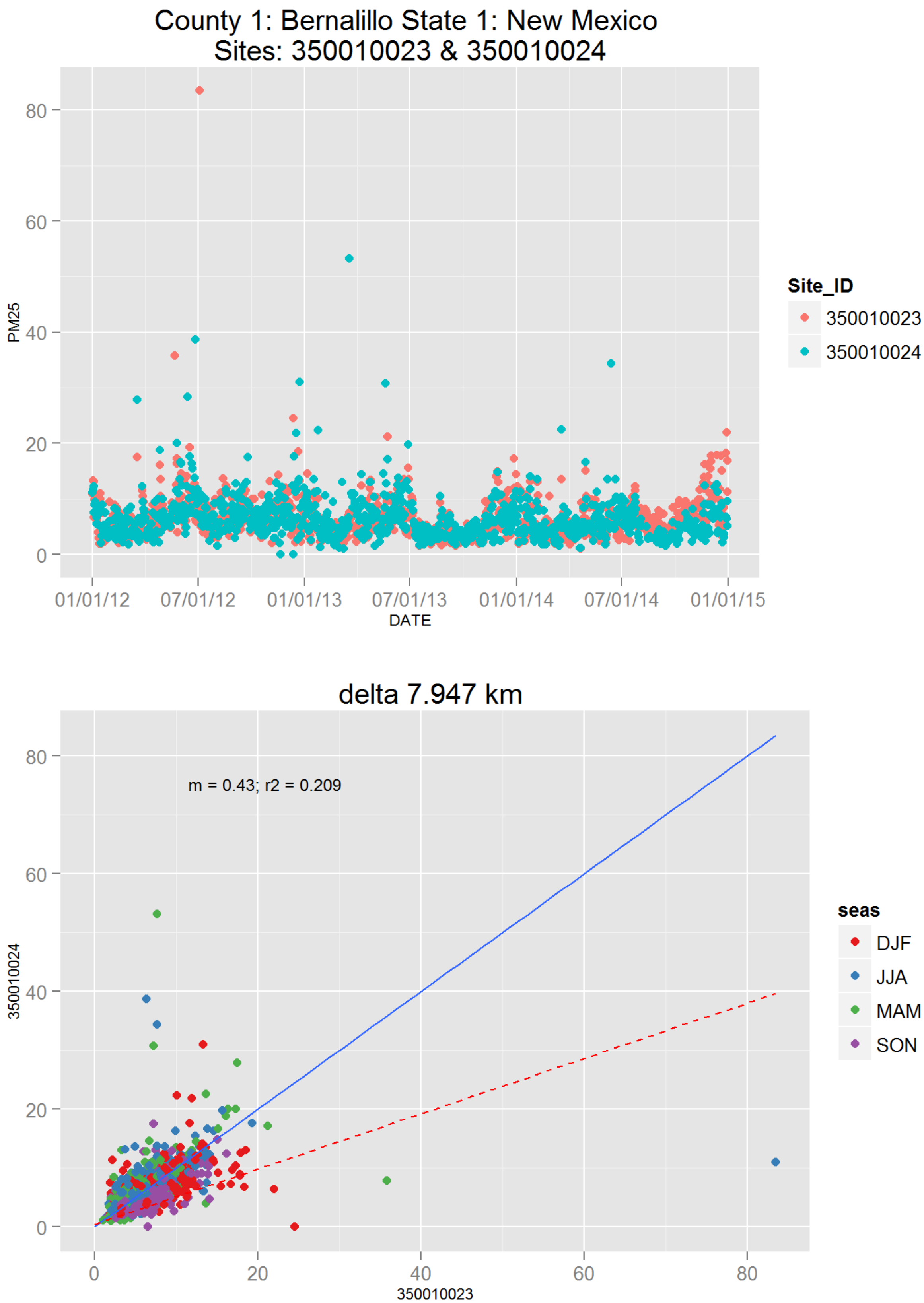


Figure 140: Comparison of $PM_{2.5}$ data for sites 350010023 and 350010023. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 141: Comparison of $PM_{2.5}$ data for sites 360810124 and 360810124. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

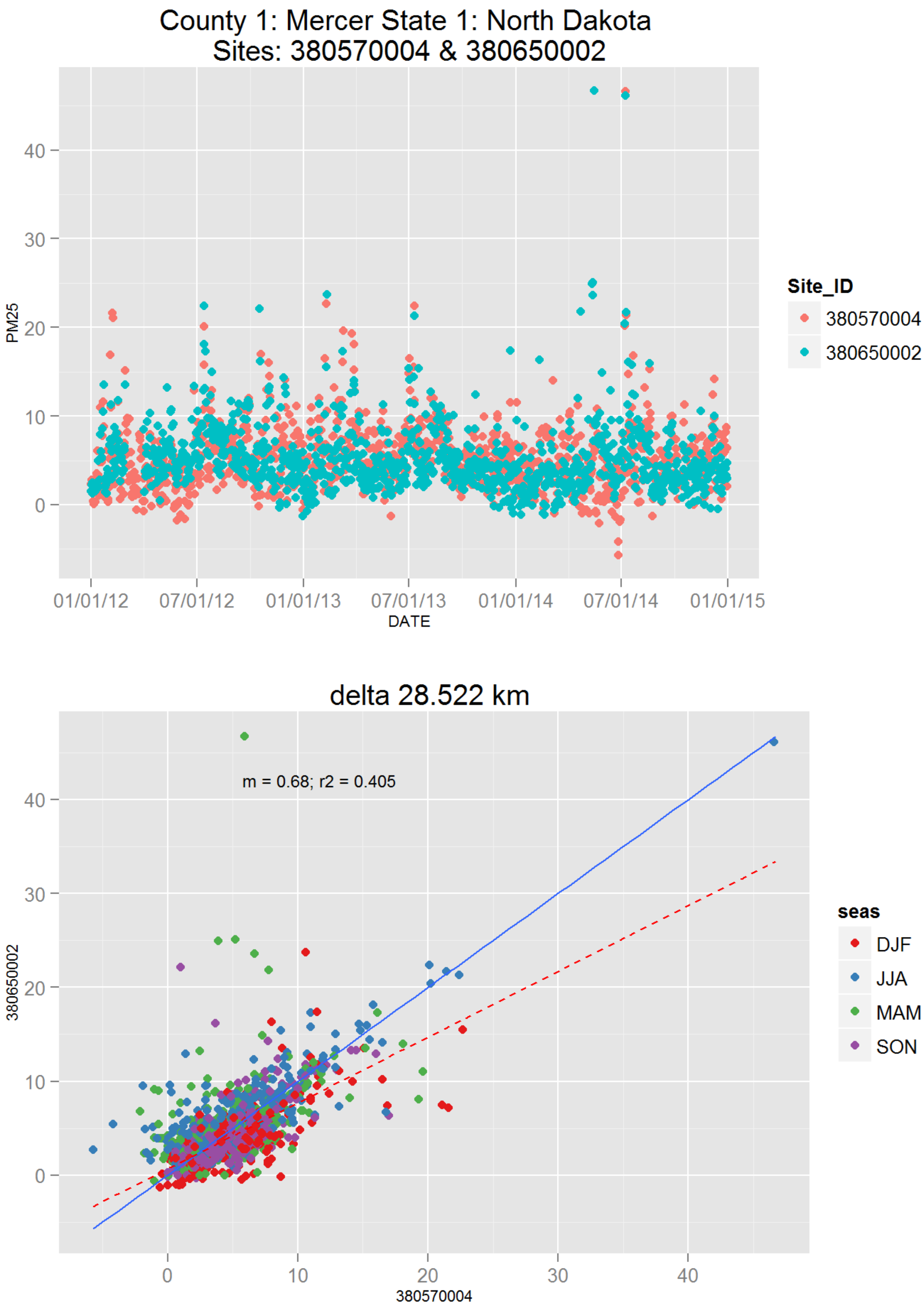
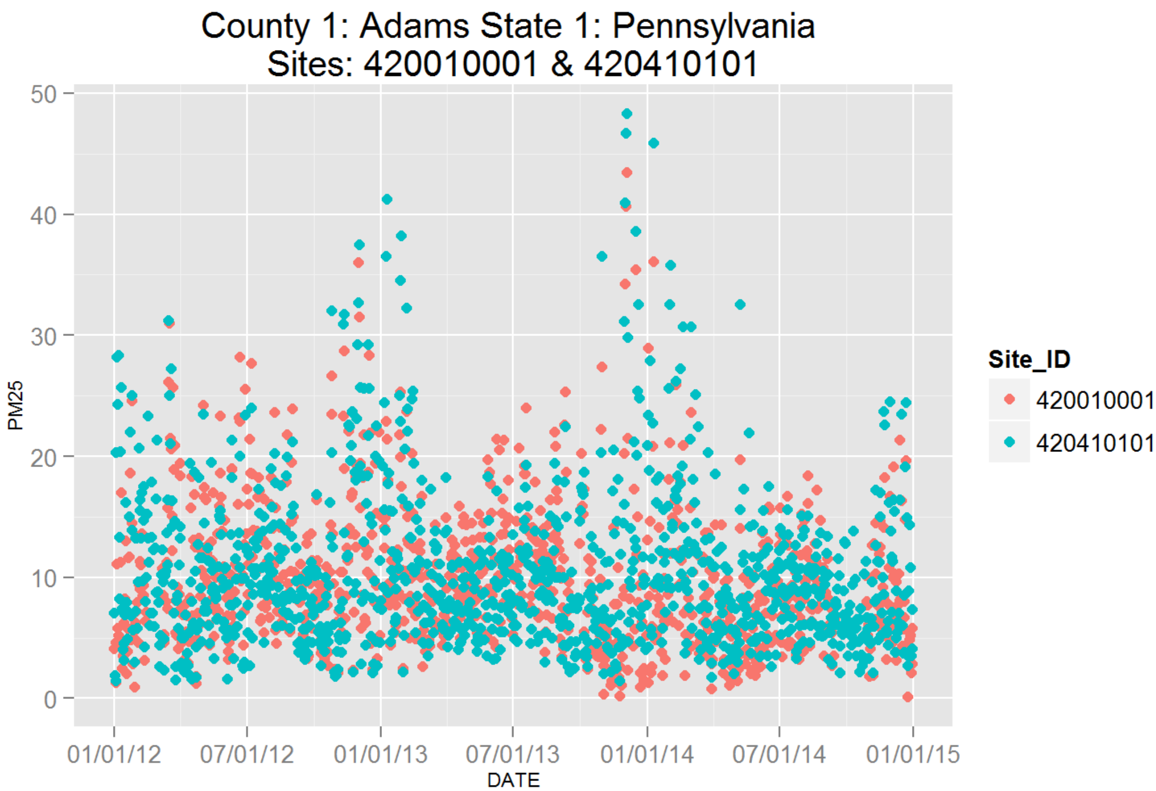


Figure 142: Comparison of $PM_{2.5}$ data for sites 380570004 and 380570004. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



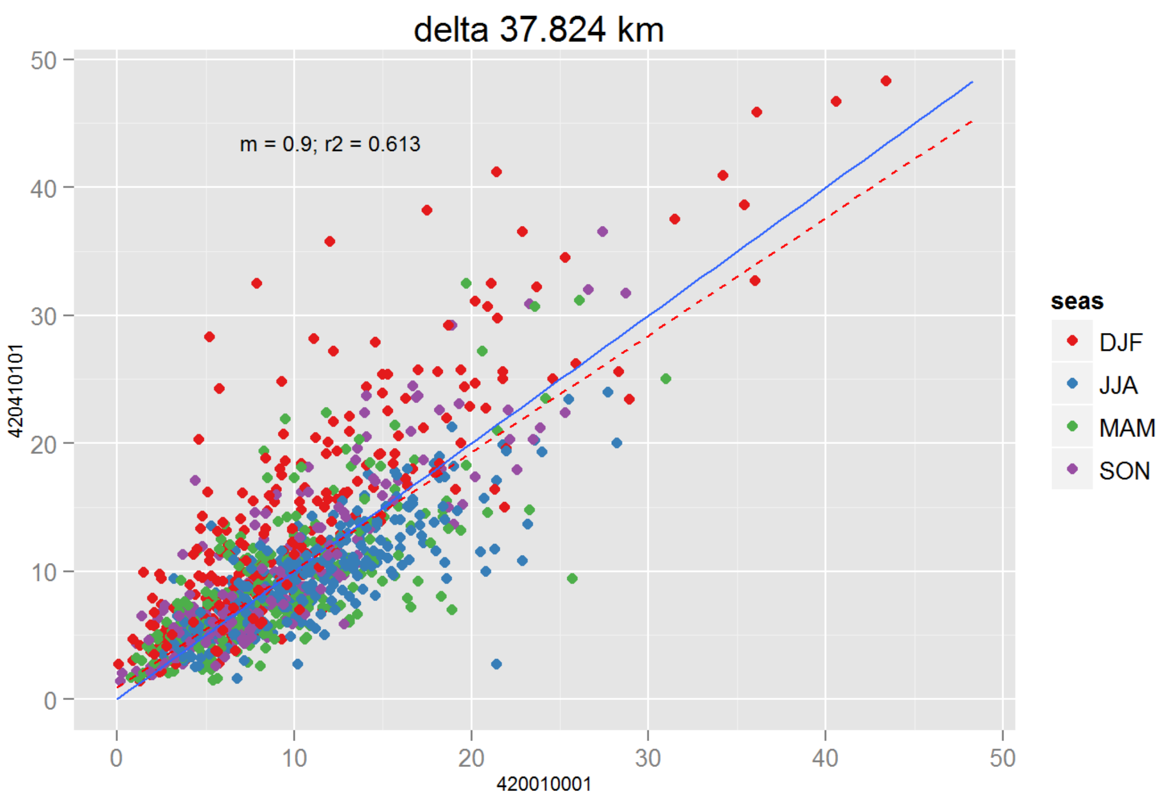
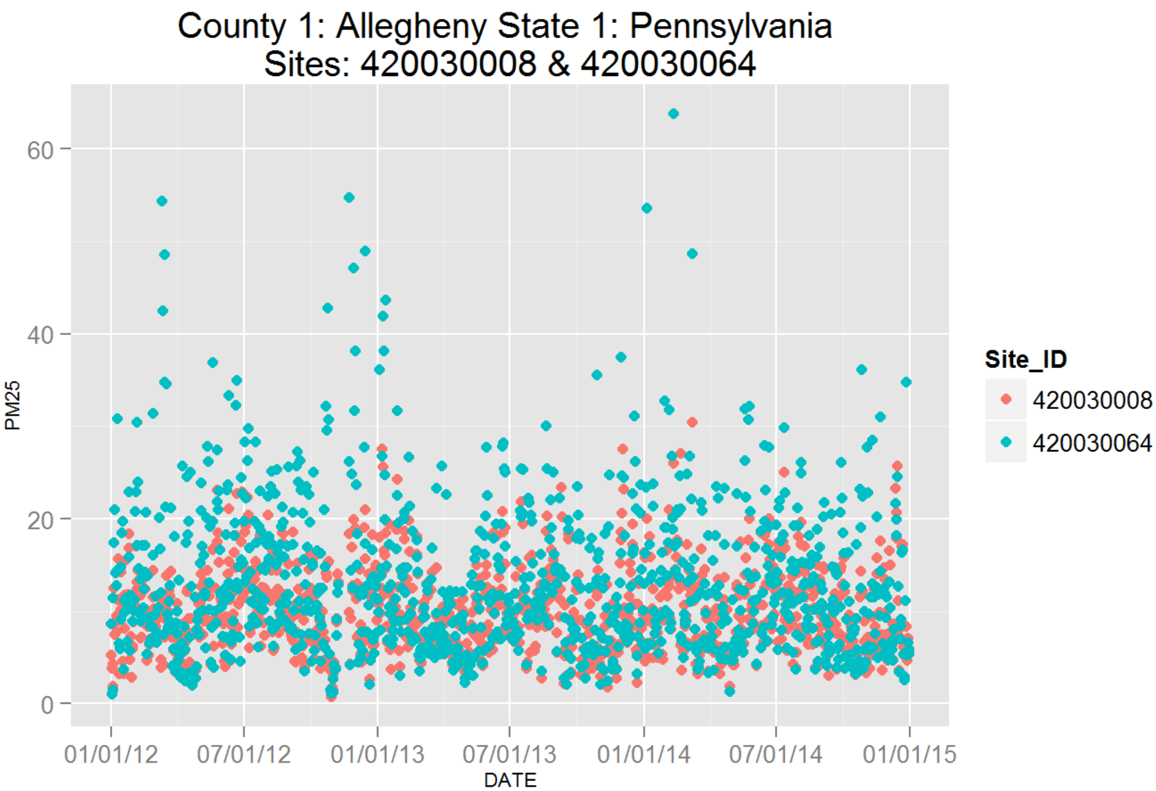


Figure 143: Comparison of $PM_{2.5}$ data for sites 420010001 and 420010001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



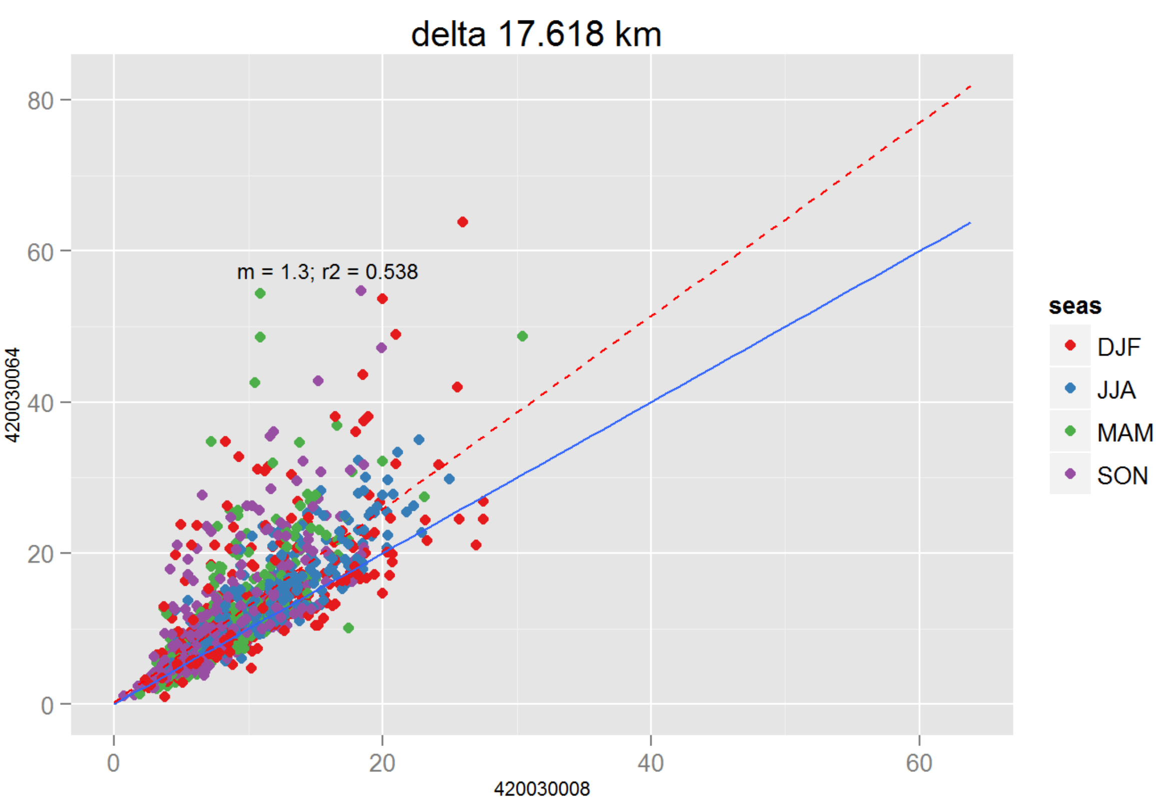


Figure 144: Comparison of $PM_{2.5}$ data for sites 420030008 and 420030008. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

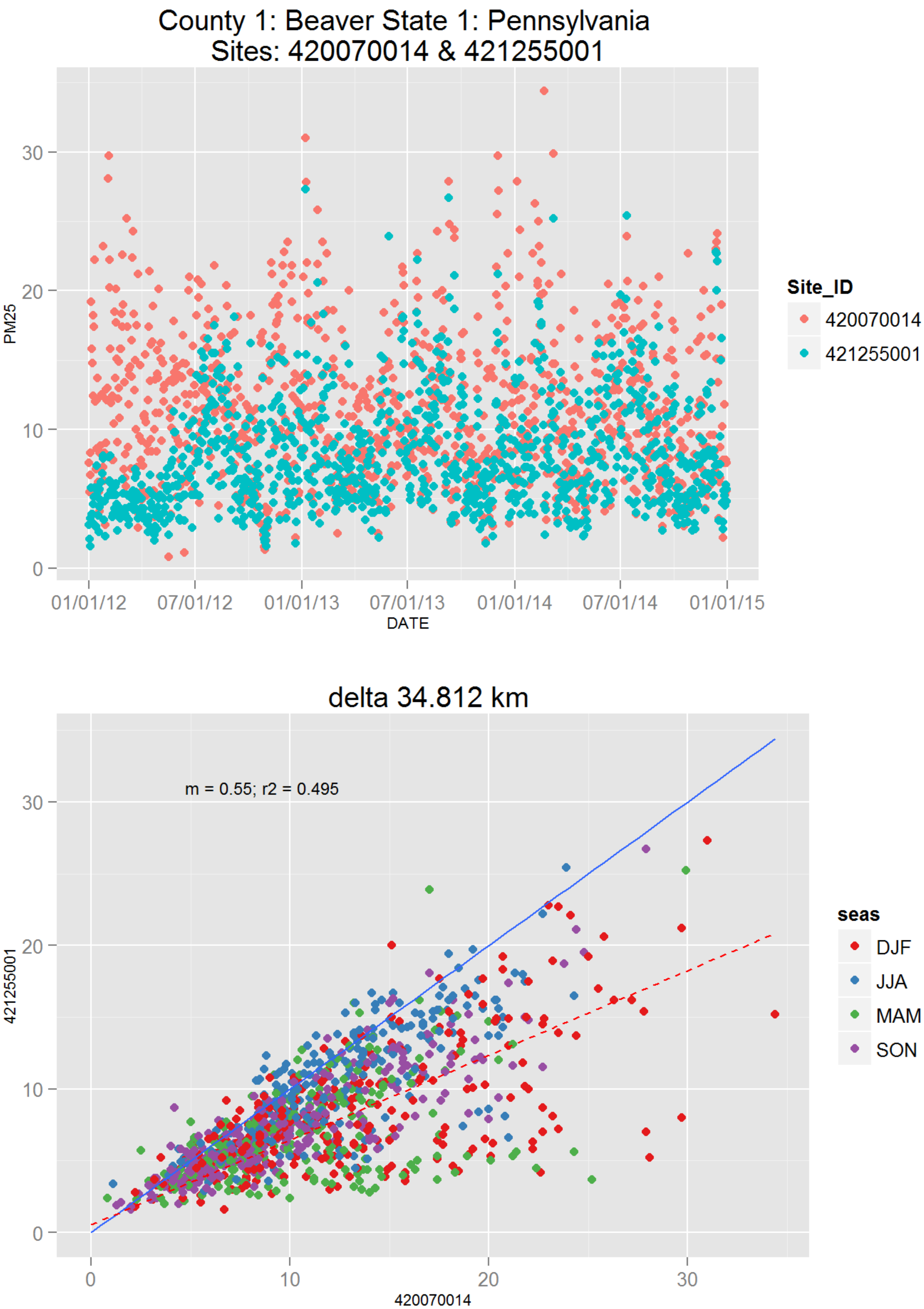


Figure 145: Comparison of $PM_{2.5}$ data for sites 420070014 and 420070014. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

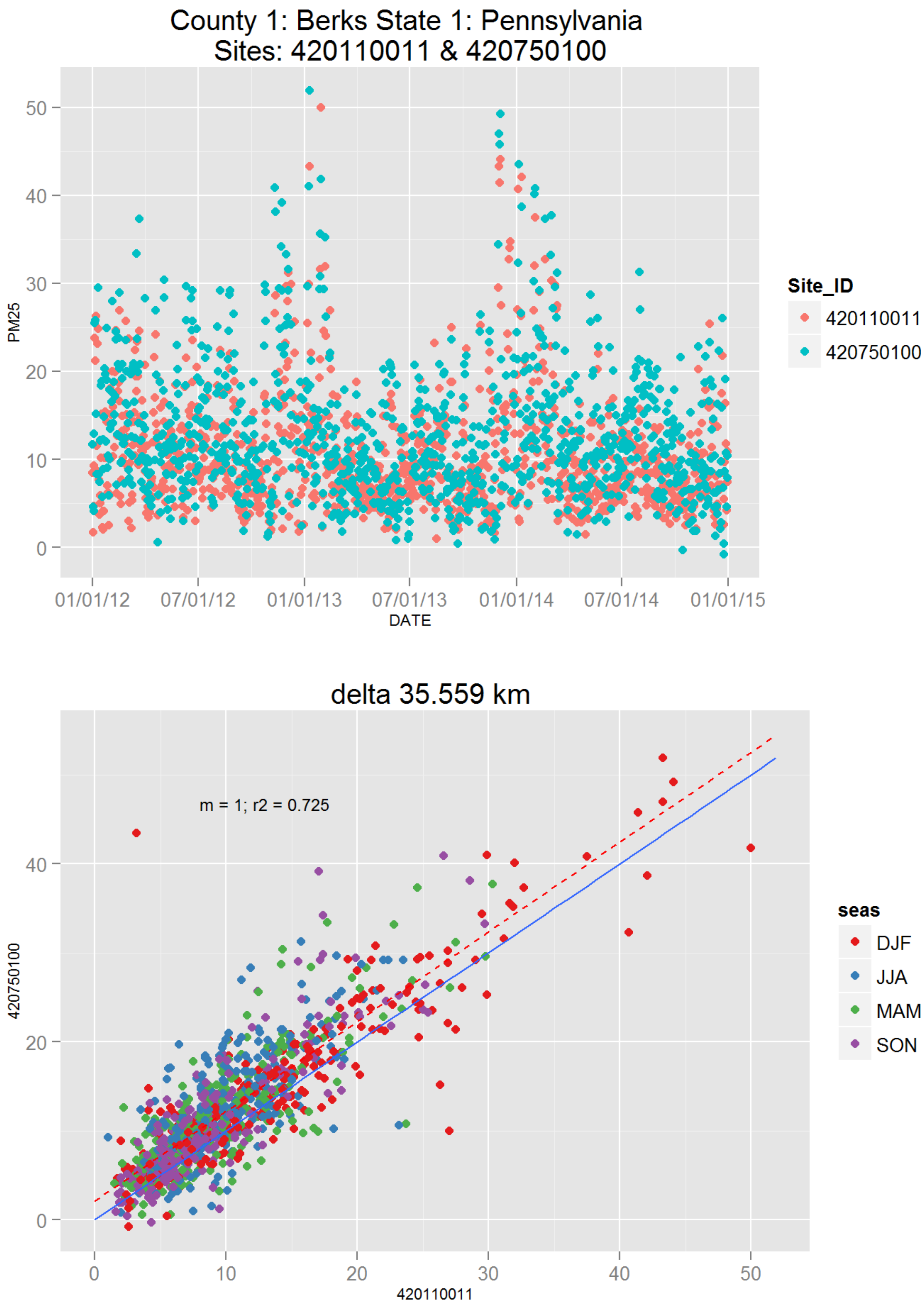


Figure 146: Comparison of $PM_{2.5}$ data for sites 420110011 and 420110011. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 147: Comparison of $PM_{2.5}$ data for sites 420410101 and 420410101. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 148: Comparison of $PM_{2.5}$ data for sites 420450002 and 420450002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

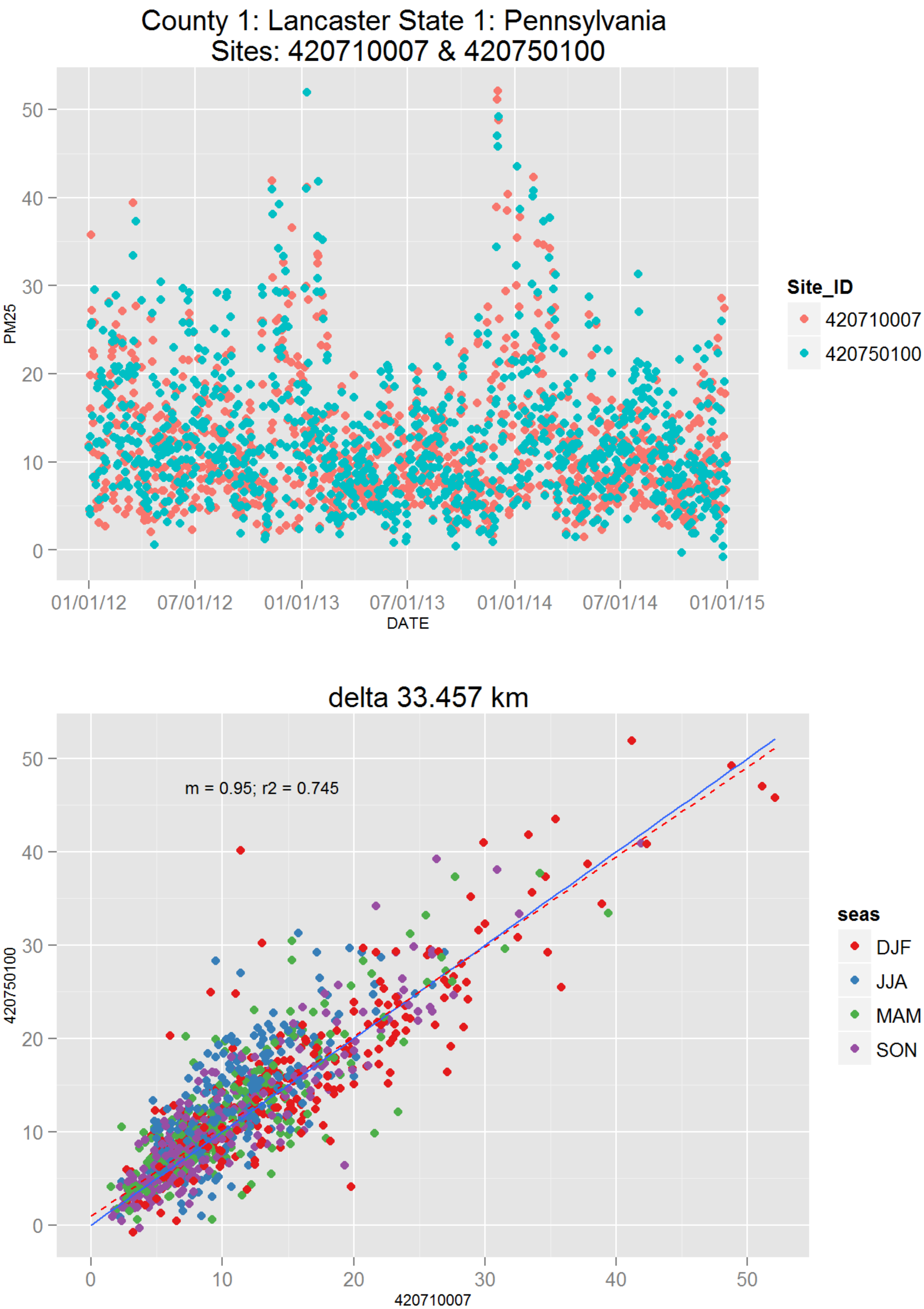


Figure 149: Comparison of $PM_{2.5}$ data for sites 420710007 and 420710007. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

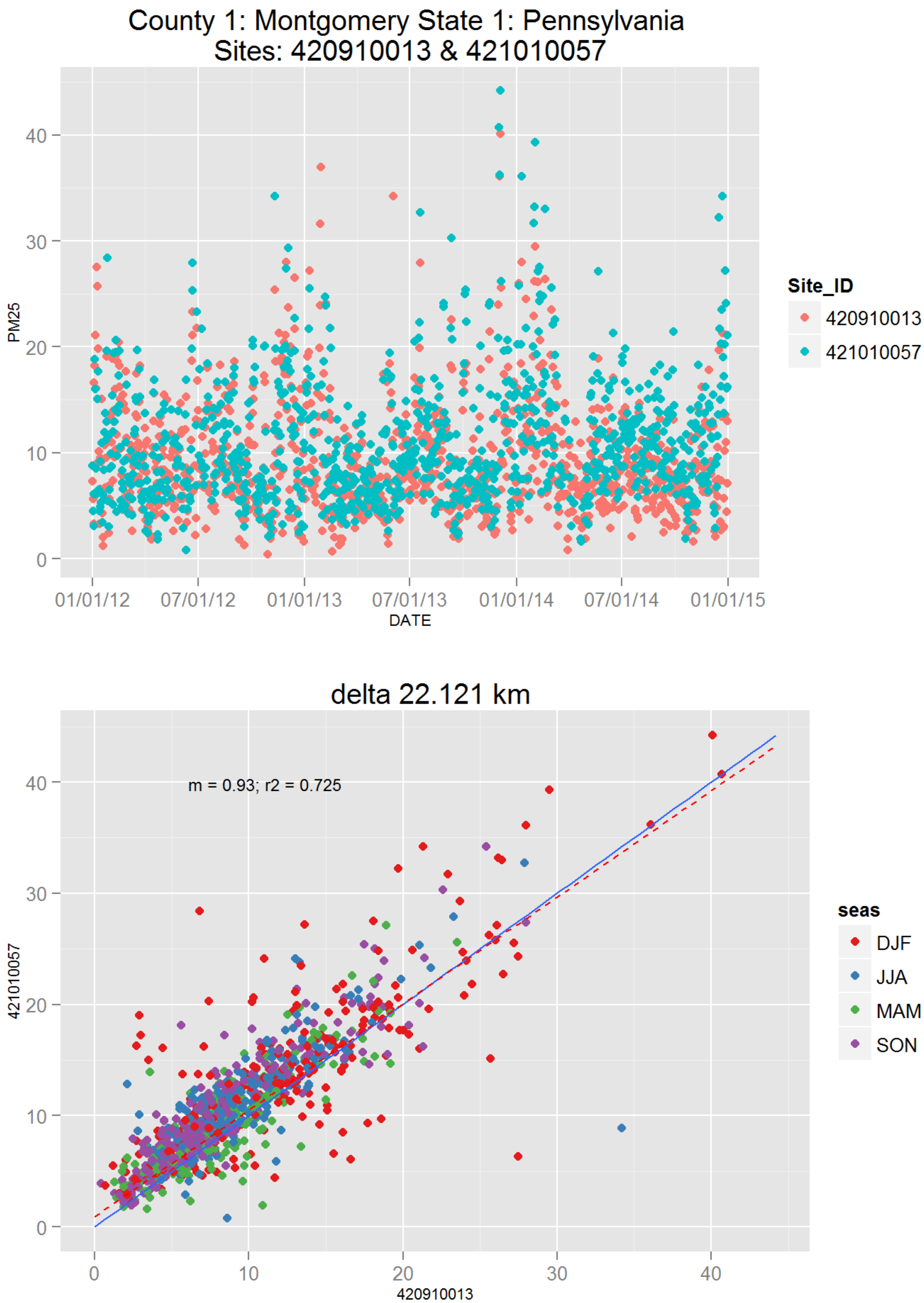


Figure 150: Comparison of $PM_{2.5}$ data for sites 420910013 and 420910013. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

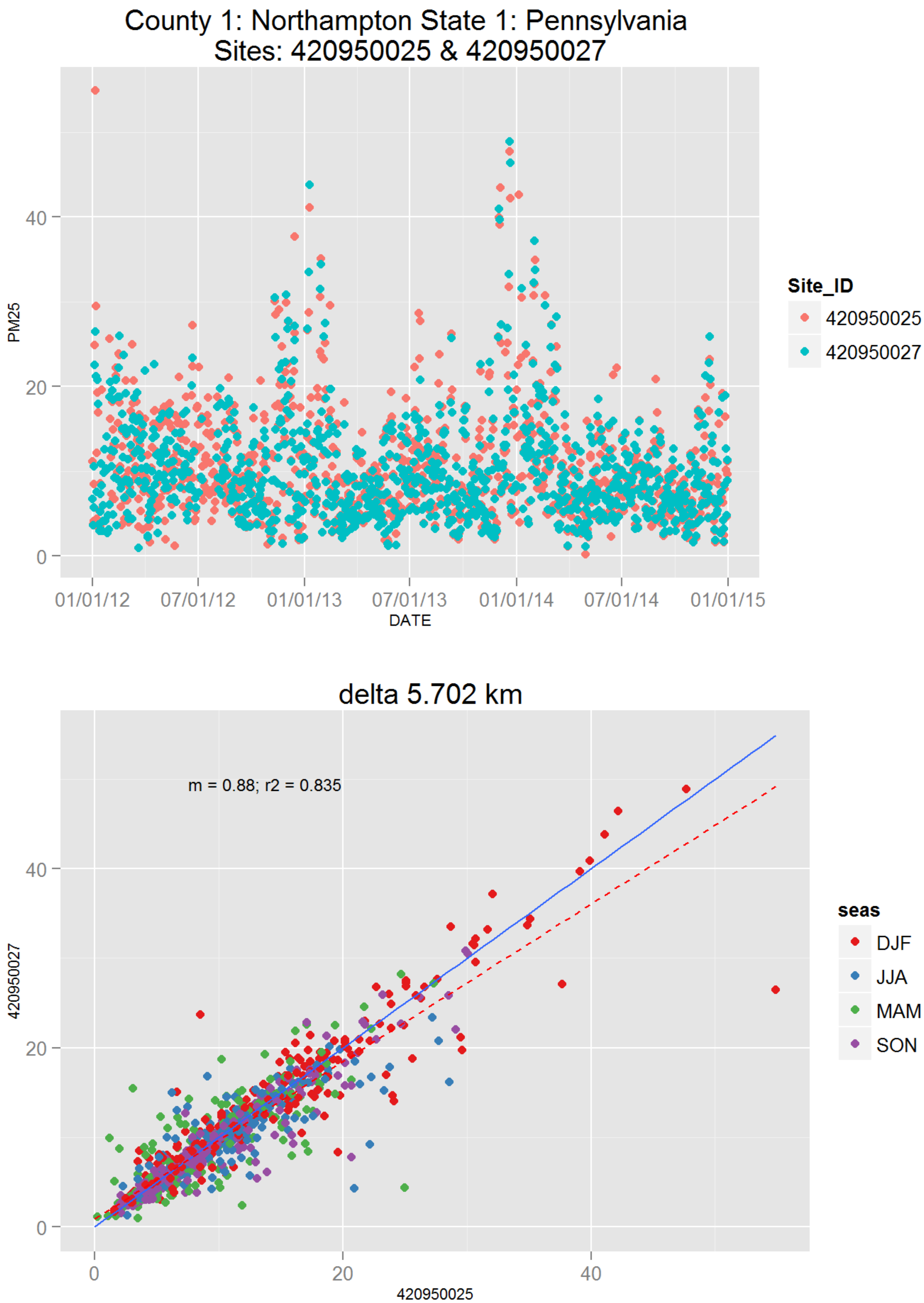
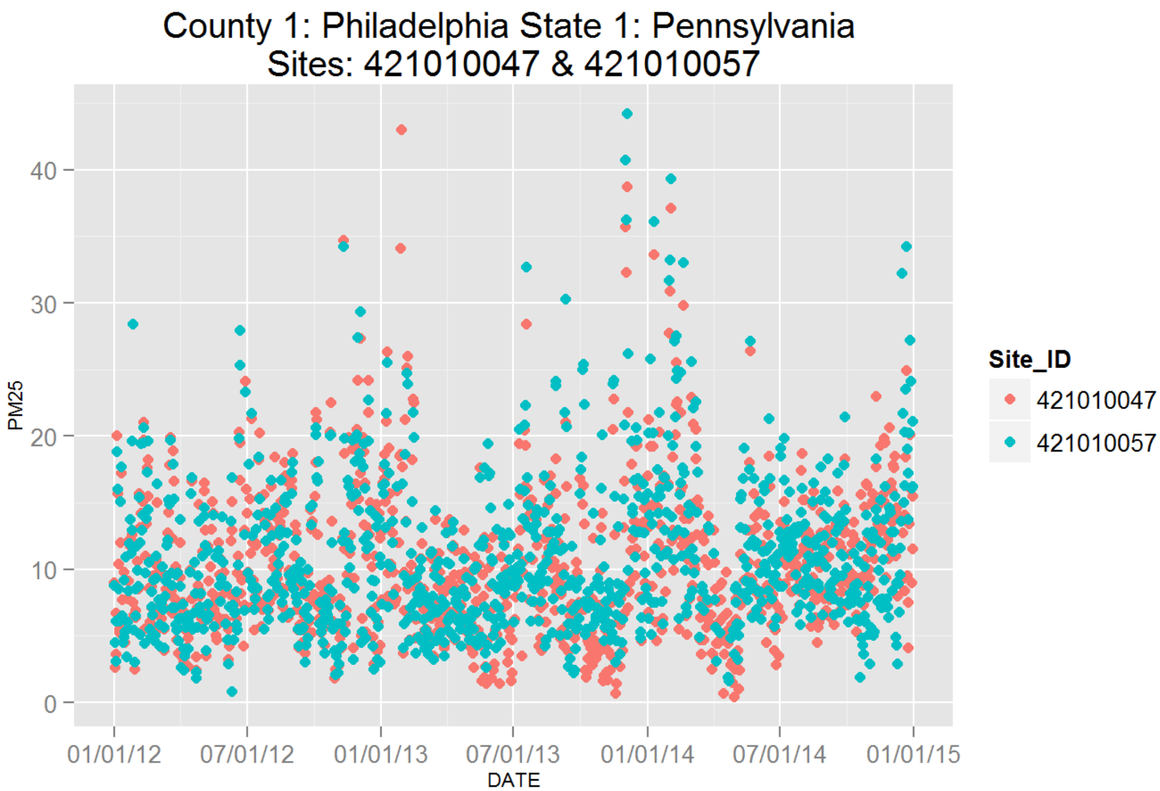


Figure 151: Comparison of $PM_{2.5}$ data for sites 420950025 and 420950025. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



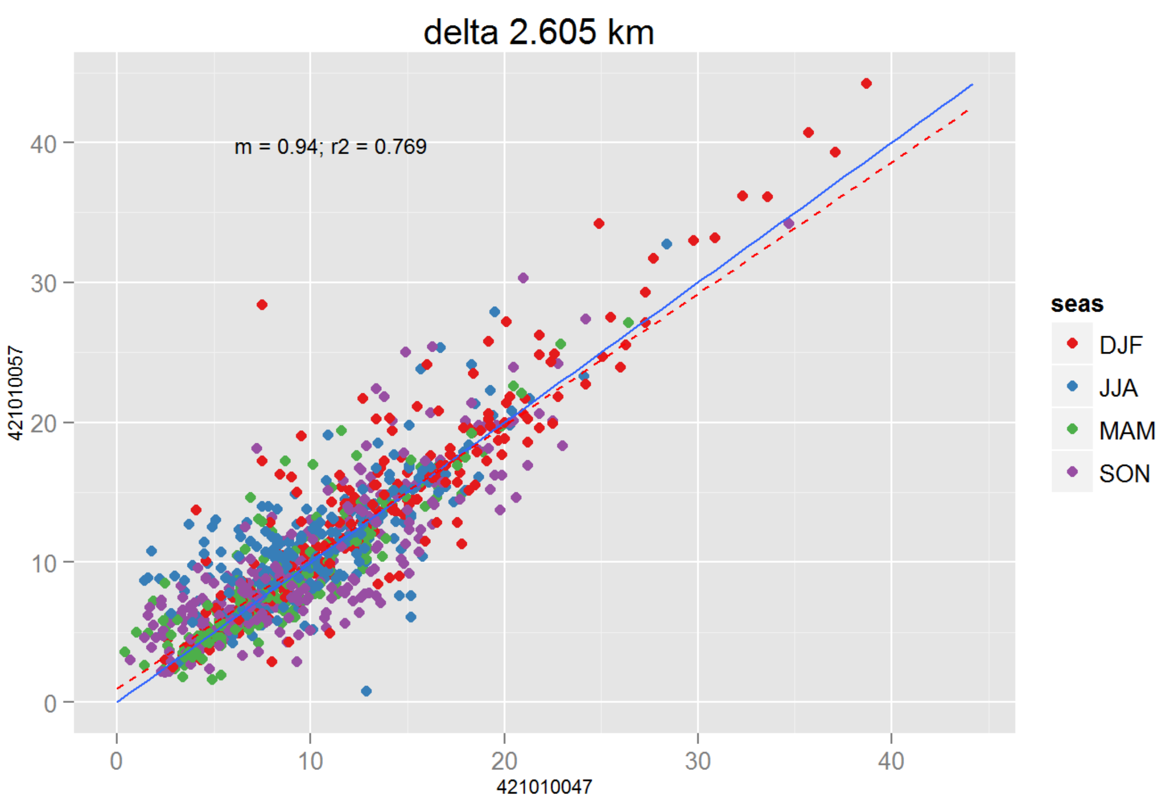


Figure 152: Comparison of $PM_{2.5}$ data for sites 421010047 and 421010047. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

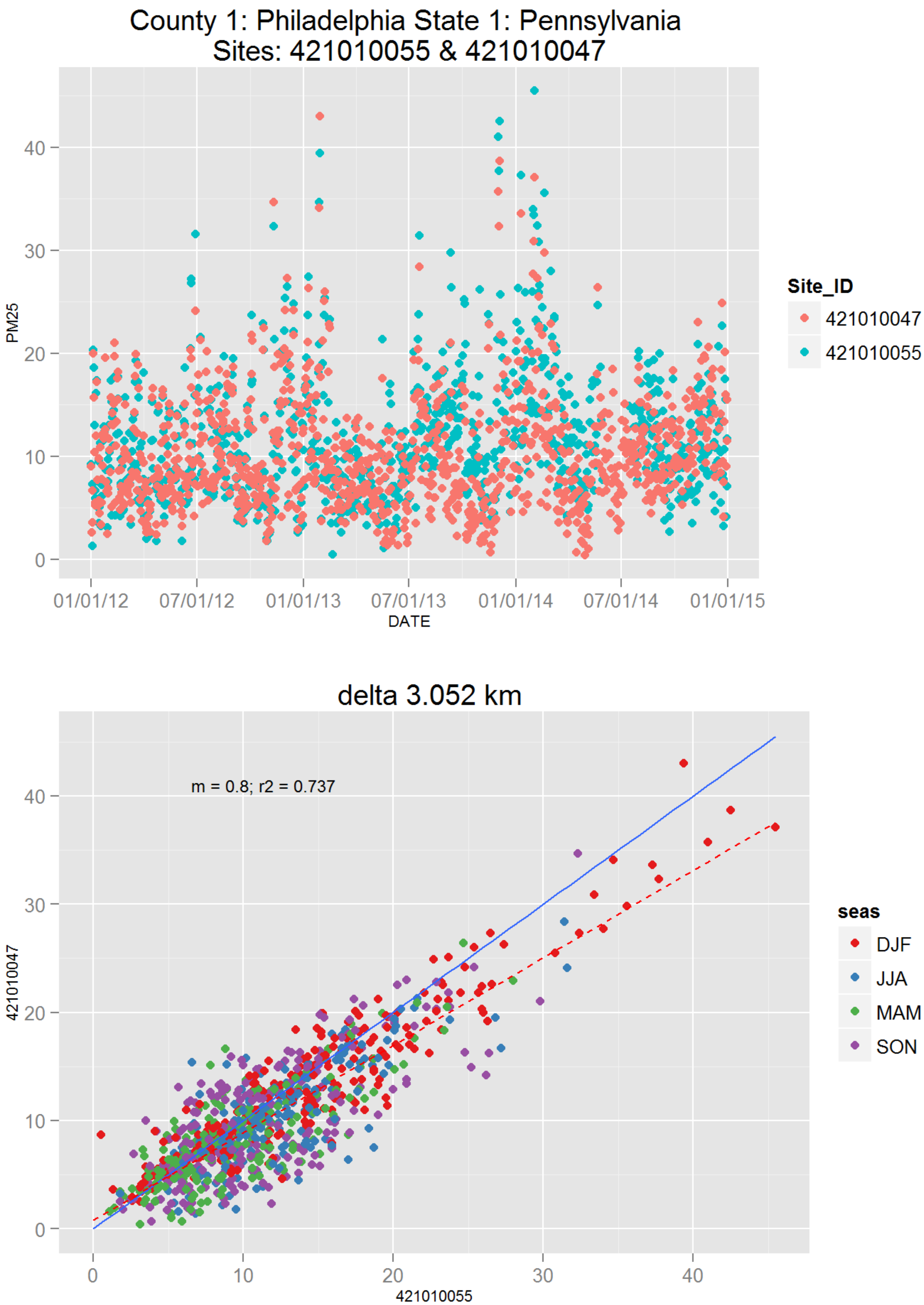


Figure 153: Comparison of $PM_{2.5}$ data for sites 421010055 and 421010055. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 154: Comparison of $PM_{2.5}$ data for sites 421250005 and 421250005. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 155: Comparison of $PM_{2.5}$ data for sites 421250200 and 421250200. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 156: Comparison of $PM_{2.5}$ data for sites 421255001 and 421255001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 157: Comparison of $PM_{2.5}$ data for sites 421290008 and 421290008. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 158: Comparison of $PM_{2.5}$ data for sites 421330008 and 421330008. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

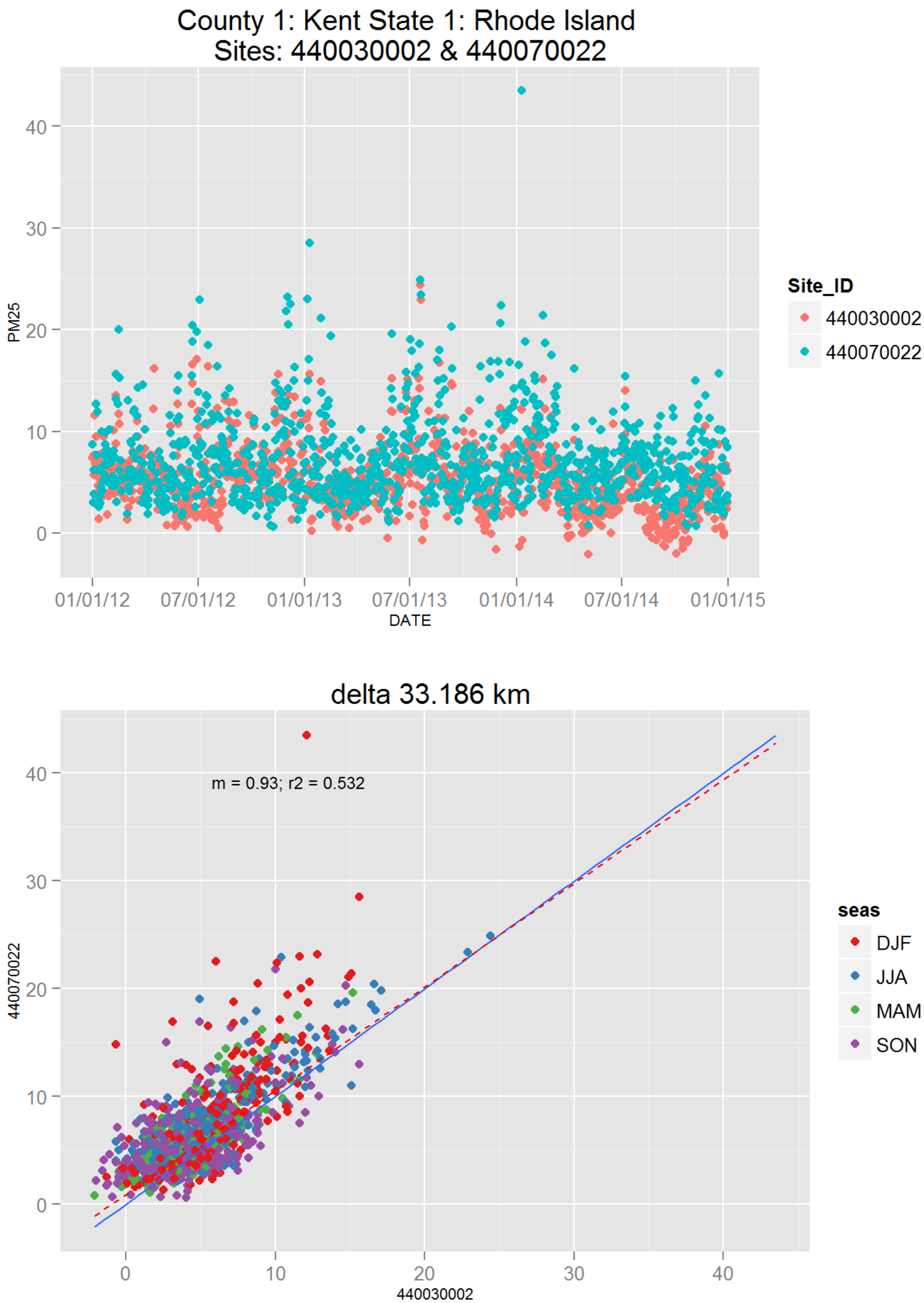


Figure 159: Comparison of $PM_{2.5}$ data for sites 440030002 and 440030002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

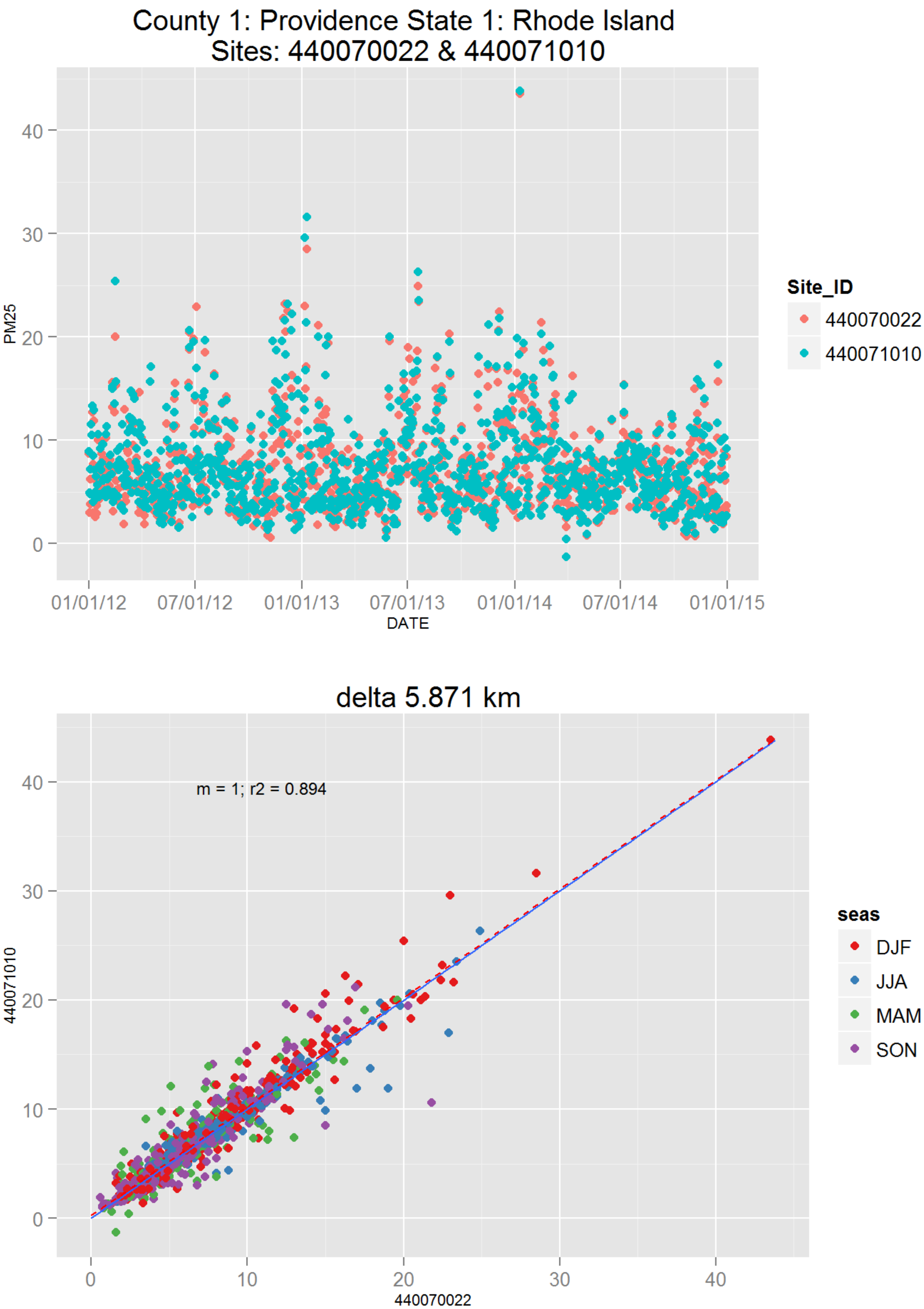


Figure 160: Comparison of $PM_{2.5}$ data for sites 440070022 and 440070022. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

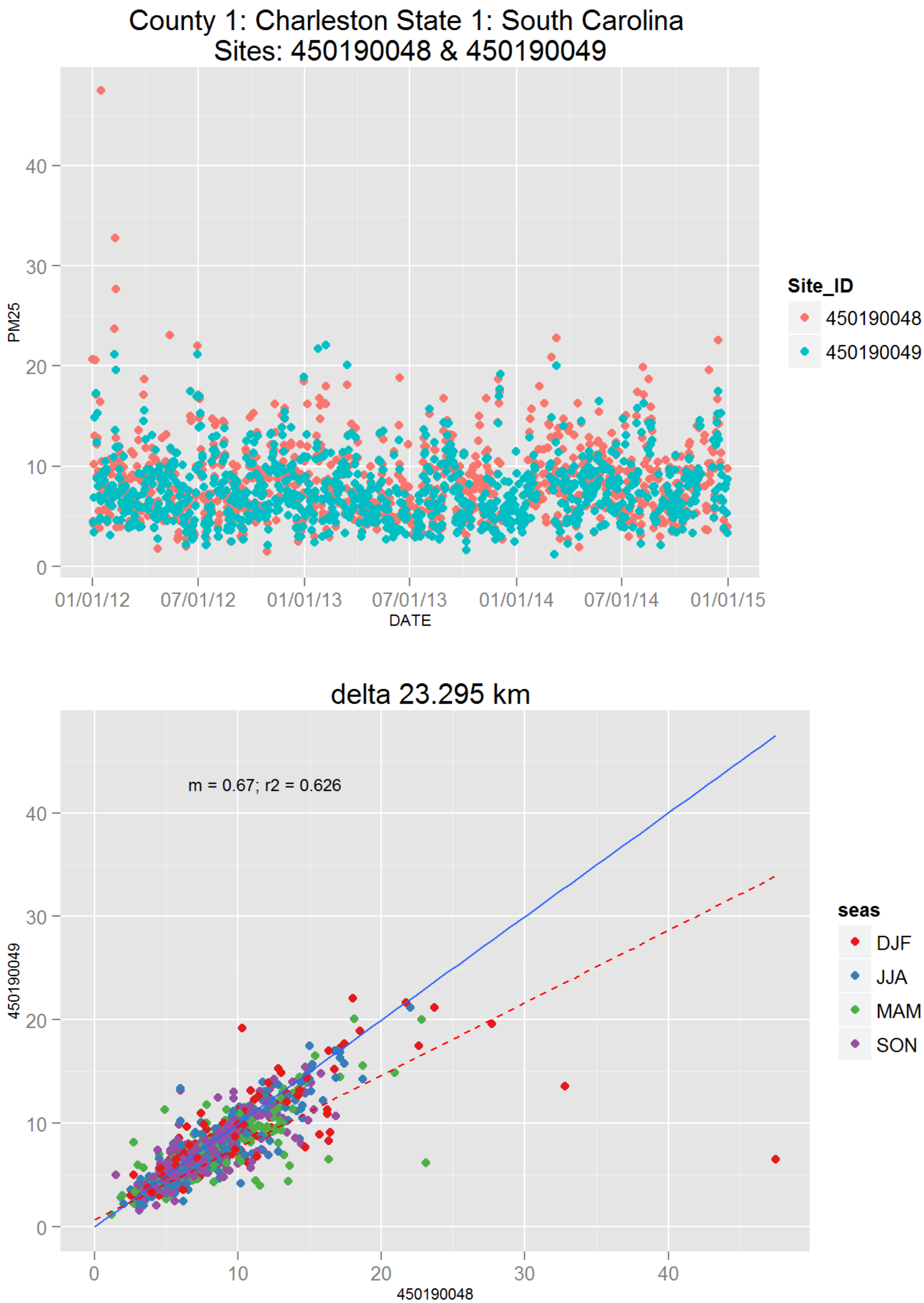


Figure 161: Comparison of $PM_{2.5}$ data for sites 450190048 and 450190048. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

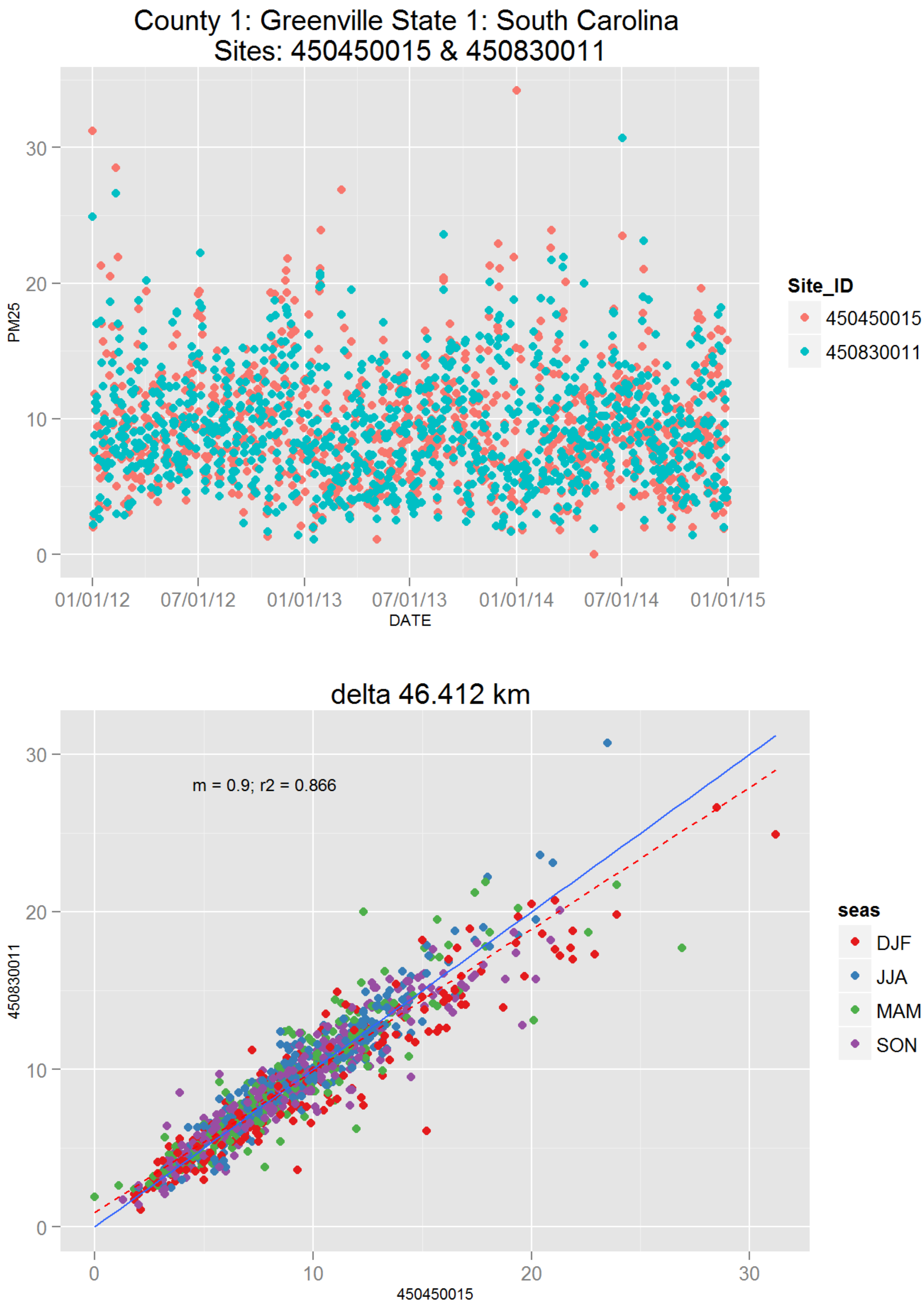


Figure 162: Comparison of $PM_{2.5}$ data for sites 450450015 and 450450015. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 163: Comparison of $PM_{2.5}$ data for sites 450630008 and 450630008. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

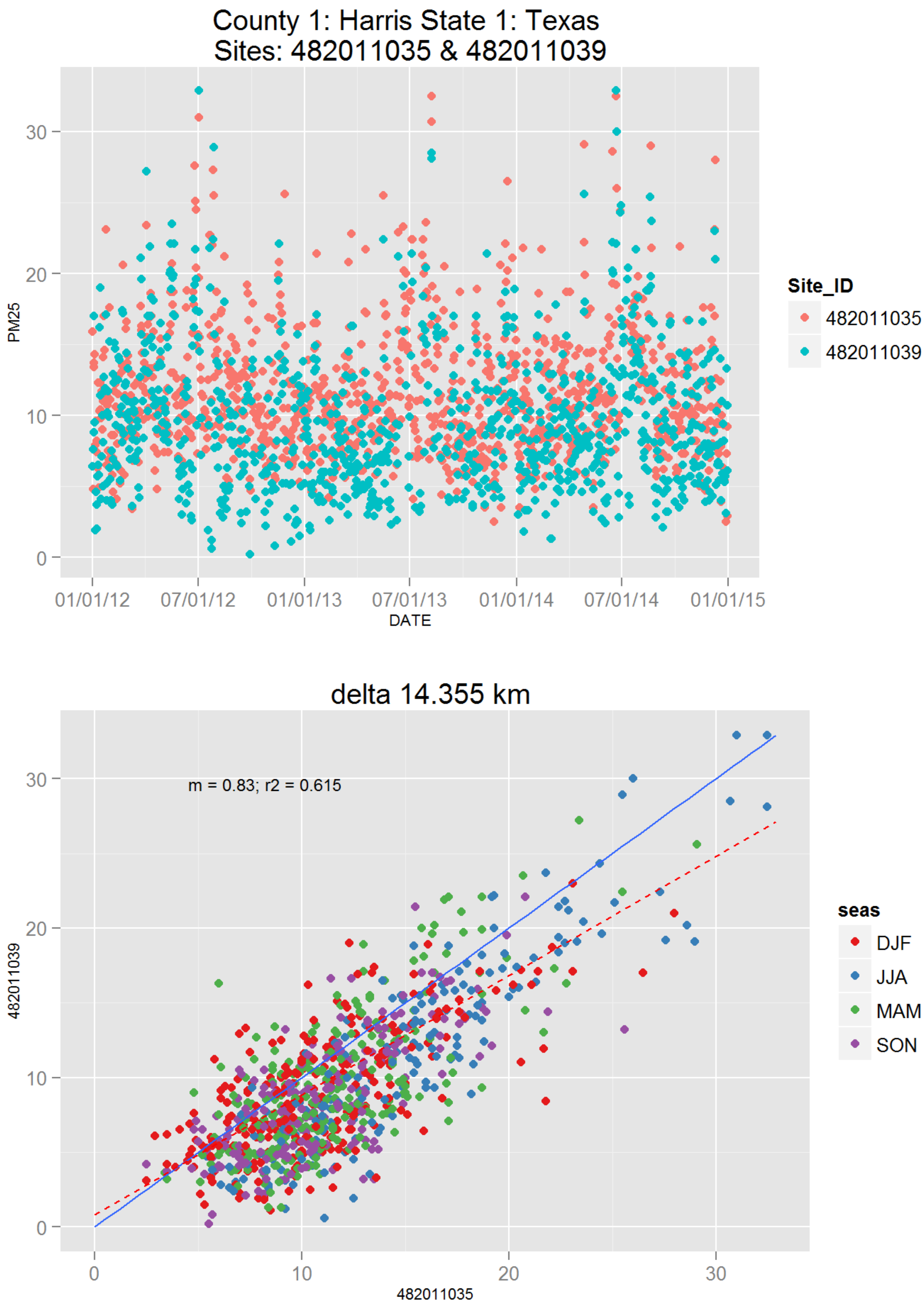


Figure 164: Comparison of $PM_{2.5}$ data for sites 482011035 and 482011035. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

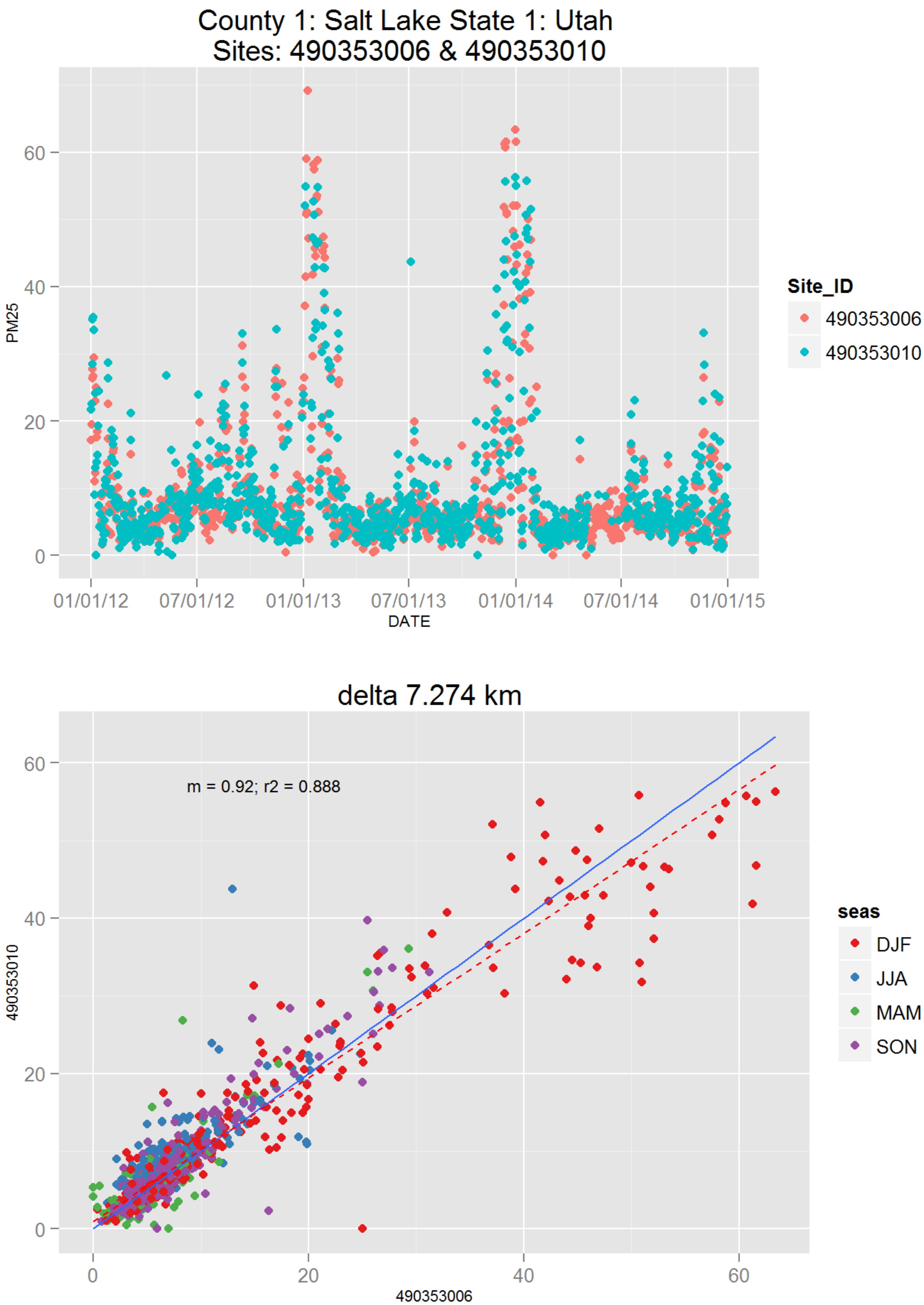


Figure 165: Comparison of $PM_{2.5}$ data for sites 490353006 and 490353006. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

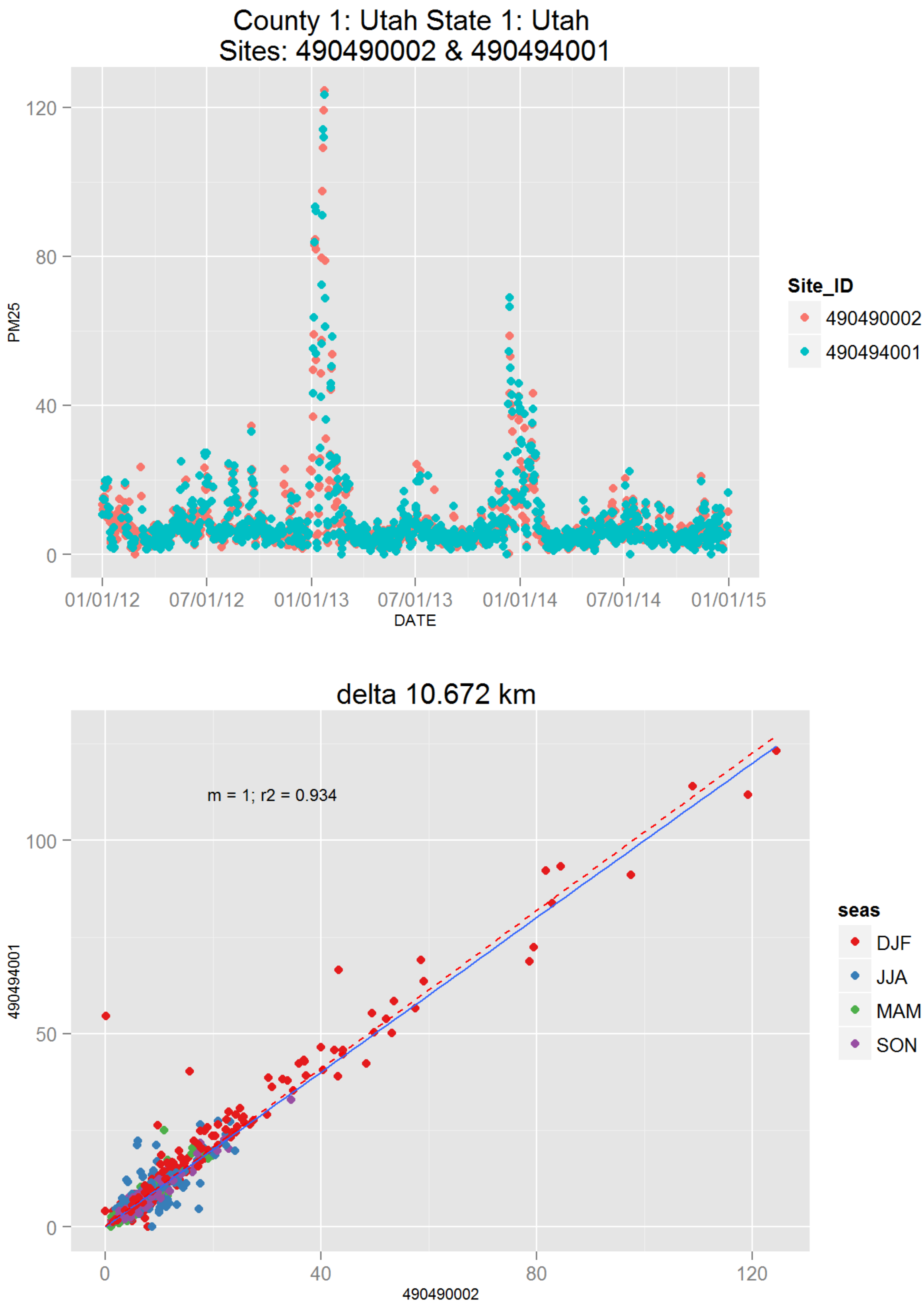
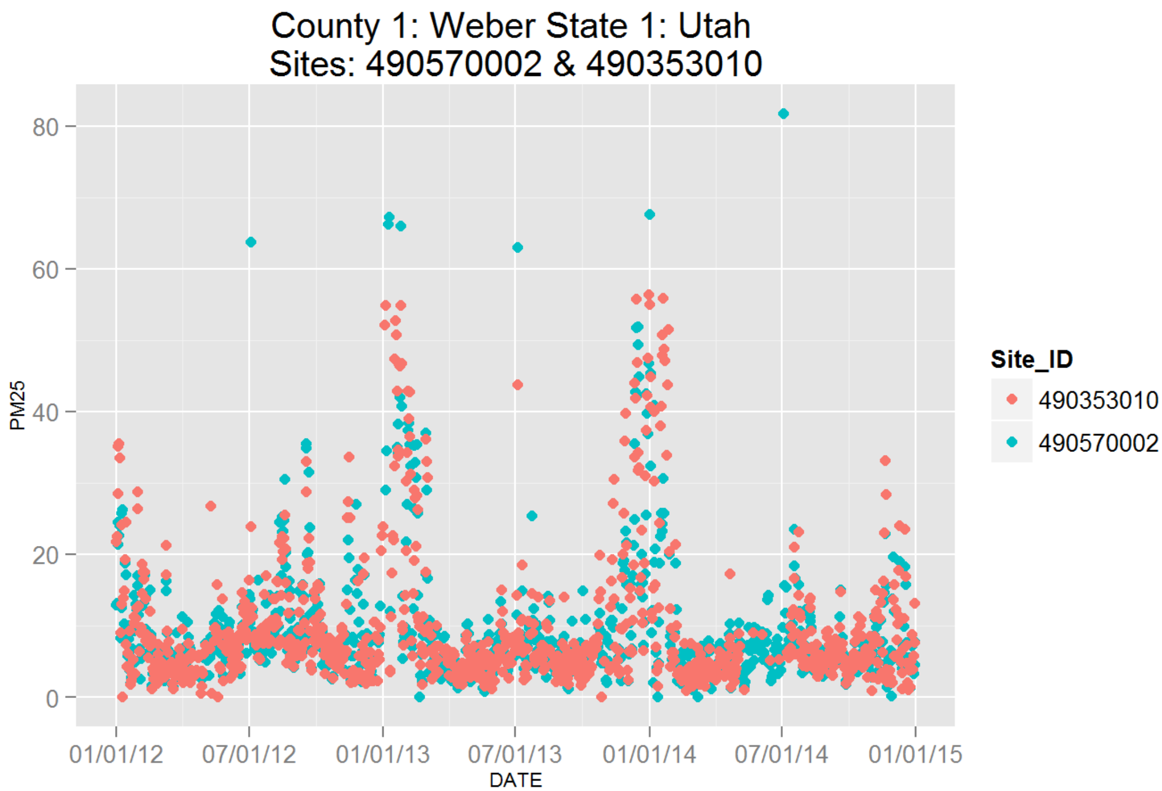


Figure 166: Comparison of $PM_{2.5}$ data for sites 490490002 and 490490002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



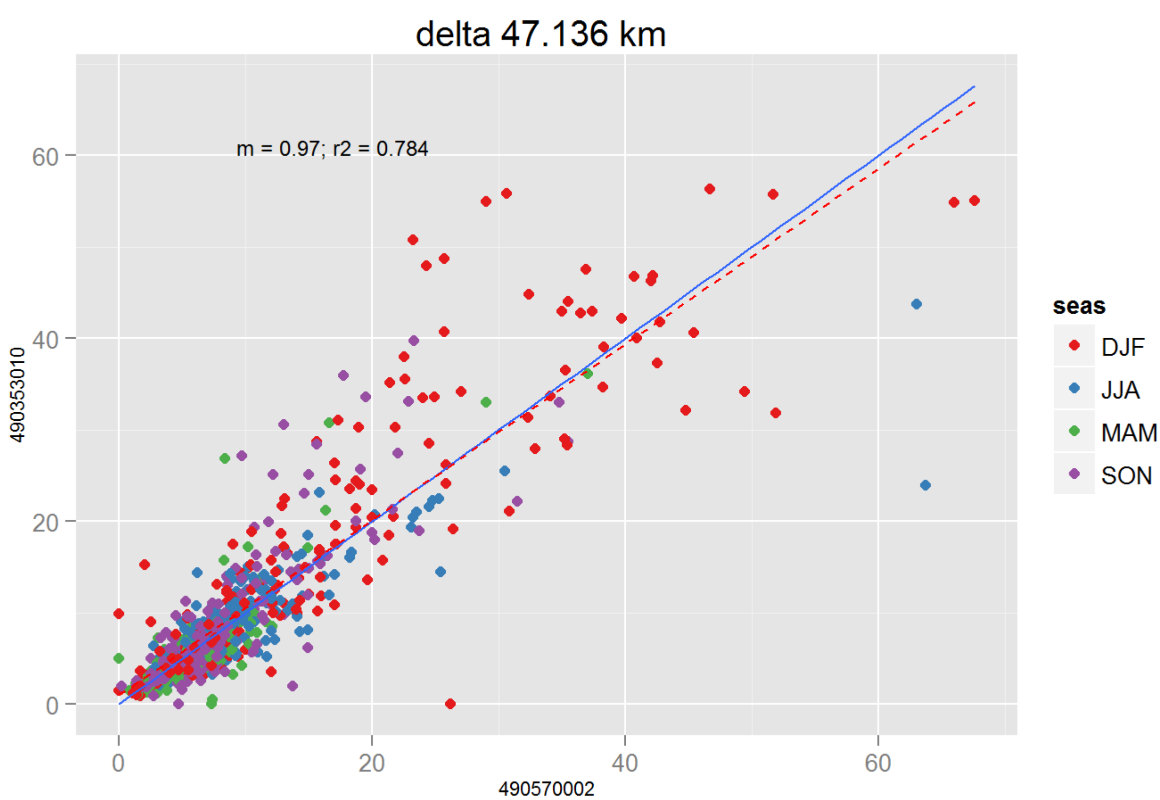


Figure 167: Comparison of $PM_{2.5}$ data for sites 490570002 and 490570002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

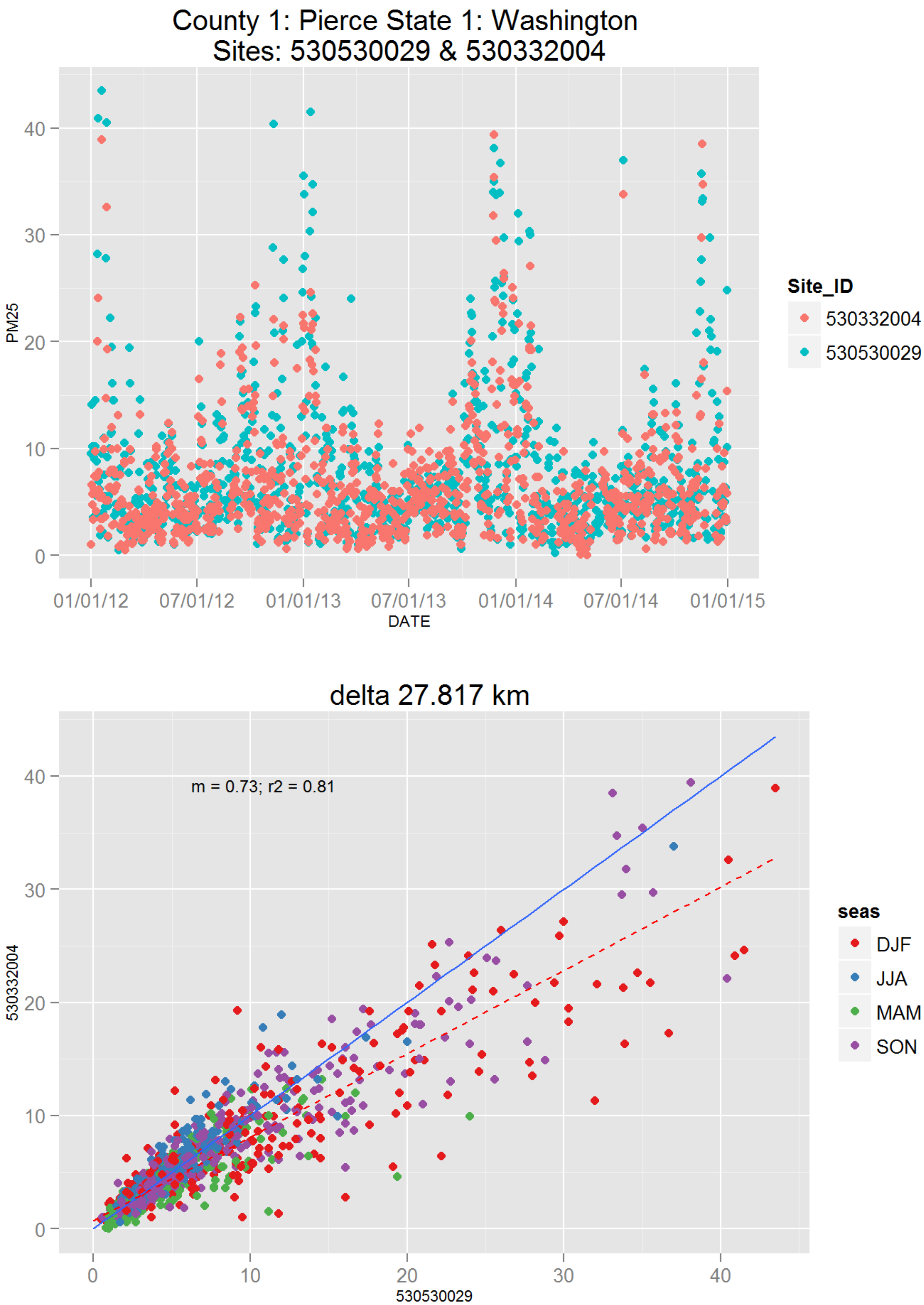
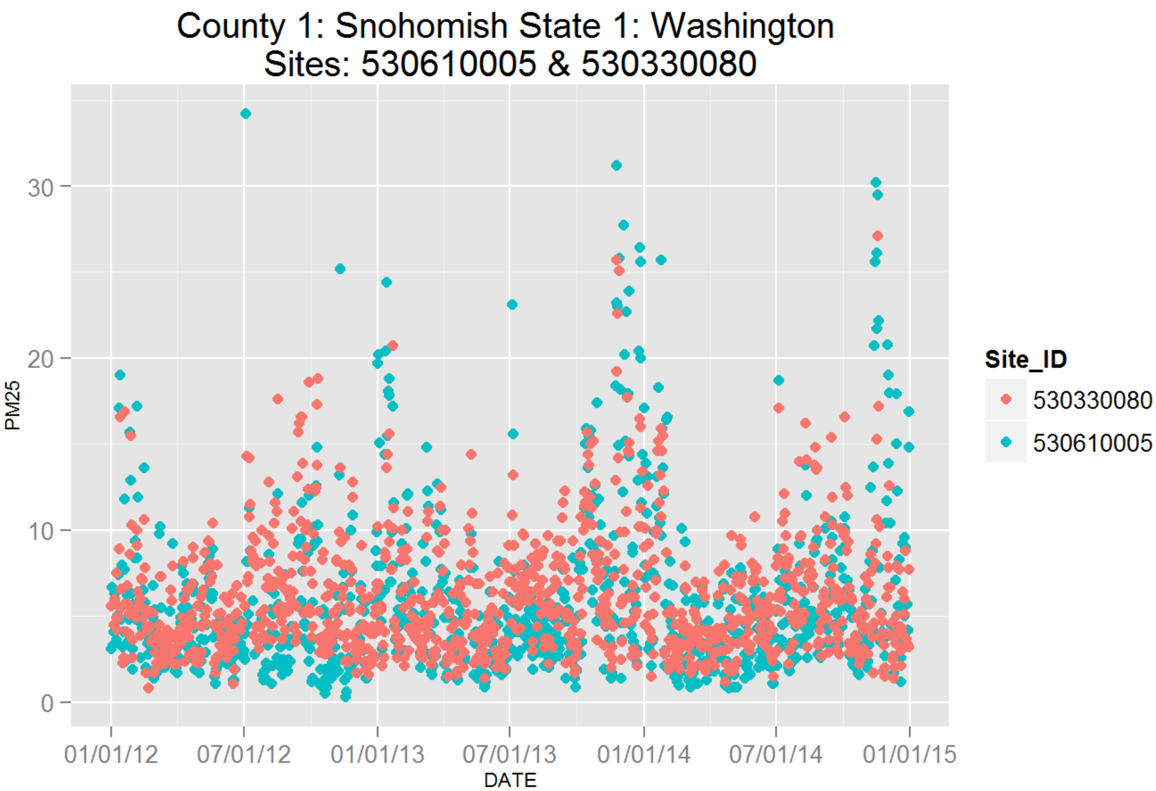


Figure 168: Comparison of $PM_{2.5}$ data for sites 530530029 and 530530029. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



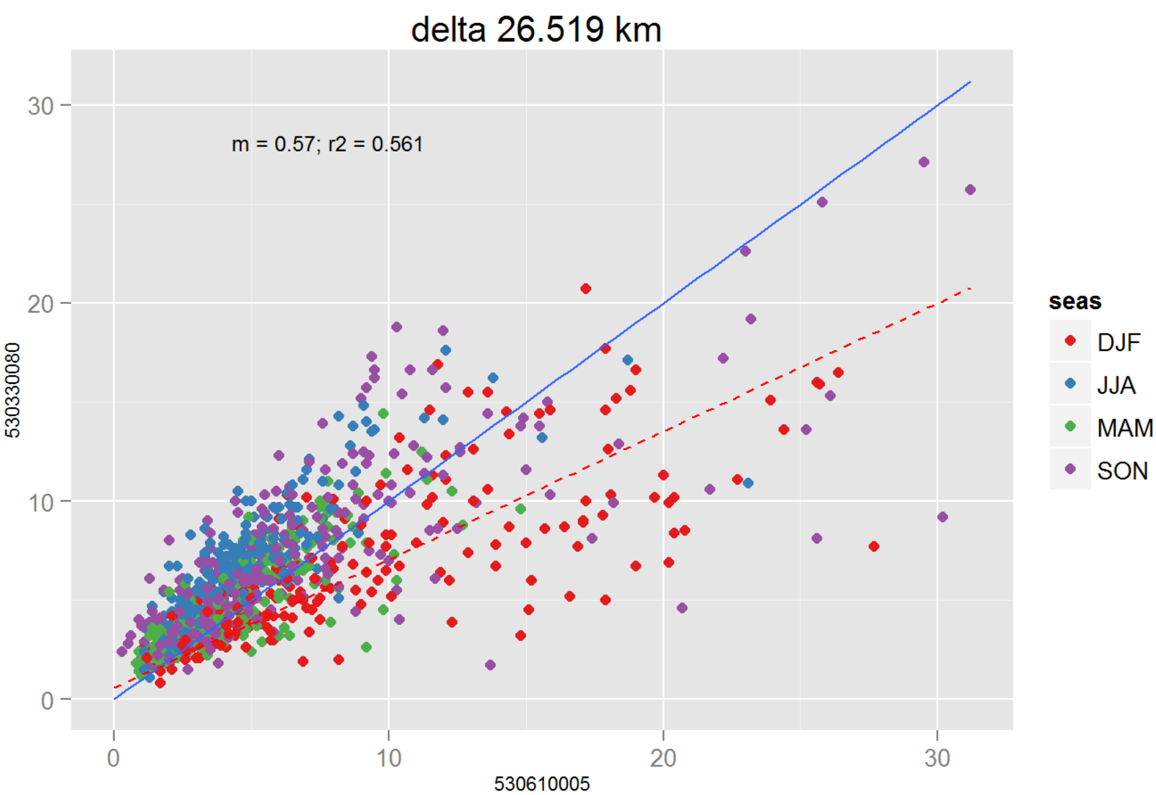


Figure 169: Comparison of $PM_{2.5}$ data for sites 530610005 and 530610005. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

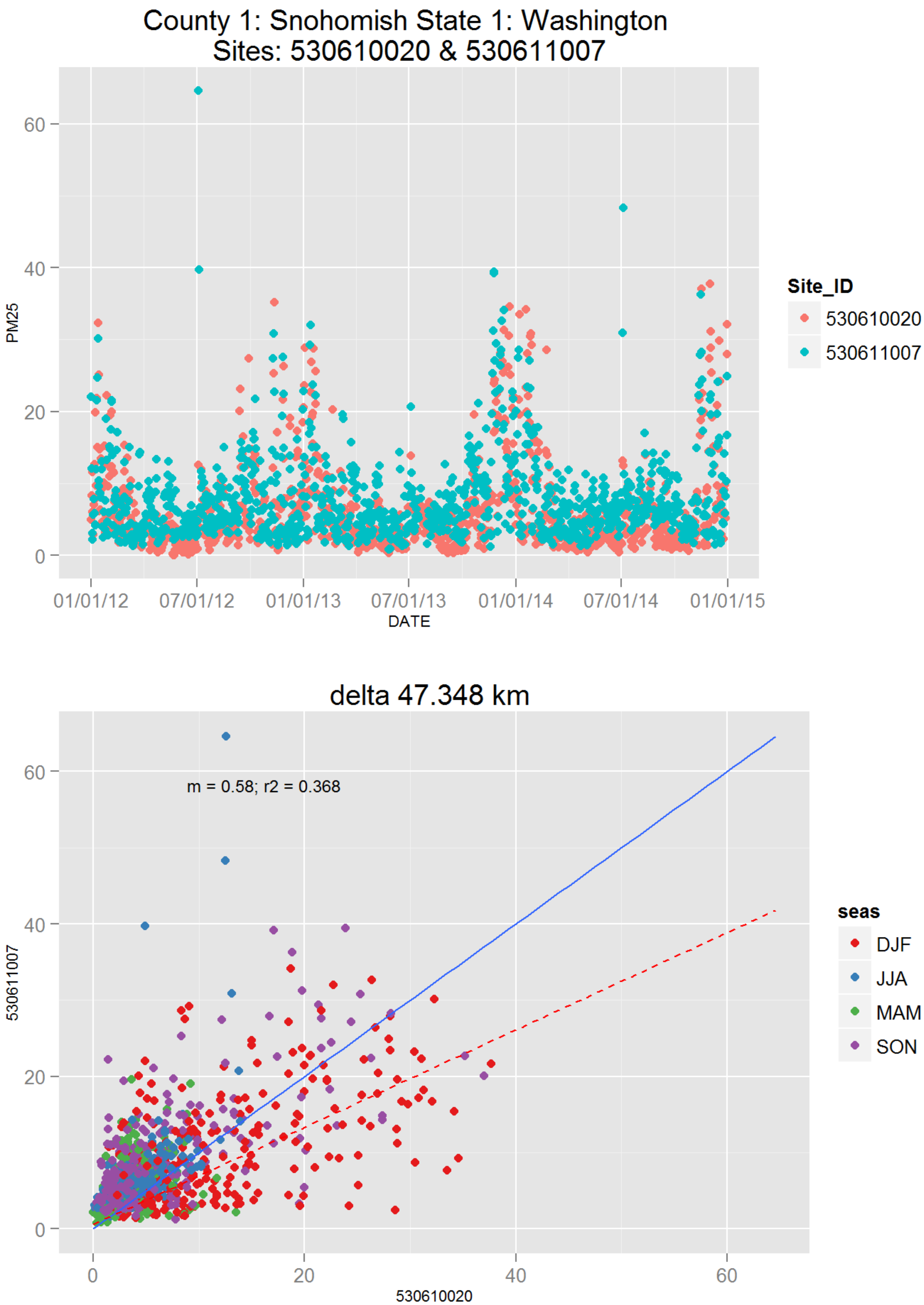


Figure 170: Comparison of $PM_{2.5}$ data for sites 530610020 and 530610020. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

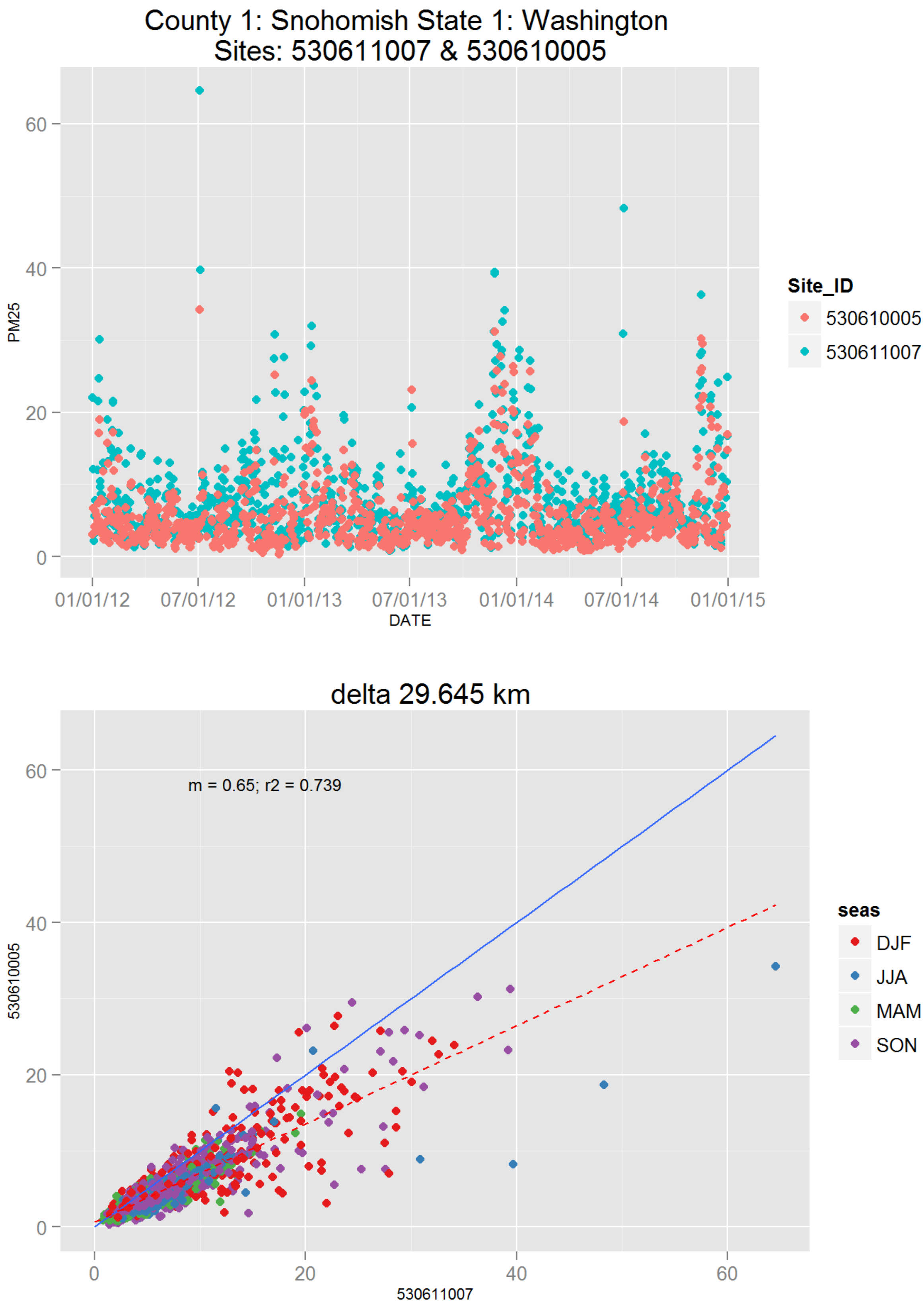


Figure 171: Comparison of $PM_{2.5}$ data for sites 530611007 and 530611007. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 172: Comparison of $PM_{2.5}$ data for sites 550090005 and 550090005. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

5 Comparison of air quality variability for ozone sensitivity tests

Results from the ozone sensitivity analysis discussed in Section 2.2.3.

5.1 All available data, no quarterly subsets

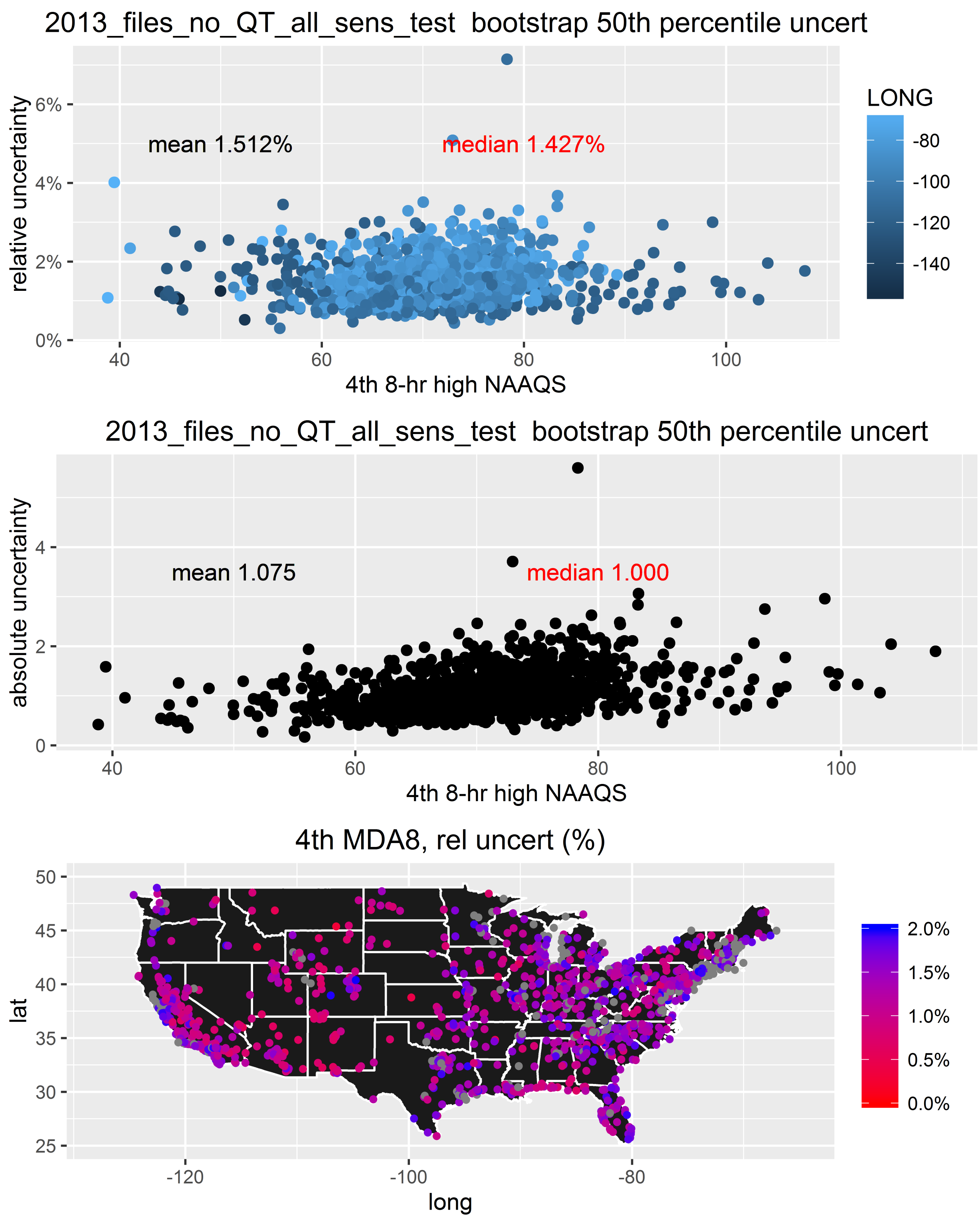


Figure 173: Bootstrap results for the ozone 2013 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

5.2 All available data, with quarterly subsets

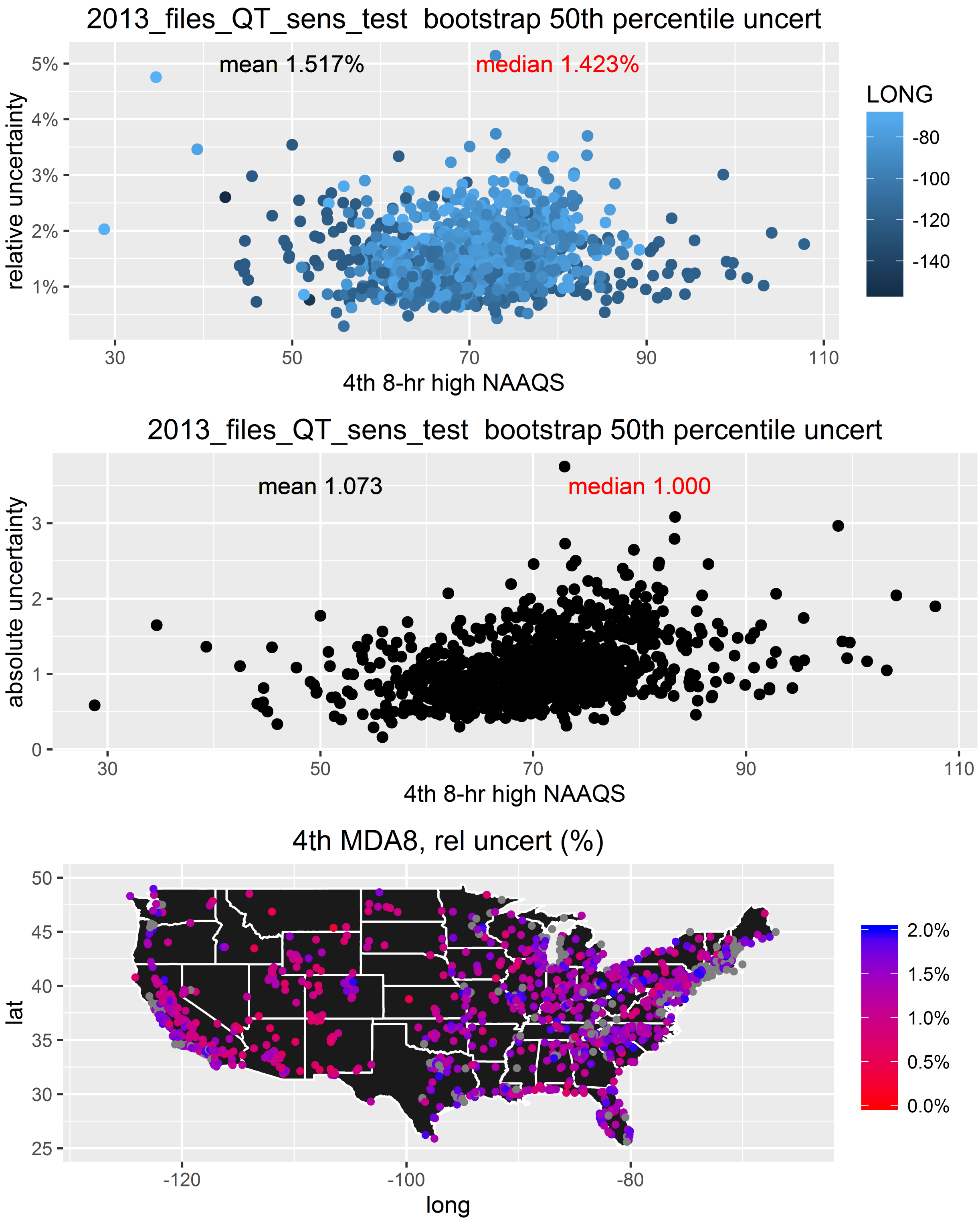


Figure 174: Bootstrap results for the ozone 2013 DVs, showing the 50%, 65%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

6 Analysis of temporal lag on ozone data and results from a blocked bootstrap sensitivity analysis

This section presents results from an analysis to examine the temporal correlation of air quality levels, i.e., the tendency of high concentration days to occur after other days with high concentrations. Such behavior, if present, would be a function of both emission trends (e.g., weekday traffic versus weekend traffic) and meteorology (e.g., high pressure systems often hinder the transport of pollutants and also accompany higher temperatures, which tend to increase the formation of ozone). The primary motivation for this assessment is to determine whether the implementation of a block bootstrap procedure is needed for the bootstrapping analysis described in Section 2.2.3 in order to account for possible temporal correlation, and if so, what is the appropriate block size. If not properly accounted for, correlation can affect the assessment of uncertainty (i.e., the standard errors used to calculate confidence intervals). While in this analysis confidence intervals were constructed using empirical percentiles, it is important to consider whether autocorrelation may be affecting the distributional characteristics of the bootstrapped data. Thus a sensitivity analysis is considered.

A block bootstrap method can be used in the presence of autocorrelation to replicate the correlation structure in the data. Blocks are designed such that the dependence between adjacent or closely spaced measurements is contained within a block, and there is induced independence between measurements in adjacent blocks. Block size selection can be tricky, as the blocks should be large enough to induce independence but small enough to retain important characteristics of the data, including natural variation and overall trends (i.e., the variance-bias trade-off for avoiding over-smoothing). There is no one agreed upon method for selection of block size for bootstrapping procedures. Many considerations can come into play, including practical issues and subject-matter scientific expertise. The analysis presented here first attempts to determine the "length of lag" in the ambient ozone data (i.e., how long do correlations of concentrations between MDA8 values persist). Based on the lag analysis, a secondary "blocked" bootstrap analysis was completed which sampled blocks of days corresponding to the lag found in the initial analysis. Ultimately, a 7-day lag was selected from the lag analysis. The resultant bootstrap results were similar to the original non-parametric bootstrap, which sampled individual days rather than blocks of 7-days.

6.1 Analysis procedure and results

The R software package [R Core Team, 2017] was used to conduct the lag analysis. The acf (autocorrelation function) and pacf (partial autocorrelation function) were used to determine the autocorrelation of the time series of MDA8 values at each measurement site for all data available from 2016. The results from the network-wide correlations were summarized in Figures 175 and 176.

The results from the acf analysis suggest that autocorrelations drop off after lag 3, as the mean and median correlation coefficients at 4, 5 and 6 days lag are equivalent (top panel of Figure 175). While the correlations are still within the 95% confidence interval returned from acf out to the 6-day lag, the fact that the distribution of the differences between the individual correlation and the confidence intervals (middle and bottom panel of Figure 175) are virtually identical starting at 3 days of lag suggest that correlations at this level would be found at any lag period. The pacf analysis accounts for the autocorrelation found in the previous lag periods (i.e., the correlation found for the 2-day lag removes the correlation found from the 1-day lag). The results from this analysis suggest that the autocorrelation is only significant to one day. Taking these results into account, a 3-day lag should appropriately account for any autocorrelation in the ozone data. This is implemented in the bootstrapping analysis via a 7-day block size to account for +- 3 days surrounding the sampled daily value. Thus, prior to bootstrapping, the data is grouped into fixed blocks of size n $\bar{7}$ and the sampline with replacement is performed on the blocks. This is also consistent with block sizes used in Inoue and Shintani [2006] and Hall and Horowitz [1996]. A 7-day block size also addresses the consideration of weekly (7-day) pollution patterns across weekday to weekend that may exist.

A second bootstrap analysis was completed for the 2016 ozone data using a block sampling method, with the 7-day block sample size, in order to determine the effect of possible temporal autocorrelation on the bootstrap confidence intervals. The analysis was conducted with the R "boot" package with the tsboot (time seties bootstrap) package with block resampling with fixed block lengths. Simple blocking, rather than overlapping blocks of randomly varying widths, should suffice for initial consideration of possible effect of dependence on bootstrapped confidence intervals (Lahiri [1999] and Andrews [2002]). The results from this bootstrap approach are shown in Figures 177 and 178 (which can be compared to the results from the non-paremetric bootstrap in Sections 3.1 and 4.2 in the main document). The results detailed in Figure 178 indicate slightly greater variability in the blocked bootstrap result, with the mean variability from the blocked bootstrap was 1.62%, versus 1.42% from the non-parametric bootstrap, and the median was 1.55%, versus 1.47% from the non-parametric bootstrap. while there are a few sites with notable larger variability, as with the non-parametric bootstrap, there is no large-scale trend in the variability. The only location of note is perhaps the Uinta Basin in Utah, where a cluster of sites are grayed out in the map, indicating variability greater than the color scale. These sites have the highest variability from the blocked bootstrap. The Uinta Basin is known to have a unique patter in high-ozone days, with the maximum concentrations occurring in the winter during unique meteorological events, such that the high days are always clustered together. As a result, this highly unique ozone pattern has distinctly different results in the blocked bootstrap as compared to the non-parametric bootstrap.

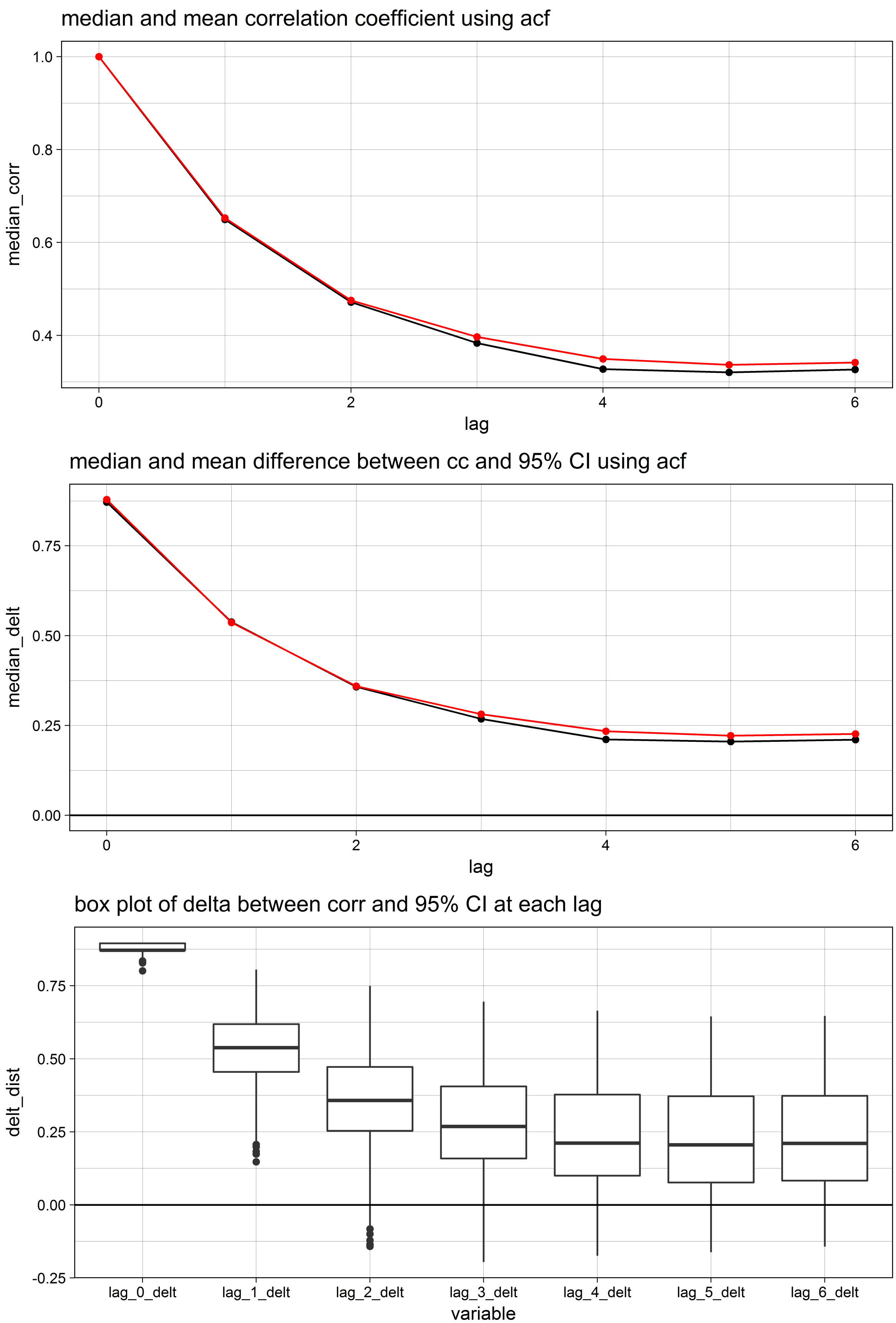


Figure 175: Mean (red lines) and median (black lines) correlations from the acf analysis for ozone data from 2016.

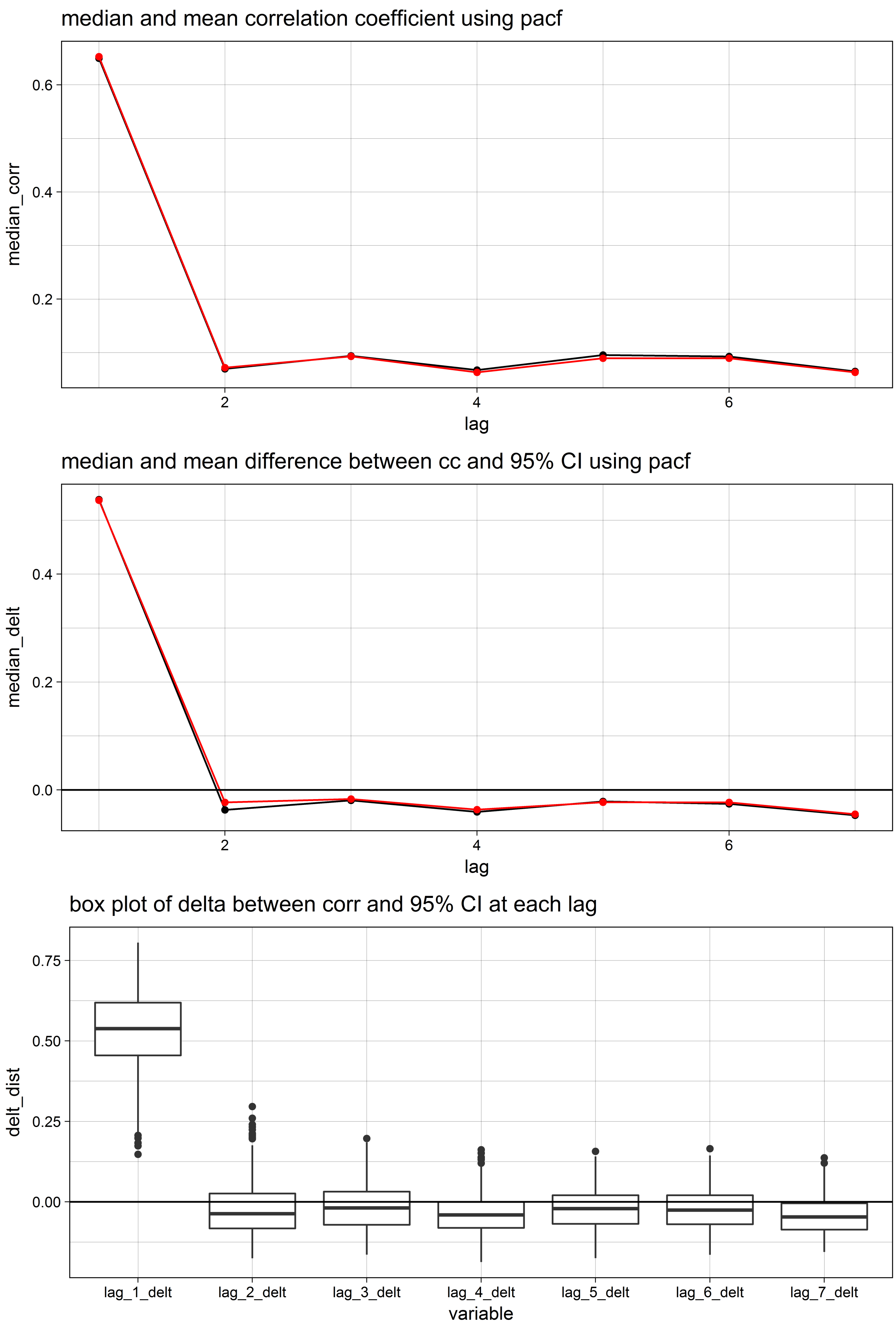


Figure 176: Mean (red lines) and median (black lines) correlations from the pacf analysis for ozone data from 2016.

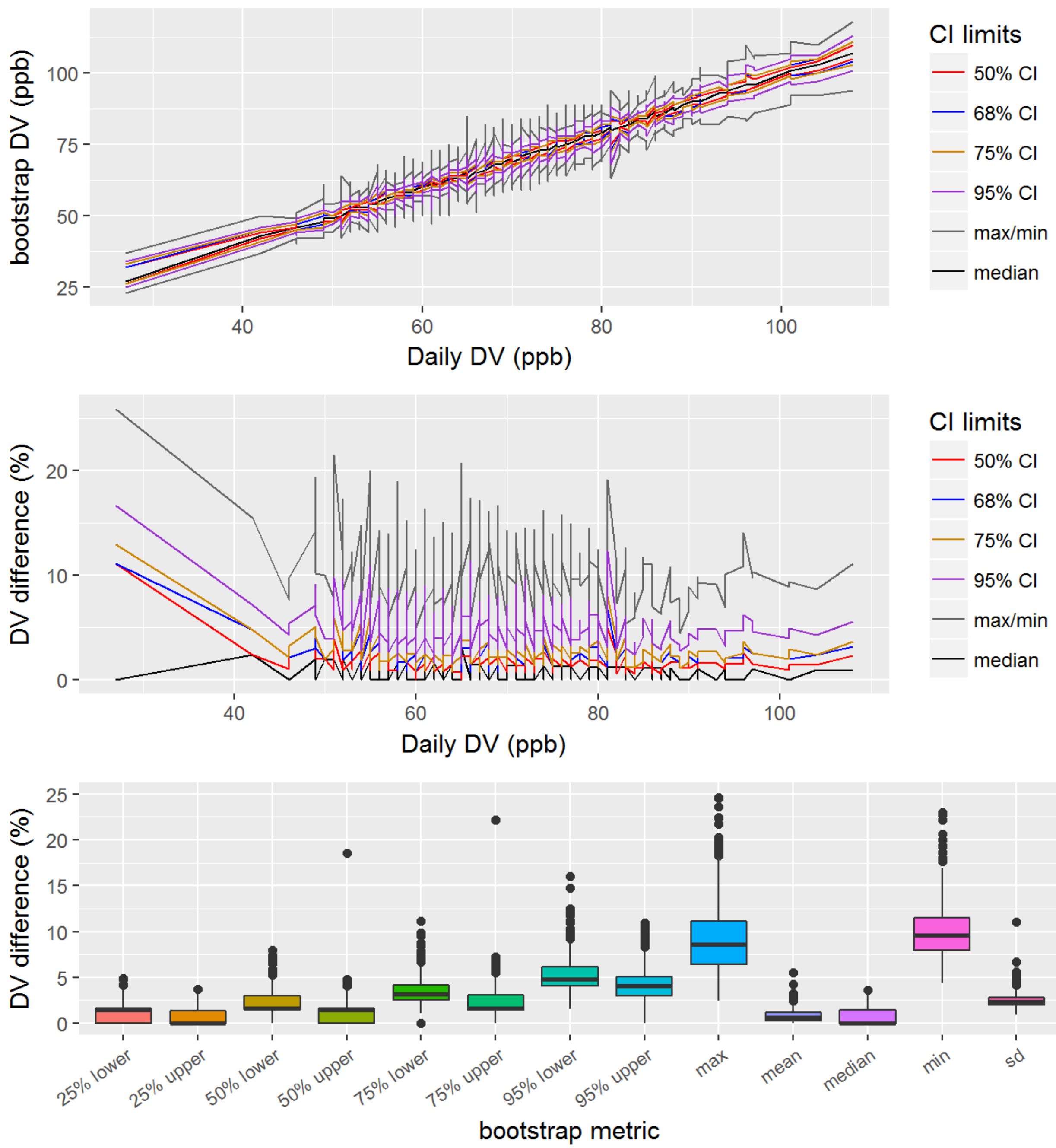


Figure 177: Blocked-bootstrap results for the ozone 2016 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

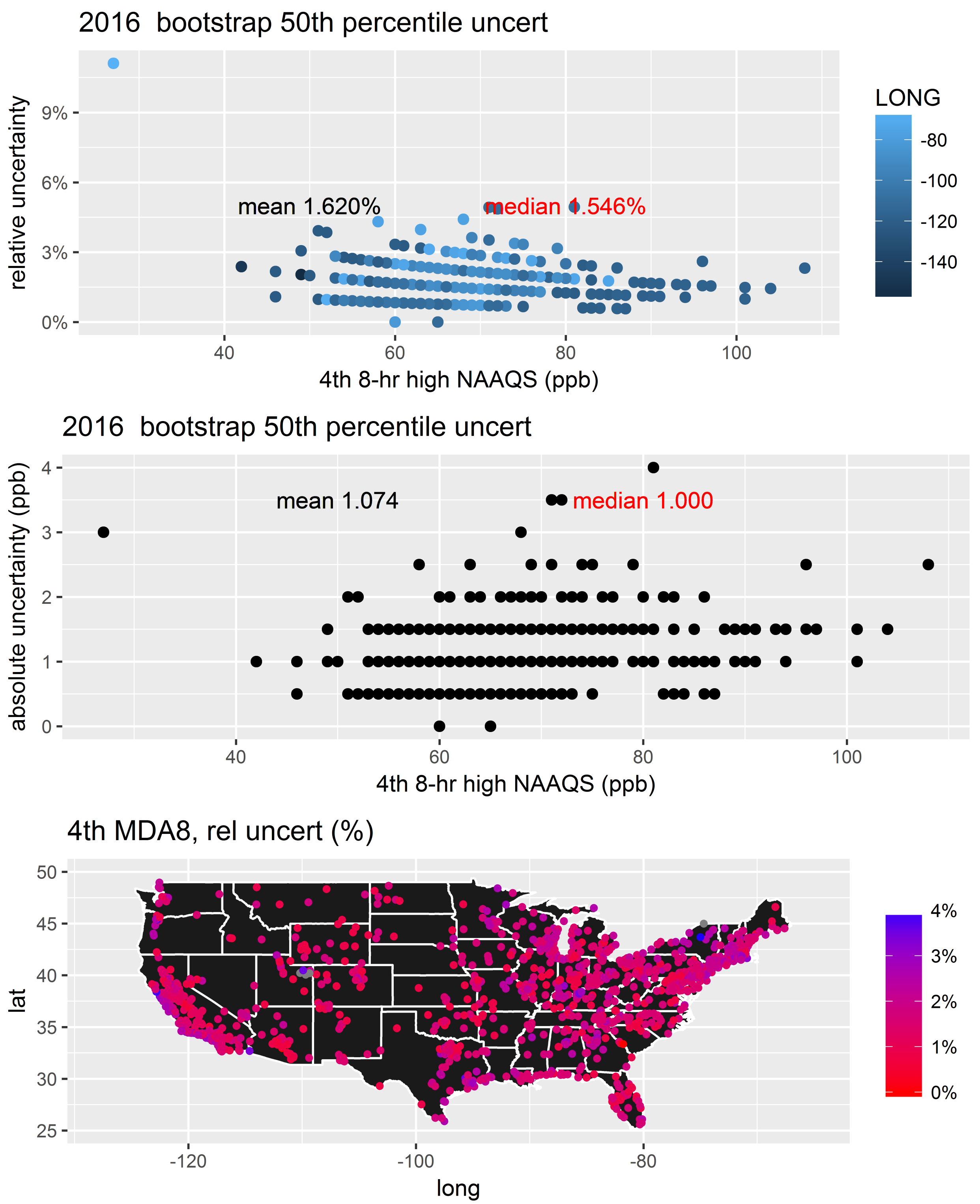


Figure 178: Blocked-bootstrap results from the 50% CIs for the 2016 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2015 ozone DV at each site.

7 Results from cluster analyses and other spatial groupings

This section presents results from several cluster analyses and other analyses conducted to examine the presence of spatial groupings or trends. If strong correlation in the variability can be found in natural spatial groupings, there may be reason to consider the variability at a regional, rather than national, level. The primary purpose of this analysis is to attempt to identify natural spatial groupings and determine if there strong correlations in the variability within these spatial groupings and if the variability between spatial groupings are significantly different. Since there is no clear pathway to determine the spatial correlations, the analysis presented here consists of several iterations of cluster analysis as well as an analysis of variability based on well-established climate regions to explore this issue from various perspectives.

7.1 Cluster analyses

Cluster analysis is an analysis technique that attempts to group data by similar characteristics of the data in question. This is generally done by assigning quantitative values to each characteristic and measuring and minimizing the "distance" between the existing clusters. The "distance" parameter can be calculated in a variety of ways, but the most common (and the one used here) is simply the Euclidian distance between the input variables. Two types of clustering algorithms are applied, a K-means algorithm and a hierarchical algorithm. The K-mean algorithm uses a pre-determined number of clusters and initially randomly assigns all items to clusters. The distance between cluster centers and all individuals are calculated, then individuals are reassigned to their closest cluster. The algorithm repeats a set number of times or until a minimum convergence threshold is reached. Hierarchical algorithms do not use a predetermined a number of clusters, but instead start with each individual as part of their own cluster. The first step in a hierarchical analysis combines the two closest clusters (which are just the two closest members at the first step). Each subsequent step combines the next closest clusters, until only 2 clusters are left. The R software package [R Core Team, 2017] was used to conduct the cluster analysis, using the kmeans and hclust functions. The analysis was performed on the results from the 2014-2016 PM variability results, as described in the following sections.

7.1.1 Cluster analysis with latitude, longitude, and variability values

This cluster analysis used the latitude, longitude (both in degrees eastwest and northsouth), and the relative variability (as a percentage of the site's DV). Thus, the distance between individuals and clusters is defined as the difference between the latitude, longitude, and relative variability. Since the longitudes and latitude varies on a much larger scale between sites (longitude ranges from -64 to -160 degrees, latitude ranges from 17 to 64 degrees) than the relative variability (0-5 percent for the annual and 0-75 percent for the daily DVs), the spatial input component will have a greater impact on the resulting than the site-level variability (clusters for the annual and daily DVs were computed separately). That is, the spatial closeness will be the primary factor in forming these clusters, but the analysis will then try to group nearby sites with similar levels of variability. Hierarchical and K-mean clustering were applied independently.

The clusters formed from this analysis is shown in Figures 179 and 180 and statistics are summarized in Tables 1- 4. The K-means analysis used 10 clusters, which was picked based on the number of EPA Regions. The figure also shows the hierarchical cluster results at 10 clusters for comparison. The clusters from the hierarchical analysis have relatively little recognizable geographic correlation. For example, cluster 1 (orange circles in both the annual and 24-hr figures) consists of a group of sites over California and Arizona and a group over the south eastern US (Florida, Georgia, Alabama), with a major discontinuity in this grouping, with no data points in New Mexico, Texas, Louisiana, and Mississippi. Table 1 and 2 show the statistics from the hierarchical clusters for the annual and 24-hr standards. The table includes a comparison of the mean variability from each cluster to the mean from the entire dataset using a Welch Modified Two-Sample t-Test (determined from the tsum.test function from the BSDA package in R) to determine if the means are significantly different. For the annual standard, the p values are all fairly high, with the smallest value just over 0.1, which is well above the nominal p value of 0.5 typically identified as an indicator that the means may be different. For the 24-hr standard, there are 2 clusters with p values less than 0.05. Cluster 7 has a p value of 0.02, which may be different than the annual mean, but the sites in this cluster (light blue squares with an "x") are spread across the country, i.e., they are not spatially distinct. Cluster 8 has the smallest p value (0.006) and has the smallest mean variability. For the most part, this cluster is in the same region (purple asterisk), in the eastern US, from North Carolina up to New York. However, this cluster is interspersed with several other clusters. Thus, while it has distinct variability values and is spatially correlated, it is not spatially distinct.

The results from the K-mean cluster analysis are starkly different from the hierarchical analysis. The clusters are all geographically distinct and the results of the t-test indicate that several of the clusters are distinctly different from the mean dataset. For the annual standard, half of the clusters have p-values less than 0.05 (2, 3, 4, 8, and 10) while 7 clusters have p-values less than 0.05 for the daily standard (1, 2, 3, 4, 5, 6, and 9, though cluster 1 and 9 only have a few members and cluster 2 is close enough to 0.05 to discount as significantly different, leaving only 4 clusters of note). On the surface, this suggests there are regional differences in the variability. However, the differences between the results from the annual and daily standards suggest the result is less certain. For example, cluster 4 in the annual analysis stands out as having the largest mean variability and a very small p-value, suggesting the variability in this subset is significantly different from the mean dataset. However, these sites are part of a larger cluster in the daily results (cluster 8), which include California sites, and has lower mean variability than the mean from the dataset (though not significantly different, with a p-value of 0.31). Another example of inconsistency between the annual and daily results is cluster 3 in the annual results, which roughly correlates to cluster 6 in the daily results. In this case, the clusters represent approximately the same geographic region. However, for the annual result, cluster 3 has mean variability that appears to be significantly higher than the mean dataset's (p-value of 0.016), but significantly lower mean variability for the daily standard than the mean dataset's (p-value of 0.018). Thus, these particular geographic areas have higher than average variability in the long-term, but lower than average variability in the short-term. The inconsistent results from the K-means analysis make it difficult draw specific conclusions about the geographic nature of the variability as estimated by this analysis.

Table 1: Comparison of hierarchical clusters for lat-long-annual variability

n.sites	grp	mean	median	sd	ann.pval
107	1	1.628	1.538	0.7296	0.1969
44	2	1.917	1.786	0.7377	0.1052
43	3	1.759	1.724	0.7835	0.7972
57	4	1.759	1.744	0.5889	0.7024
43	5	1.659	1.630	0.5025	0.4108
44	6	1.745	1.765	0.4914	0.8207
73	7	1.736	1.754	0.5604	0.8929
26	8	1.674	1.714	0.6251	0.6762
49	9	1.816	1.744	0.7380	0.4171
38	10	1.660	1.453	0.8483	0.6384
524	All	1.727	1.705	0.6700	1.0000

Table 2: Comparison of hierarchical clusters for lat-long-24-hr variability

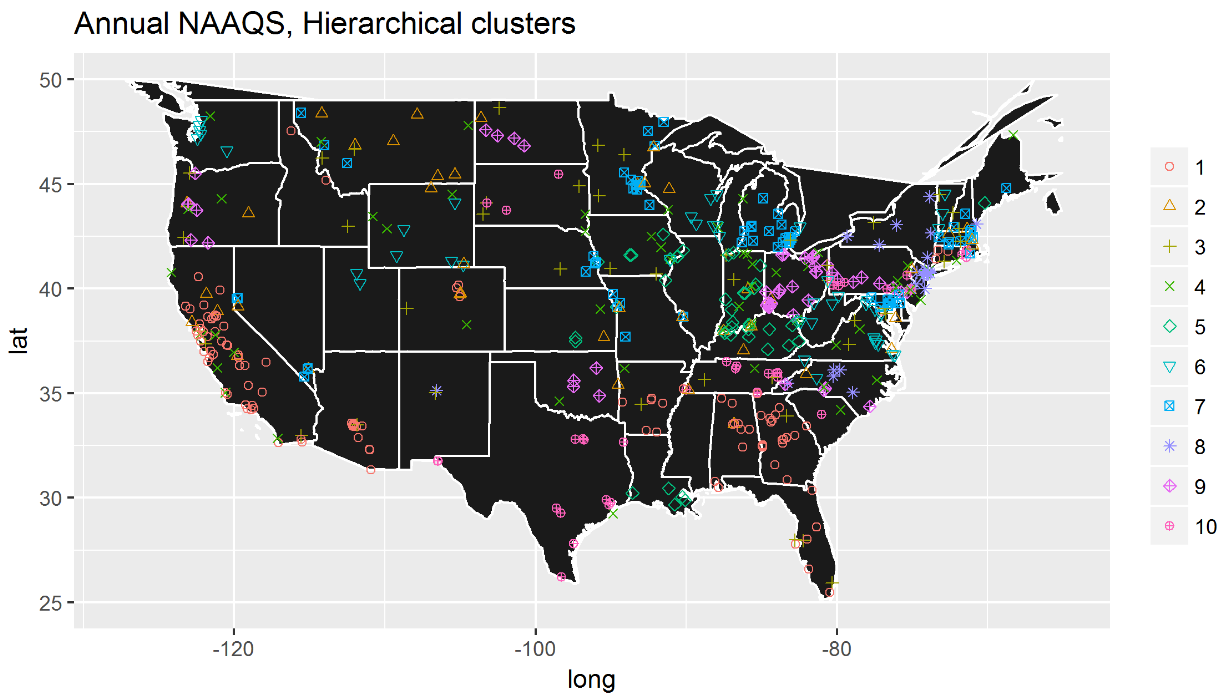
n.sites	grp	mean	median	sd	TF.pval
107	1	5.985	3.846	8.882	0.718242
44	2	6.396	5.013	5.046	0.365465
43	3	6.339	5.556	6.332	0.501141
57	4	5.150	4.054	3.758	0.370173
43	5	6.192	4.762	6.745	0.618803
44	6	5.098	4.692	3.403	0.336031
73	7	4.720	4.167	2.598	0.021686
26	8	4.338	4.006	1.917	0.005972
49	9	5.989	4.545	8.297	0.787386
38	10	6.217	4.583	6.185	0.594067
524	all	5.659	4.447	6.163	1.000000

Table 3: Comparison of K-means clusters for lat-long-annual variability

n.sites	grp	mean	median	sd	ann.pval
61	1	1.634	1.613	0.5826	2.508e-01
26	2	1.425	1.291	0.5835	1.618e-02
96	3	1.863	1.796	0.4657	1.599e-02
26	4	2.487	2.395	0.7843	4.550e-05
58	5	1.661	1.658	0.5383	3.941e-01
38	6	1.619	1.695	0.6162	3.044e-01
57	7	1.573	1.471	0.5912	7.068e-02
66	8	1.423	1.373	0.4362	2.550e-06
57	9	1.663	1.471	0.8888	6.006e-01
39	10	2.266	2.174	0.7903	1.556e-04
524	all	1.727	1.705	0.6700	1.000e+00

Table 4: Comparison of K-means clusters for lat-long-24-hr variability

n.sites	grp	mean	median	sd	TF.pval
3	1	61.111	58.333	12.729	1.701e-02
59	2	4.794	4.762	2.367	3.604e-02
44	3	9.083	8.477	3.979	2.320e-06
104	4	3.807	3.333	2.131	8.750e-08
33	5	4.041	2.941	3.060	9.146e-03
109	6	4.751	4.545	2.849	1.831e-02
38	7	4.994	4.762	2.825	2.154e-01
68	8	5.165	4.202	3.342	3.111e-01
7	9	30.768	30.952	9.002	3.049e-04
59	10	5.020	3.846	3.136	1.938e-01
524	all	5.659	4.447	6.163	1.000e+00



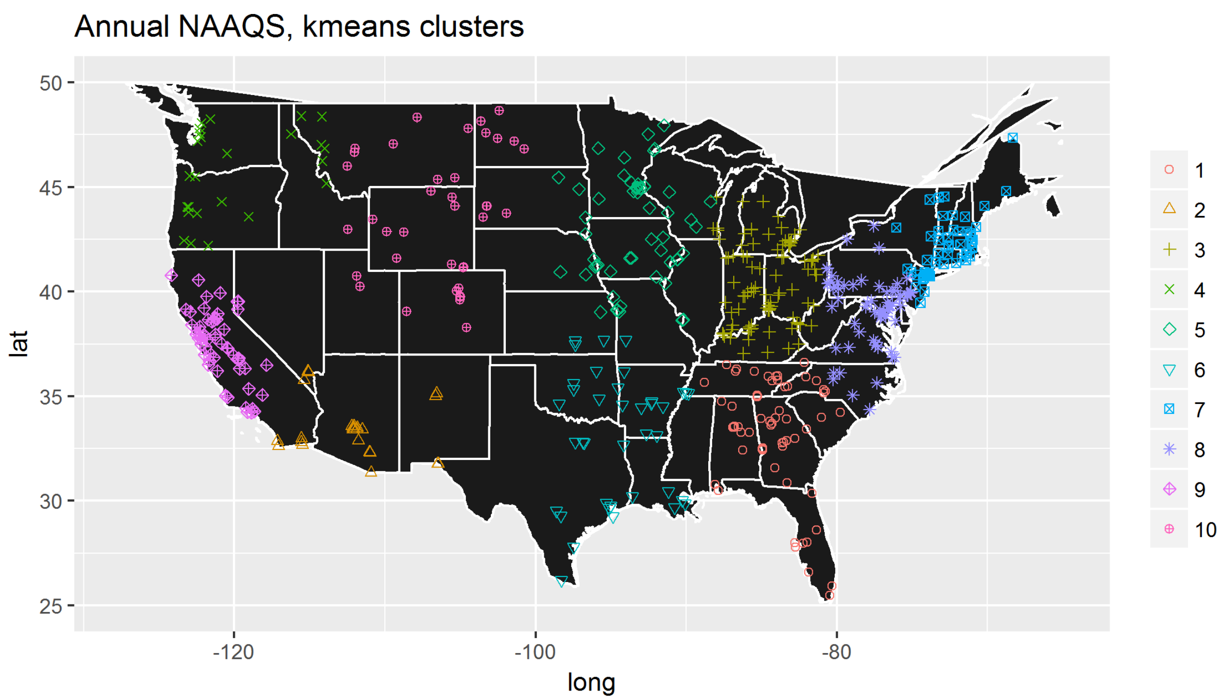


Figure 179: Hierarchical and K-means clusters for the annual variability for 2016.

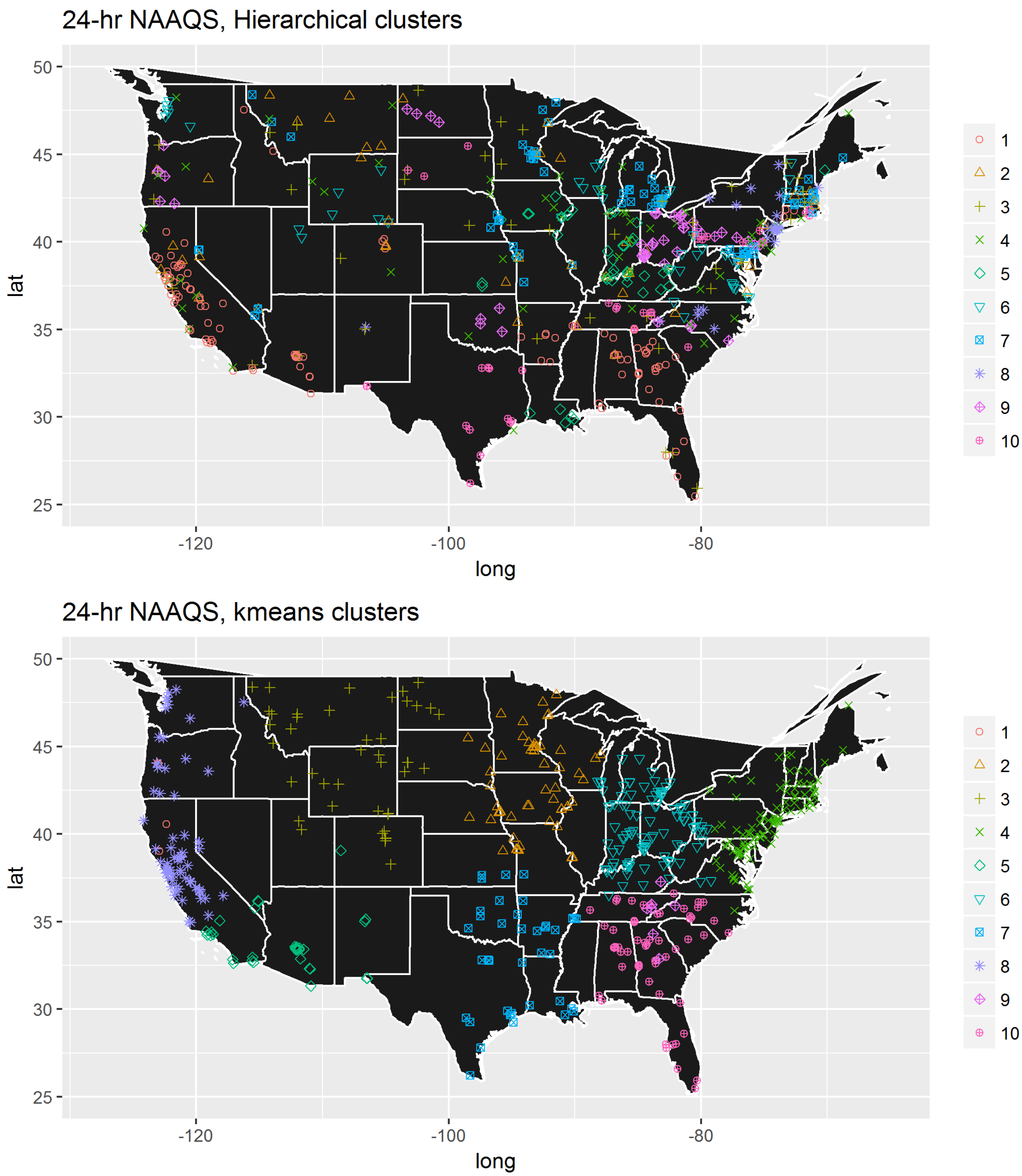


Figure 180: Hierarchical and K-means clusters for the 24-hr variability for 2016.

7.1.2 Cluster analysis with time series of variability values (2014-2016)

This cluster analysis used the relative variability (as a percentage of the site's DV) from each site over 3 DV periods (2014-2016). Thus, the distance between individuals and clusters is defined as the difference between each year's variability values (i.e., the variability from 2014 data, the variability from 2015 data, the variability from 2016 data) for a particular standard. Unlike the previous analysis, all input variables are on the same scale, such that no one parameter is driving the cluster formation. Therefore, this analysis attempts to group sites with similar levels of variability over time in order to see if those variability trends have spatial correlation. Since this approach incorporates the variability over time, it reflects the final composite variability value determined in the main analysis, which is the average over 3 DV periods. Hierarchical and K-mean clustering were applied independently.

The clusters formed from this analysis is shown in Figures 181 and 182 and statistics are summarized in Tables 5- 8. The K-means analysis used 10 clusters, which was picked based on the number of EPA Regions. The figure also shows the hierarchical cluster results at 10 clusters for comparison. As with the latlongvariability analysis presented in the previous section, the clusters from the hierarchical analysis have relatively little recognizable geographic correlation. However, most of the clusters have mean variability levels that are distinctly different from the mean dataset (note that the mean values presented here represent the mean from all years), though this approach also resulted in more clusters with very few members, such that the annual results only had 5 clusters with 20 or more members and the 24-hr results only had 3 clusters with 20 or more members. The spatial distribution of the results from the K-means analysis was similar to the hierarchical results, in that relatively little recognizable geographic correlation. However, the K-mean algorithm resulted in more meaningful clusters in terms of number of members and the statistical significance. Thus, while the cluster analysis conducted with the 3-year variability trends resulted in groups that were distinct with respect to their variability levels, it showed essentially no spatial correlation, suggesting that geographic differences in variability do not need to be taken into account.

Table 5: Comparison of hierarchical clusters for 2014-2016 variability, annual

n˙sites	grp	mean	median	sd	ann˙pval
86	1	1.027	1.053	0.2387	1.418e-49
84	2	1.475	1.449	0.2964	9.408e-08
30	3	2.422	2.381	0.2443	2.756e-19
147	4	1.877	1.852	0.2730	1.498e-05
20	5	1.488	1.559	0.4272	3.258e-02
7	6	2.185	2.439	0.8897	2.129e-01
12	7	3.017	3.008	0.3734	3.841e-08
9	8	3.099	3.125	0.4706	1.585e-05
1	9	4.843	5.128	0.8896	NA
2	10	3.987	4.139	0.5126	9.888e-02
398	11	1.716	1.666	0.5855	1.000e+00

Table 6: Comparison of hierarchical clusters for 2014-2016 variability, 24-hr

n˙sites	grp	mean	median	sd	TF˙pval
272	1	3.630	3.333	1.662	1.499e-08
49	2	8.034	7.895	2.365	1.311e-11
58	3	4.848	4.545	2.134	7.556e-01
9	4	11.821	11.765	2.320	1.106e-05
4	5	9.097	5.409	7.870	3.701e-01
1	6	18.480	20.312	4.399	NA
1	7	55.556	75.000	41.107	NA
1	8	25.966	29.630	7.354	NA
2	9	13.220	14.460	6.831	3.363e-01
1	10	41.250	46.667	20.340	NA

n˙sites	grp	mean	median	sd	TF˙pval
398	11	4.957	4.129	4.143	1.000e+00

Table 7: Comparison of K-means clusters for 2014-2016 variability, annual

n˙sites	grp	mean	median	sd	ann˙pval
23	1	1.027	1.053	0.2387	1.068e-14
3	2	1.475	1.449	0.2964	2.939e-01
61	3	2.422	2.381	0.2443	1.781e-38
66	4	1.877	1.852	0.2730	3.753e-04
48	5	1.488	1.559	0.4272	1.382e-03
5	6	2.185	2.439	0.8897	3.038e-01
68	7	3.017	3.008	0.3734	4.822e-50
28	8	3.099	3.125	0.4706	3.779e-16
27	9	4.843	5.128	0.8896	9.066e-17
69	10	3.987	4.139	0.5126	2.528e-56
398	11	1.716	1.666	0.5855	1.000e+00

Table 8: Comparison of K-means clusters for 2014-2016 variability, 24-hr

n˙sites	grp	mean	median	sd	TF˙pval
2	1	3.630	3.333	1.662	4.577e-01
59	2	8.034	7.895	2.365	2.016e-13
44	3	4.848	4.545	2.134	7.770e-01
104	4	11.821	11.765	2.320	4.609e-65
37	5	9.097	5.409	7.870	3.098e-03
46	6	18.480	20.312	4.399	1.172e-26
2	7	55.556	75.000	41.107	3.319e-01
22	8	25.966	29.630	7.354	6.561e-12
73	9	13.220	14.460	6.831	7.360e-16
9	10	41.250	46.667	20.340	6.812e-04
398	11	4.957	4.129	4.143	1.000e+00

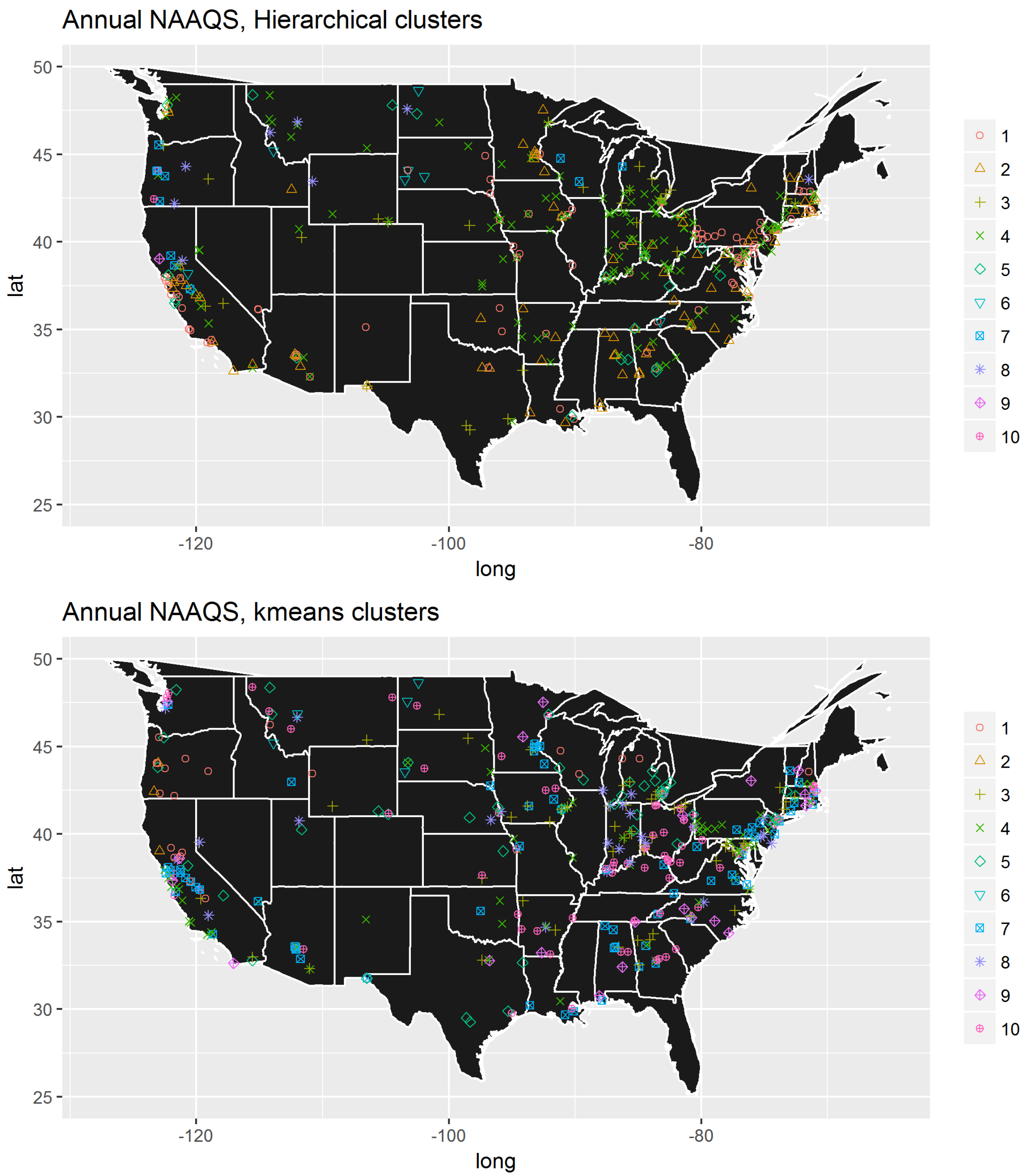


Figure 181: Hierarchical and K-means clusters for the annual variability for 2014-2016.

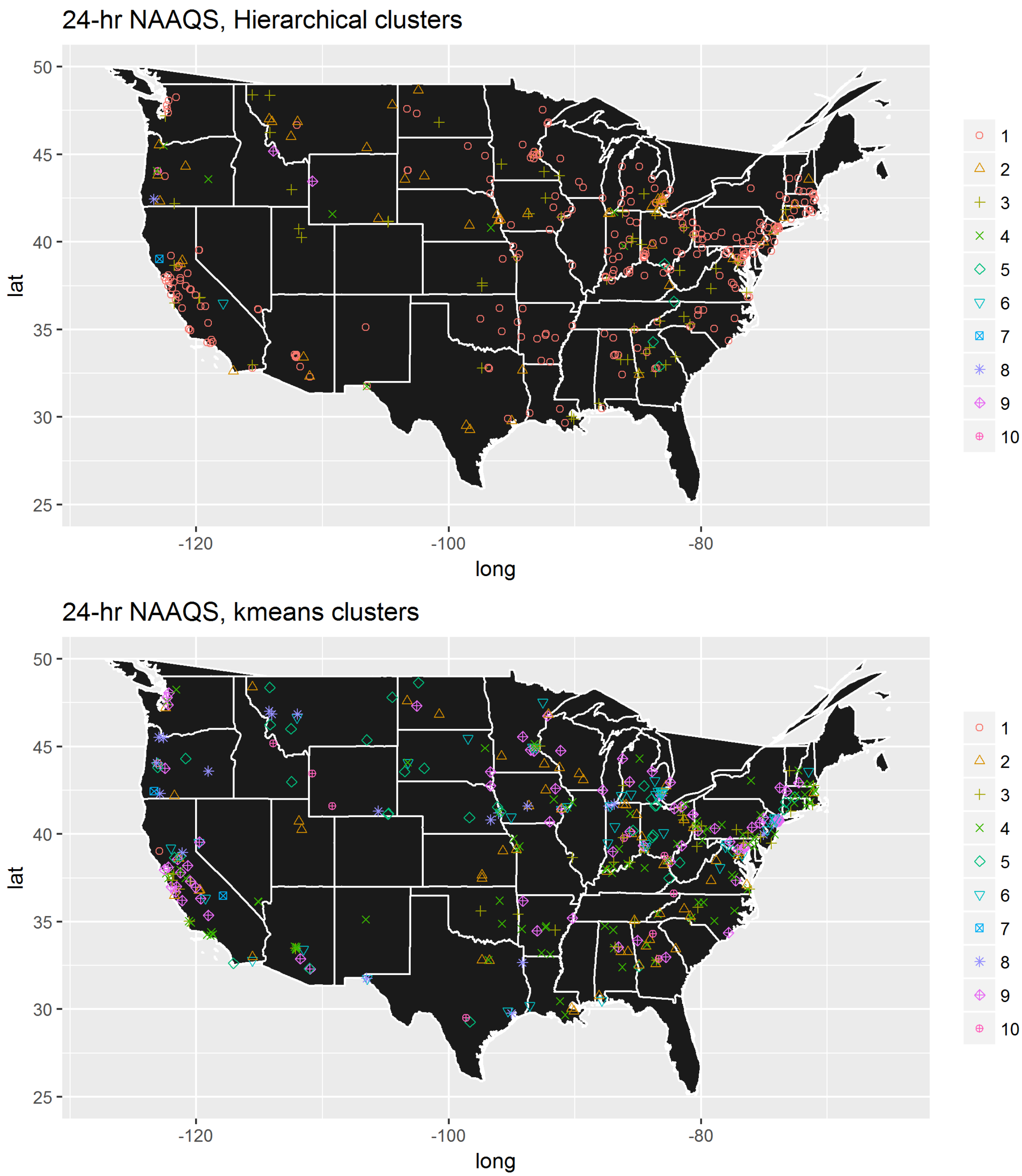


Figure 182: Hierarchical and K-means clusters for the 24-hr variability for 2014-2016.

7.1.3 Cluster analysis with time series of variability values (2012-2016)

This cluster analysis used the relative variability (as a percentage of the site's DV) from each site over 5 DV periods (2012-2016). Thus, the distance between individuals and clusters is defined as the difference between each year's variability value (i.e., the variability from 2012 data, ..., the variability from 2016 data) for a particular standard. Unlike the previous analysis, all input variables are on the same scale, such that no one parameter is driving the cluster formation. Therefore, this analysis attempts to group sites with similar levels of variability over time in order to see if those variability trends have spatial correlation. Since this approach incorporates the variability over time, it partly reflects the final composite variability value determined in the main analysis, which is the average over 3 DV periods. The extended period is evaluated in addition to the 3 DV periods presented above in order to improve correlations that may exist with a longer data record. Hierarchical and K-mean clustering were applied independently.

The clusters formed from this analysis is shown in Figures 183 and 184 and statistics are summarized in Tables 9- 12. The K-means analysis used 10 clusters, which was picked based on the number of EPA Regions. The figure also shows the hierarchical cluster results at 10 clusters for comparison. The results from this analysis are fairly similar to the results using the 3 DV periods. The clusters are not spatially distinct; many clusters have few members, though most have distinct variability levels. Thus, while the cluster analysis conducted with the 5-year variability trends resulted in groups that were distinct with respect to their variability levels, it showed essentially no spatial correlation, suggesting that geographic differences in variability do not need to be taken into account.

Table 9: Comparison of hierarchical clusters for 2012-2016 variability, annual

n˙sites	grp	mean	median	sd	ann˙pval
76	1	1.138	1.099	0.2967	1.272e-27
30	2	2.292	2.273	0.3028	6.841e-12
55	3	1.533	1.531	0.3474	9.338e-04
105	4	1.809	1.802	0.3136	6.702e-02
1	5	3.295	2.632	1.2706	NA
16	6	2.929	2.871	0.3333	1.926e-11
2	7	3.691	3.562	1.0601	2.315e-01
1	8	3.117	3.670	0.8943	NA
1	9	4.096	4.348	0.5319	NA
3	10	2.185	2.232	0.7266	3.884e-01
290	11	1.727	1.678	0.5550	1.000e+00

Table 10: Comparison of hierarchical clusters for 2012-2016 variability, 24-hr

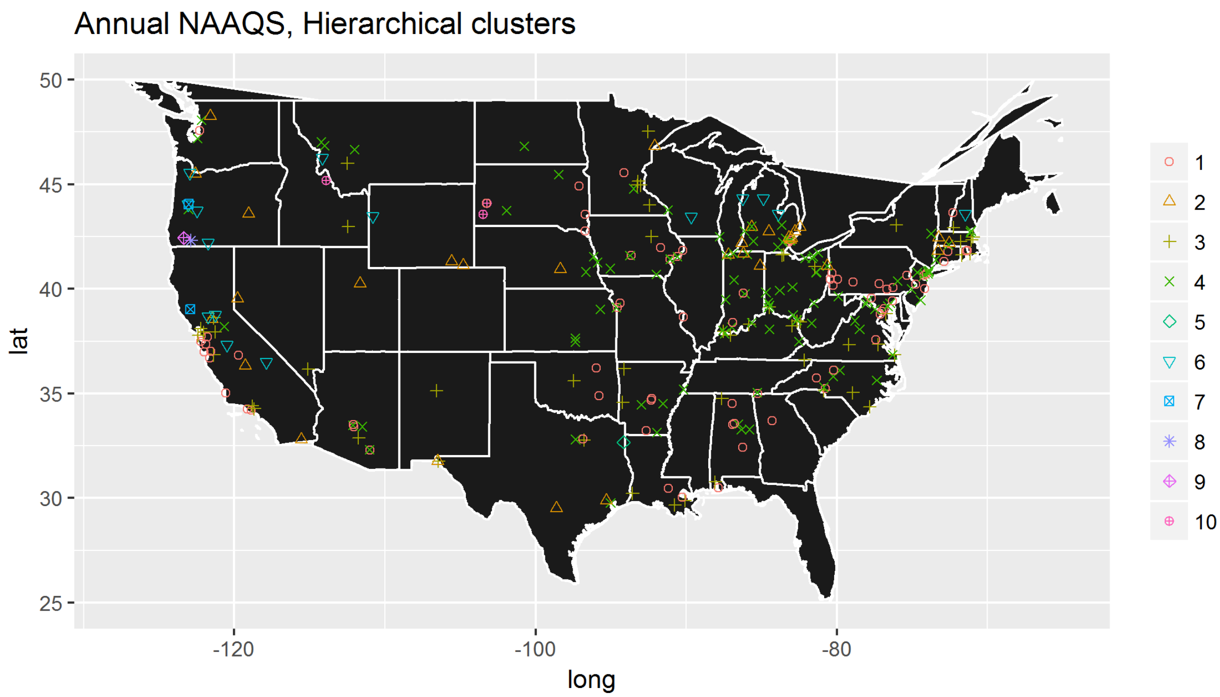
n˙sites	grp	mean	median	sd	TF˙pval
193	1	3.460	3.226	1.469	6.142e-09
25	2	7.603	7.143	2.917	1.610e-04
32	3	5.251	4.762	2.776	5.508e-01
33	4	7.187	7.143	2.787	1.074e-04
1	5	36.947	10.000	37.861	NA
1	6	13.357	7.692	10.734	NA
2	7	14.391	12.812	5.456	2.455e-01
1	8	36.444	10.000	39.143	NA
1	9	23.033	28.571	9.620	NA
1	10	27.333	18.750	23.875	NA
290	11	4.927	4.132	3.802	1.000e+00

Table 11: Comparison of K-means clusters for 2012-2016 variability, annual

n˙sites	grp	mean	median	sd	ann˙pval
46	1	1.138	1.099	0.2967	1.113e-18
39	2	2.292	2.273	0.3028	5.306e-15
56	3	1.533	1.531	0.3474	8.645e-04
51	4	1.809	1.802	0.3136	1.358e-01
34	5	3.295	2.632	1.2706	2.922e-08
3	6	2.929	2.871	0.3333	2.210e-02
3	7	3.691	3.562	1.0601	8.448e-02
21	8	3.117	3.670	0.8943	5.988e-07
21	9	4.096	4.348	0.5319	5.424e-16
16	10	2.185	2.232	0.7266	2.442e-02
290	11	1.727	1.678	0.5550	1.000e+00

Table 12: Comparison of K-means clusters for 2012-2016 variability, 24-hr

n˙sites	grp	mean	median	sd	TF˙pval
1	1	3.460	3.226	1.469	NA
72	2	7.603	7.143	2.917	1.189e-09
67	3	5.251	4.762	2.776	4.262e-01
20	4	7.187	7.143	2.787	2.264e-03
26	5	36.947	10.000	37.861	2.218e-04
75	6	13.357	7.692	10.734	2.885e-09
2	7	14.391	12.812	5.456	2.455e-01
3	8	36.444	10.000	39.143	2.978e-01
2	9	23.033	28.571	9.620	2.285e-01
22	10	27.333	18.750	23.875	2.491e-04
290	11	4.927	4.132	3.802	1.000e+00



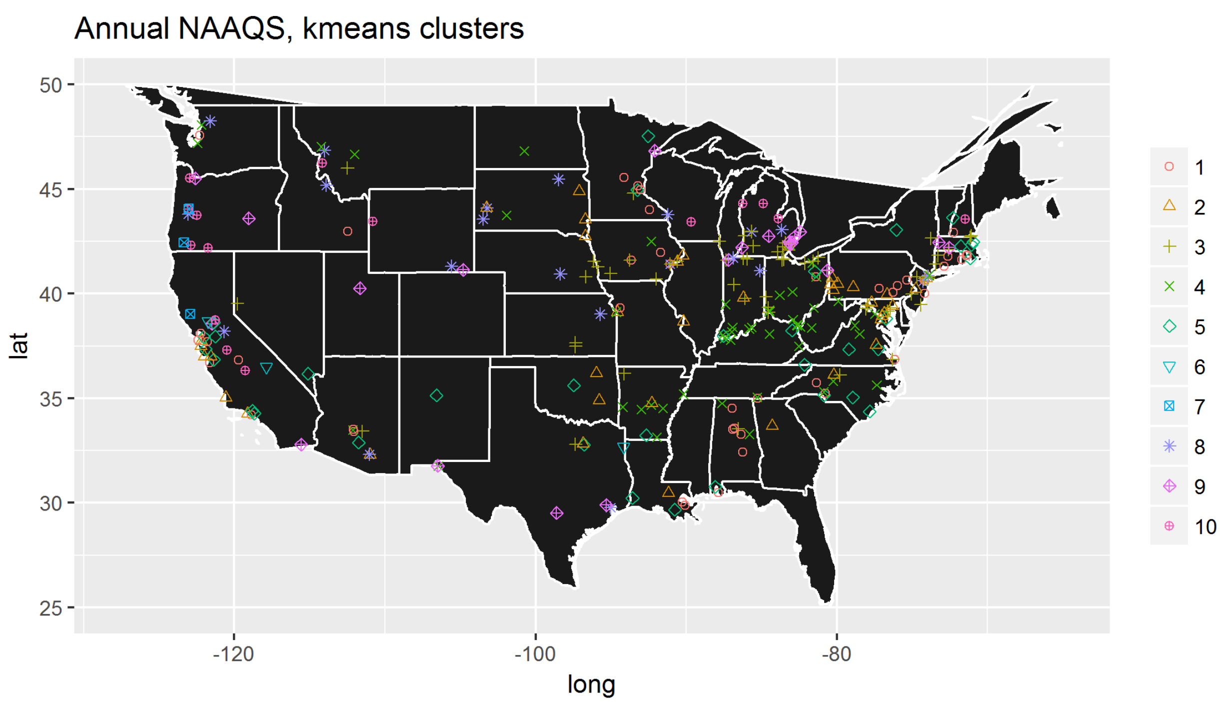
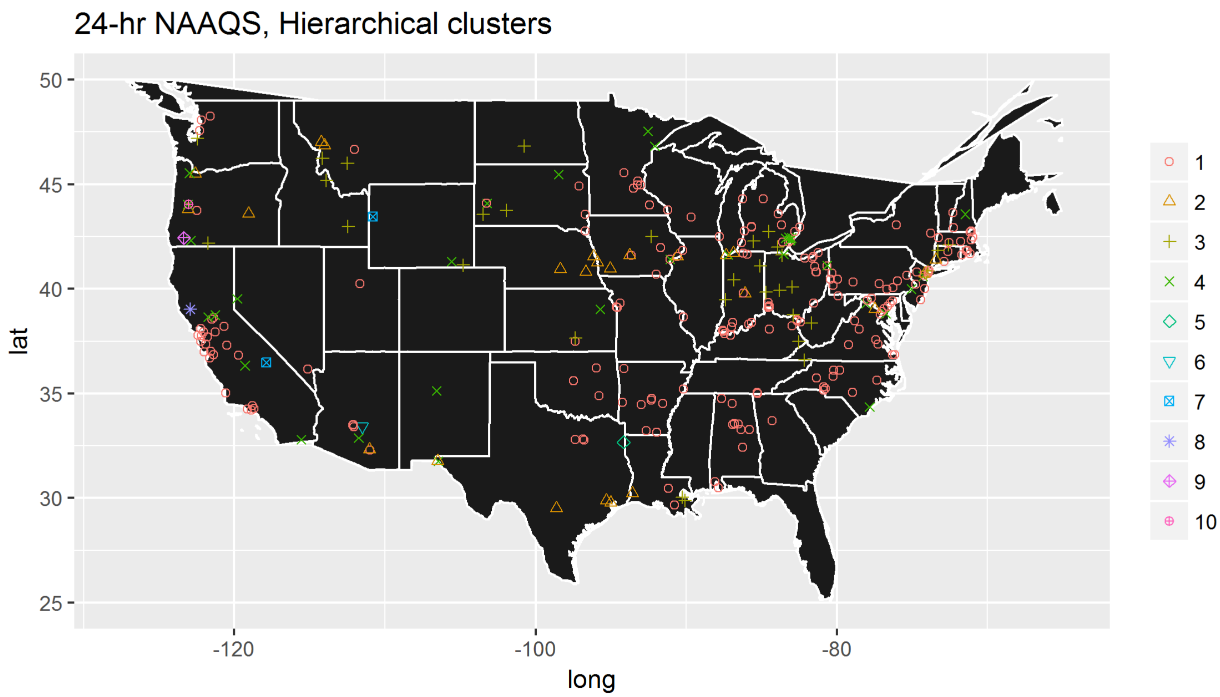


Figure 183: Hierarchical and K-means clusters for the annual variability for 2012-2016.



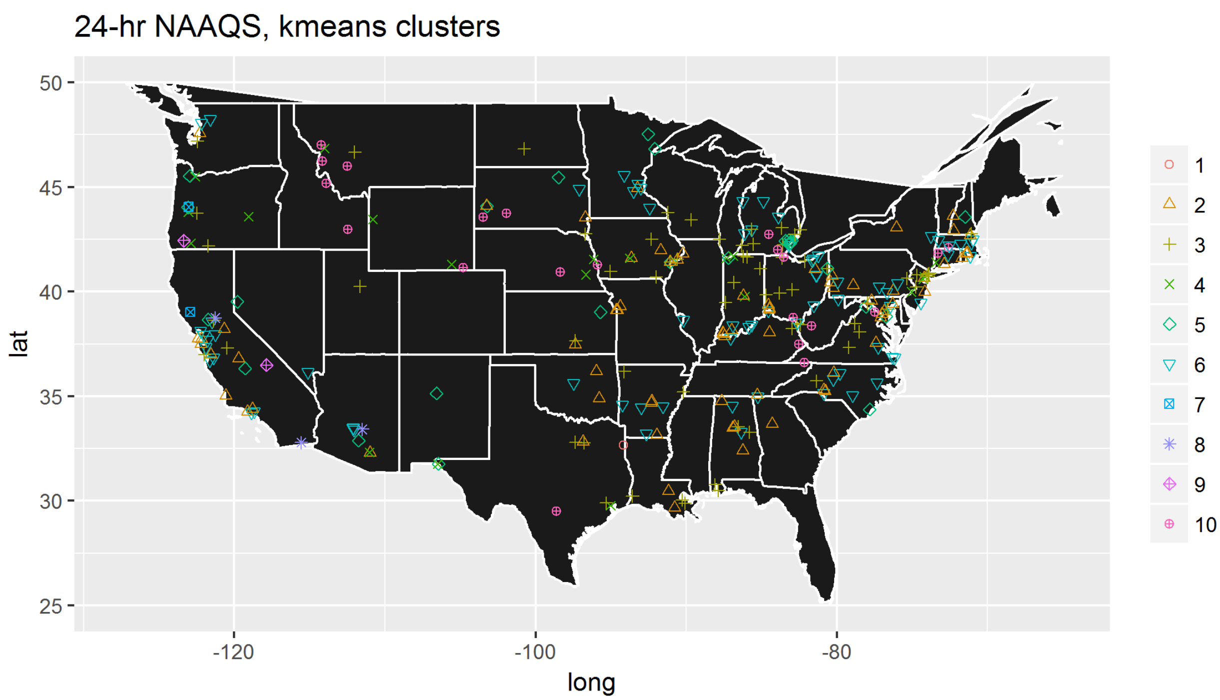


Figure 184: Hierarchical and K-means clusters for the 24-hr variability for 2012-2016.

7.2 Spatial analysis using NOAA Climate Regions

The National Oceanic and Atmospheric Administration (NOAA) has identified 9 "climate regions" [Thomas and Koss, 1984], which have been identified to have distinct climatologically characteristics (more information available at the NOAA website). This spatial grouping thus represents an independent spatial grouping with which to evaluate regional variability characteristics. The mean annual and 24-hr variability values from sites within these regions are compared in Figure 185 and detailed in Tables 13 and 14. The Pacific Northwest (region 4) and the Central Northwest (region 9) stand out as having higher variability, which was seen in the first K-means cluster analysis (using latitude, longitude, and the variability). The p-values for the annual results are less than 0.05 for these two regions (though the p-value for region 9 is just barely less than 0.05 and the p-value for region 4 is still relatively large). The p-values for these two regions from the 24-hr results are well above the nominal value of 0.05 and so are not significantly different from the mean dataset. Thus, the results again make it difficult draw specific conclusions about the geographic nature of the variability as estimated by this analysis, though the overall interpretation of these results is that most regions are not significantly different from the mean dataset.

Table 13: Comparison of variability within NOAA climate regions, annual

Region Number	n sites	ann mean	ann median	ann sd	ann pval
1	96	1.724119413	1.724137931	0.538832598	0.986301656
2	69	1.864963958	1.875	0.501458469	0.547340281
3	94	1.457900874	1.388888889	0.545905166	0.245378188
4	22	2.473011166	2.405978785	0.845285667	0.013781951
5	39	1.719076105	1.704545455	0.622791499	0.970268689
6	70	1.563384611	1.583124478	0.459051349	0.469019357
7	25	1.623411728	1.595744681	0.621819606	0.676247393
8	66	1.610690042	1.405159932	0.84597489	0.627212461
9	42	2.255001105	2.198067633	0.808280411	0.047892639
All	523	1.727970703	1.727970703	0.670071968	1

Table 14: Comparison of variability within NOAA climate regions, 24-hr

Region Number	n sites	TF mean	TF median	TF sd	TF pval
1	96	5.861069073	4.545454545	6.966363904	0.924845491
2	69	4.690627066	4.545454545	1.965280928	0.632945983
3	94	3.628720758	3.125	2.092141946	0.326832786
4	22	10.13095985	7.417582418	11.55838373	0.165608839
5	39	5.280129178	5	3.051945505	0.853550834
6	70	5.893319944	4.653679654	5.146873277	0.911597545
7	25	5.383605813	4	3.836720273	0.896970478
8	66	5.93939837	3.923076923	10.5914199	0.90680014
9	42	8.69767221	7.692307692	3.979459168	0.166016078
All	523	5.660892679	4.347826087	6.168440212	1

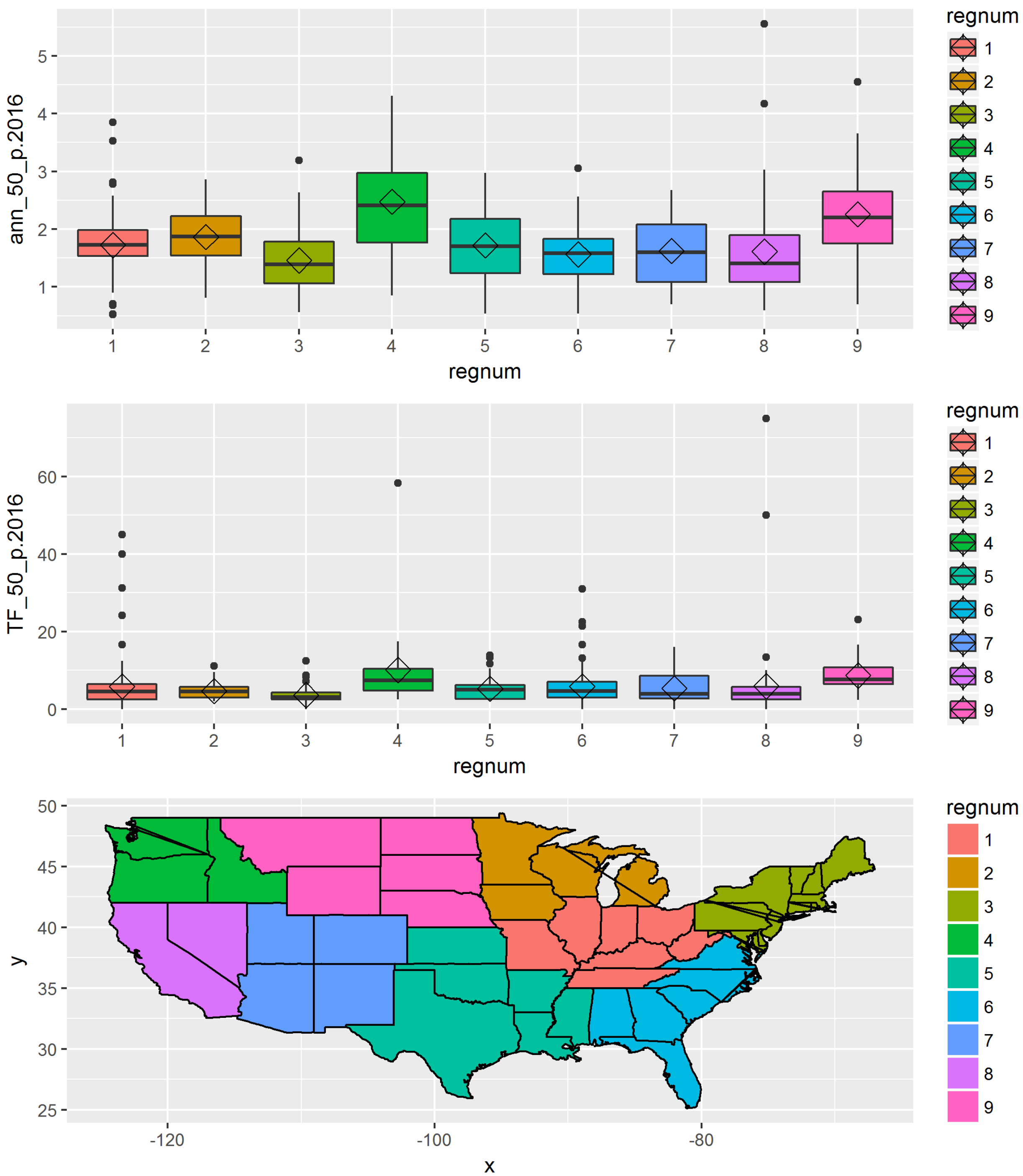


Figure 185: Comparison of variability within NOAA climate regions for 2016.

References

D. W. Andrews. The Block-block Bootstrap: Improved Asymptotic Refinements. Cowles Foundation Discussion Papers 1370, Cowles Foundation for Research in Economics, Yale University, May 2002. URL `https://ideas.repec.org/p/cwl/cwldpp/1370.html`.

P. Hall and J. L. Horowitz. Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64(4):891–916, 1996. ISSN 0165-0270. doi: 10.2307/2171849. URL `http://www.jstor.org/stable/2171849`.

A. Inoue and M. Shintani. Bootstrapping gmm estimators for time series. *Journal of Econometrics*, 133(2):531–555, 2006. URL `http://EconPapers.repec.org/RePEc:eee:econom:v:133:y:2006:i:2:p:531-555`.

S. N. Lahiri. Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404, 1999.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2017.

R. K. Thomas and W. J. Koss. Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983. *Historical Climatology Series 4-3*, 1984.

United States Environmental Protection Agency	Office of Air Quality Planning and Standards Air Quality Assessment Division Research Triangle Park, NC	Publication No. EPA-454/R-18-001 April, 2018

Peer Review Report for the Technical Basis for the EPA's Development of Significant Impact Thresholds for $PM_{2.5}$ and Ozone

EPA-454/S-18-001
March 2018

Peer Review Report for the Technical Basis for the EPA's Development of Significant Impact Thresholds for $PM_{2.5}$ and Ozone

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Assessment Division
Research Triangle Park, NC

Overview

As part of the OMB review process for the draft guidance document for $PM_{2.5}$ and ozone SILs[1], the EPA agreed to conduct a peer review of the technical basis document (TBD).[2] This summary of the peer review provides the charge questions supplied to the peer reviewers, a summary of the comments received from the reviewers, and overviews of changes made to the TBD and additional analyses conducted in response to reviewer comments.

Peer review process

The peer review was conducted under an EPA contract to the University of North Carolina at Chapel Hill that has been used for technical review purposes similar to this work in the past. The peer review was overseen by Dr. Sarav Arunachalam, Research Associate Professor with the Center for Environmental Modeling for Policy Development. The EPA provided a list of six potential reviewers, from which the contractor obtained agreements from three reviewers to conduct the peer review. The peer reviews were conducted by environmental statisticians on faculty at major U.S. universities. The three reviewers were (bios for each reviewer are provided in Appendix A to this document):

- **Candace Berrett, PhD**; Assistant Professor, Department of Statistics, Brigham Young University
- **Veronica Berrocal, PhD**; John G Searle Assistant Professor of Biostatistics, University of Michigan School of Public Health
- **Bo Li, PhD**; Associate Professor, Department of Statistics, University of Illinois at Urbana-Champaign

Charge questions

The charge questions were developed by EPA in consultation with OMB. The final set of charge questions sent to the reviewers were as follows:

1) Are the relevant technical aspects of the statistical procedure clearly described?
    a. Are input data (EPA's AQS) and their characteristics sufficiently described?
    b. Is it clear what is being estimated?
    c. Is the bootstrap procedure described in sufficient detail to allow reproduction?
2) Are the descriptions of statistical concepts clear and accurate?

[1] Guidance for Comment: Significant Impact Levels for Ozone and Fine Particle in the Prevention of Significant Deterioration Permitting Program, https://www.epa.gov/nsr/draft-guidance-comment-significant-impact-levels-ozone-and-fine-particle-prevention-significant

[2] Technical Basis for the EPA's Development of Significant Impact Levels for $PM_{2.5}$ and Ozone, Office of Air Quality Planning and Standards, RTP, NC, 2017, EPA 454/R-17-002.

    a. Are the descriptions of statistical significance and significance testing clearly and sufficiently described to assist the layperson in understanding the analysis?
    b. Do the examples provided in the TBD illustrate the concepts of statistics sufficiently for the layperson to understand the analysis?
3) Are the assumptions and choices in the analysis clearly described and supported?
    a. Are the assumptions and choices in the analysis sufficiently documented?
    b. Does the document sufficiently describe the sensitivity of results to the choices and assumptions in the analysis? For example, are the technical considerations that support the policy decision to aggregate the variability to a single national value clearly articulated?
4) Are the procedures appropriate for the analytical goals?
    a. Is bootstrapping an appropriate technique to quantify the variability in the air quality design value statistics? Is the bootstrapping analysis a reasonable approach to inform a policy determination of Significant Impact Levels (*i.e.*, threshold levels)?
5) In your assessment, is there need for further analysis or clarification? Do you have suggestions for improving the document?

The peer review occurred parallel to the public comment period, from August 1 through September 30, 2016. The peer reviewers were given approximately 30 days to review the package, which included all three SILs documents (*i.e.*, in addition to the TBD, the policy memo[3] and legal memo[4] were provided to the reviewers). Each individual peer reviewer provided their comments to the UNC contractor, who then anonymized and delivered the reviews to the EPA as PDF documents, similar to how peer review comments would be handled by a scientific journal. The peer review responses are provided in their entirety in Appendix B of this report.

Summary of reviewer responses

The reviewer comments were largely supportive of the TBD and the analysis presented therein.

Reviewer 1

Reviewer 1 offered a few editorial comments but was very supportive of the methods, presentation, and conclusions from the analysis. Their response to charge question 3b was particularly expressive:

"The bootstrap is applied appropriately, and the selection of 50% conference interval to obtain conservative SILs is reasonable. The selection of a single national value is not optimal considering the spatial variability, but taking the consistency of policy into consideration and given the fact that there are no large scale trends in ambient air variability are present, it is not

[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.

unreasonable to have a single national value. Using the median rather the mean provides a more robust SIL for NAAQS."

Reviewer 2

Reviewer 2 offered a few editorial comments but also had several comments related to spatial variability. In particular, in contrast to Reviewer 1, Reviewer 2 felt that the spatial variability and dependence was not sufficiently accounted for.

Reviewer 3

Reviewer 3 offered a few editorial comments but also had several comments related to clarity and specificity, particularly with respect to the statistical terminology. The reviewer also had one technical comment related to the considerations of temporal dependence on the sampling and how this was accounted for in the bootstrap technique.

Summary of responses to peer review and public comments

The EPA made a number of revisions to the TBD, including (1) updating the analysis to include more recent data, (2) editing a number of sections for clarity and accuracy, and (3) conducting new and updated analyses to investigate issues raised by the reviewers.

Updated analysis

The bootstrapping methods for $PM_{2.5}$ data processing for the calculation of the $PM_{2.5}$ design values were also adjusted slightly to better align the methods with standard practice for calculating design values. Specifically:

- The rounding conventions for calculating $PM_{2.5}$ design values were applied in accordance with the EPA's regulations.[5] The original document applied the appropriate truncation conventions for ozone (*i.e.*, truncate to the whole ppb)[6]; however, the rounding rules for $PM_{2.5}$ were not correctly applied (*i.e.*, round design values to the nearest whole μg/m^3 for the daily NAAQS and the nearest 10th of a μg/m^3 for the annual NAAQS).

[5] Appendix N to Part 50—Interpretation of the National Ambient Air Quality Standards for $PM_{2.5}$.
[6] Appendix U to Part 50 - Interpretation of the Primary and Secondary National Ambient Air Quality Standards for Ozone.

- The selection of the 98th percentile value for the daily $PM_{2.5}$ value was corrected to use Table 1 provided in the CFR[5] rather than calculating the 98th percentile value based on the number of samples.

These updates had no impact on the recommended annual $PM_{2.5}$ SIL values (0.2 μg/m^3), while the daily $PM_{2.5}$ SIL value increased slightly, from 1.3 μg/m^3 to 1.5 μg/m3, primarily due to the updated 98th percentile selection approach, rather than the application of the rounding concentrations.

Editorial comments

Updates to the TBD were made to address editorial comments from all three peer-reviewers as well as in response to comments received during the public comment period. The majority of these were minor edits so they are not highlighted here but reflected in the revised TBD. However, significant editorial changes were made in section 4.1 in response to both Reviewer 3 and public comments. Section 4.1 was heavily revised with much of the discussion moved to the policy document in order to clarify the difference in the technical analysis and the policy choices made from the available options derived from the technical data. Specifically, we updated section III of the policy document to more clearly describe what information informed the selection of the EPA recommended SIL values and what the policy decision was based upon.

In response to Reviewer 3, the EPA conducted an additional analysis that examined the impact of persistence in ambient concentrations (*i.e.*, concentrations on one day being similar to concentrations on the following or previous day, which could occur due to similar meteorological conditions). The analysis focused on ozone because the EPA believes this pollutant would most likely show the impact of temporal persistence. While the EPA had conducted a form of this analysis during the development of the SILs package, the new analysis more rigorously analyzed the impact of the persistence of pollution events by analyzing the temporal correlation between ambient data at individual sites using standard statistical techniques and aggregated this correlation across the country. In simple terms, the analysis calculates correlation coefficients (using linear regressions) between data from day 1 with data from day 2, between data from day 1 and data from day 3, etc. The correlation between these pairs of days can inform the degree to which concentrations on a particular day can be predicted by concentrations from the previous days and how long pollution events might typically occur. The lag found from the correlation analysis (*i.e.*, 7 days) was used to conduct a block-sampling of the data and a re-run of the bootstrap analysis. The block sampling modified the bootstrap analysis to include the 3 days before and after each randomly selected day, such that blocks of consecutive days were analyzed. This procedure, thus, accounts for any temporal persistence that may be present in the air quality variability. The results at the 50% confidence interval were minimally different from the original, non-parametric analysis that assumed no lag. This additional analysis and the results are documented in section 6 of the appendix to the TBD.

Reviewer 3 also made specific comments on the spatial correlation among sites; in particular, that they did not agree with the EPA's assertion that there is not a significant spatial correlation among sites. Reviewers 1 and 2 also commented on the presence of a correlation between the spatial variability. Reviewer 1 specifically commented that there were no large scale trends, which was also the EPA's conclusion. In general, the EPA believes that the disagreement by Reviewer 3 is a matter of phrasing in the original TBD. There is a spatial correlation in both ozone and $PM_{2.5}$ in that most areas show relatively small variability and that there is not a strong spatial correlation in the location of sites with high variability. The document was revised to emphasize our intent in describing the spatial correlation. However, we also conducted several analyses to explore the existence of spatial groupings, *i.e.*, to determine if there are natural grouping of monitors with similar levels of variability. Three separate analyses were conducted, as follows:

- A cluster analysis was done using the latitude, longitude, and variability at each site in order to allow spatial variables to form natural groupings with similar levels of air quality variability.

- The NOAA climate regions were used to segregate data into predefined spatial groups based on similar weather patterns. The air quality variability of each climate region was then compared on a region-to-region basis and with the data aggregated to the national level to determine if the subsets were quantitatively different from one another.

Each analysis was conducted separately based on the air quality variability from both the annual and 24-hr $PM_{2.5}$ standards for the 2014-2016 data. The first two analyses were conducted using a "K-means" clustering algorithm and a hierarchical clustering algorithm. The K-means algorithm uses a pre-determined number of clusters and initially randomly assigns all sites to clusters. The difference between the cluster centers and all individuals are calculated, then sites are reassigned to the most similar cluster. The algorithm repeats a set number of times or until a minimum convergence threshold is reached. Hierarchical algorithms do not use a predetermined a number of clusters, but instead start with each site as part of their own cluster. The first step in a hierarchical analysis combines the two most similar clusters (which are just the two most similar sites at the first step). Each subsequent step combines the next closest clusters, until only two clusters are left, which contain all the individual sites.

The results of these additional analyses, which attempted to identify naturual groupings of sites based on similar levels of variability (*e.g.*, sites with consistently high variability), are presented in section 7 of the appendix to the TSD. The three separate analysis described above were conducted with each averaging period, resulting in 14 different sets of clusters. The results across these 14 sets of clusters varied widely. Several analyses did identify a unique region based on a specific clustering technique and averaging period, but these results were not consistent across clustering techniques or averaging periods. For example, the latitude, longitude, and variability analysis (first option in the list above) indicated several unique regions based on the 24-hr standard using the K-means algorithm. However, the K-means algorithm did not identify unique regions for the annual standard. Similarly, for this dataset, the hierarchical analysis identified sites with unique levels of variability for the 24-hr standard, but these sites were not spatially grouped (*e.g.*, the most unique group spanned at least 5 states, ranging from North Carolina to

Maine). Many of the analyses did not identify any unique groupings at all. When the results are considered as a whole, they support the EPA's original position that there are no large scale trends and that a national SIL is reasonable.

Appendix A: Peer reviewer bios

Candace Berrett, PhD; Assistant Professor, Department of Statistics, Brigham Young University
cberrett@stat.byu.edu; 801-422-7055; http://statistics.byu.edu/directory/berrett-candace

- Publications Chair, Section on the Environment, American Statistical Association, 2015-2016
- Program Chair, Environmental Sciences Section, International Society of Bayesian Analysis, 2014-2015
- Berrett, C. and Calder, C. A. (2016) "Bayesian spatial binary classification," Accepted for publication in Spatial Statistics.
- Co-PI, 2014, "Spatial Uncertainty: Data, Modeling, and Communication," National Institutes of Health (NIH).

Veronica Berrocal, PhD; John G Searle Assistant Professor of Biostatistics, University of Michigan School of Public Health

berrocal@umich.edu; 734-763-5965; https://sph.umich.edu/faculty-profiles/berrocal-veronica.html

Relevant Selected Publications:

- Professor of Spatial Statistics and Modern Statistical Methods, University of Michigan
- Young Investigator Award, Section on the Environment, American Statistical Association, 2015
- Chair, Section on Statistics and the Environment, American Statistical Association, 2017
- Associate Editor, **Journal of Agricultural, Biological, and Environmental Statistics, 2015**
- V. J. Berrocal, A. E. Gelfand, and D. M. Holland. (2010). A spatio-temporal downscaler for output from numerical models. Journal of Agricultural, Biological and Environmental Statistics, 15, 176-197.
- V. J. Berrocal, A. E. Gelfand, and D. M. Holland. (2014). Assessing exceedance of ozone standards: A space-time downscaler for fourth highest ozone concentrations. Environmetrics 25(4) · May 2014
- V. J. Berrocal, A. E. Gelfand, D. M. Holland, J. Burke, M. L. Miranda. (2011). On the use of a PM2:5 simulator to explain birthweight. Environmetrics, 22, 553-571.

Bo Li, PhD; Associate Professor, Department of Statistics, University of Illinois at Urbana-Champaign

libo@illinois.edu; 217-333-2167; http://www.stat.illinois.edu/people/faculty/boli.shtml

Relevant Experience and Selected Publications:

- Professor of Spatial Statistics and Analysis of Variance
- Young Investigator Award, Section on the Environment, American Statistical Association, 2011
- Associate Editor, **Journal of Agricultural, Biological, and Environmental Statistics, 2013**
- Li, B., Zhang, X. and Smerdon, J., Comparison between spatio-temporal random processes and application to climate model data (2016), *Environmetrics*, to appear.
- Li, B. and Smerdon, J. E., Defining spatial assessment metrics for evaluation of paleoclimatic field reconstructions of the Common Era (2012) *Environmetrics*, Vol. 23, 394-406.

Appendix B: Peer reviewer comments

Comments from peer reviewer 1

1) Are the relevant technical aspects of the statistical procedure clearly described? -- Yes.
a. Are input data (EPA's AQS) and their characteristics sufficiently described? -- Yes, the data is described clearly.
b. Is it clear what is being estimated? -- Yes, the ozone, annual PM2.5 and 24-hr PM2.5 NAAQS on page 8 is very clear.
c. Is the bootstrap procedure described in sufficient detail to allow reproduction? -- Yes, this is clear.

2) Are the descriptions of statistical concepts clear and accurate? -- Yes.
a. Are the descriptions of statistical significance and significance testing clearly and sufficiently described to assist the layperson in understanding the analysis? -- Although I am not a layperson in statistics, I think the concept is well explained in plain language.
b. Do the examples provided in the TBD illustrate the concepts of statistic sufficiently for the layperson to understand the analysis? -- Yes, they are straightforward to follow.

3) Are the assumptions and choices in the analysis clearly described and supported? -- Yes
a. Are the assumptions and choices in the analysis sufficiently documented? -- Yes, all details are well documented.
b. Does the document sufficiently describe the sensitivity of results to the choices and assumptions in the analysis? For example, are the technical considerations that support the policy decision to aggregate the variability to a single national value clearly articulated? -- Yes, the report carefully studied the spatial variability and the temporal variability for PM2.5 at different sampling frequencies. The bootstrap is applied appropriately, and the selection of 50% conference interval to obtain conservative SILs is reasonable. The selection of a single national value is not optimal considering the spatial variability, but taking the consistency of policy into consideration and given the fact that there are no large scale trends in ambient air variability are present, it is not unreasonable to have a single national value. Using the median rather the mean provides a more robust SIL for NAAQS.

4) Are the procedures appropriate for the analytical goals? -- Yes
a. Is bootstrapping an appropriate technique to quantify the variability in the air quality design value statistics? Is the bootstrapping analysis a reasonable approach to inform a policy determination of Significant Impact Levels (i.e., threshold levels)? -- Yes, the bootstrap is a sound statistical approach. it is very popular due to its flexibility that no parametric model or strong assumptions are required. The bootstrap is applied appropriately to quantify the variability in design values.

5) In your assessment, is there need for further analysis or clarification? Do you have suggestions for improving the document?

I read the document twice. At the first time, I was a little confused with what NAAQS represents in many places. My understanding of NAAQS is that it is a set of standards or thresholds for different statistics (or called DV here), but then it seems NAAQS is used more often as the statistics defined for NAAQS. For example, the x-axis labels in Figures 11 and 13 used NAAQS as the statistics. Although I finally realized what NAAQS often represents in the document, it might be more clear to explicitly point out that it is the statistics defined for NAAQS rather than the thresholds are actually referred to.

Page 5, The definition of "design value" is also confusing. The definition says it is "a statistic or summary metric based on the most recent one or three years ...". This seems to imply that the design value (DV) is a statistic or summary that is computed based on the sample of monitored data only for new source or modification. It seems to imply that the purpose of computing DV is to evaluate the contribution of source(s). However, later the DV is calculated based on all data measured during 2000-2014 and the results are used to derive SIL which if I understand correctly would serve the thresholds for NAAQS. I would suggest to remove "the most recent" in the definition on page 5 so it reads like "a statistic or summary metric based on one or three years ...".

Page 34 last paragraph, "using only the 1:1 monitors would produce smaller estimates of the variability". This is hard to understand intuitively. Suppose we have continuous observations in time, i.e., a continuous time series. Now we take daily values from this series for 1:1 monitors and also take values every three days for 1:3 monitors, then I expect that the daily values would exhibit no less if not more variability than the values every three days. Is there a better explanation from the characteristics of data collection for the smaller variability with 1:1 monitors? For example, since the 1:3 monitors collect data at different times during the day than the 1:1 monitors, these times may happen to have more variable PM2.5?

Page 11, line -2, ".5" seems redundant.

Page 39, first paragraph, line -5, suggests --> suggest

Comments from peer reviewer 2

Peer review of EPA's draft guidance and supporting documents recommending Significant Impact levels (SILs) for ozone and fine particle pollution that may be used in the Prevention of Significant Deterioration (PSD) permitting program

September 29, 2016

I was charged with examining the EPA's drafts of the guidance, legal, technical, and technical appendix documents regarding SILs for Ozone and $PM_{2.5}$. Overall I found the documents to contain sound and well-explained statistical methodology in order to identify ozone and $PM_{2.5}$ SILs for the US. Below I detail my responses to the charge questions.

1. **Are the relevant technical aspects of the statistical procedure clearly described?**

   a. Are input data (EPA's AQS) and their characteristics sufficiently described?

      Yes. Section 2.1 of the Technical Basis document provides details (e.g., where to access and how collected) about each data set, figures mapping the locations of the monitors, and details about the different types of monitors for each data set.

   b. Is it clear what is being estimated?

      Yes. Section 2.1 explicitly defines the DVs for primary ozone NAAQS, primary annual $PM_{2.5}$ NAAQS, and 24-hr $PM_{2.5}$ NAAQS. Section 1 describes the need for and the explanation of a SIL for each of these pollutants.

   c. Is the bootstrap procedure described in sufficient detail to allow reproduction?

      Yes. Section 2.2.3 describes the purpose of bootstrapping and a detailed procedure of how the bootstrap was implemented for each DV in this analysis. Following this outline, replication would be easily doable.

2. **Are the descriptions of statistical concepts clear and accurate?**

   a. Are the descriptions of statistical significance and significance testing clearly and sufficiently described to assist the layperson in understanding the analysis?

      Yes. Sections 1 and 2.2.1 describe statistical significance and "testing" (as it relates to confidence intervals) and connect these concepts to the SIL. Figure 3 is very useful for showing the difference between a 50% CI and 95% CI.

   b. Do the examples provided in the TBD [sic] illustrate the concepts of statistics sufficiently for the layperson to understand the analysis?

      *Yes. However, for clarification purposes, the hypothetical example on page 13 should start, "Suppose the **observed** annual mean $PM_{2.5}$ concentration…" to distinguish this number from the unobserved population mean, to which the previous paragraphs were referring.*

3. **Are the assumptions and choices in analysis clearly described and supported?**

    a. Are the assumptions and choices in the analysis sufficiently documented?

    Yes, the technical document describes all assumptions and modeling choices well.

    b. Does the document sufficiently describe the sensitivity of results to the choices and assumptions in the analysis? For example, are the technical considerations that support the policy decision to aggregate the variability to a single number clearly articulated?

    Yes, however, see part a.i and a.ii of my response to question 5.

4. **Are the procedures appropriate for the analytical goals?**

    a. Is bootstrapping an appropriate technique to quantify the variability in the air quality design value statistics? Is the bootstrapping analysis a reasonable approach to inform a policy determination of Significant Impact Levels?

    Yes. Bootstrapping is a method shown to perform well for quantifying uncertainty for a variety of statistics. That said, I have some concern about its ability to properly quantify the uncertainty for the 24-hr $PM_{2.5}$ DV, particularly for monitoring stations with 1:6 sampling frequency. At these sites, there are not many data points to capture much variability for the 98th percentile. However, these sites are relatively few and the DV is an average across three years, thus reducing potential bias. It's not a red flag, but it is something to consider moving forward with the analysis.

5. **In your assessment, is there need for further analysis or clarification? Do you have suggestions for improving the document?**

This document is well written and clearly defines statistical terms and meets the criteria defined therein. I make one suggestion for revision within the document (listed in item a.iii below; and a few typos are noted at the end of the document). While I don't think there is a need for further analysis at this time, I think future iterations of this analysis should consider two items:

a. Spatial variation.

i. The bootstrap method as implemented in this analysis does not account for the strong spatial dependence (described in Section 3.2.1). The researchers implement the bootstrap on each of the locations independent of the other locations. While this is fine for setting individual SILs, making use of spatial dependence within the bootstrap would be a more appropriate way to define a national SIL. Note that some measures have been taken to account for temporal dependence (i.e., insuring that observations sampled in each iteration of the bootstrap are observations from the same quarter), but nothing for spatial dependence.

ii. The discussion of the lack of evidence for regional SIL's is weak. Figures 11 and 14 show strong spatial dependence. Additionally, I would expect that different types of monitors (i.e., those with different sampling frequencies) will exhibit different relative uncertainties. I'd expect that monitors with less frequent measurements are more variable (and this is supported in Table 2) and therefore regional SILs could be considered for the different types of *monitors. The discussion for the desirability a national SIL is solid, but the spatial plots do not give enough evidence that regional SILs would be unreasonable to define.*

iii. *The discussion in the final paragraph of page 28 (Section 3.2.1) is poor. They are comparing two very different types of variation: variation between locations and variation within a location. This discussion should be revised or removed.*

b. *Consider a "Significant Impact Threshold." While the 50% CI for the SIL is well motivated as a value for insuring no difference (and the need for this type of a value rather than a threshold), the SIL will be used instead as a threshold limit, when in actuality, it's extremely plausible that values beyond the SIL will not "cause or contribute to an air quality violation." Providing a second level – or a threshold – of "will likely cause or contribute to an air quality violation" (e.g., a level corresponding to 99% or 99.9% CI) would be very valuable for decision makers in managing the individual cases (e.g., rather than the vague 1.2 vs 1.3 descriptions given in the current draft guidance document).*

A few typos:

- *Page 13, final paragraph: "normal distribution" and "Normal Distribution" are both used.*
- *Fourth line of the paragraph under Section 2.2.2.3: "…then* ***the*** *mean and the value…"*
- *Page 19: there's an out of place bolded "Error! Bookmark not defined."*
- *Parenthetical statement at the top of page 22: If q=50%, then the percentiles listed are correct. However, they are not correct for any value of q. The statement should read "the lower bound is the (50-q/2)% percentile and the upper bound is the (50+q/2)% percentile."*

Comments from peer reviewer 3

Response to charge questions:

1. **Are the relevant technical aspect of the statistical procedure clearly described?**

   a) **Are input data (EPA's AQS) and their characteristics sufficiently described?**

   In my opinion, the document presents the air quality data in a clear way: the description of the network design is very informative, as well as the description of the different types of spatial scale monitors employed for the two pollutants. Also the discussion of the issue of spatial and temporal variability were well presented and discussed.
   Potentially, a more extended explanation as for why the middle scale is not considered an appropriate spatial scale for PM2.5 could be useful.

   b) **Is it clear what is being estimated?**

   In general the description of the estimation procedure is rather clear, although there are parts of the documents on the estimation procedure that would benefit from a more thorough explanation.
   In more details: the document defines clearly the DV for the two pollutants and determines explicitly what the DVs are in relations to the different NAAQS. The document also clearly explains how the DVs are calculated in the resampled datasets: in particular the extended explanation on page 21 is really helpful. The explanation on how confidence intervals corresponding to different confidence levels are determined in the bootstrap framework is also rather clear. Less clear are the description of the statistics computed and presented in the Results section. Specifically, the document often refers to "difference between the bootstrapping CI value and the actual design value for a single monitoring site". This is quite confusing since a CI is an interval and thus defined by two bounds, while the actual design value at a monitoring site is a number, hence the term difference is rather ambiguous: is the difference between the design value and the upper bound of the bootstrapping CI or the difference between the design value and the lower bound of bootstrapping CI? The label on the horizontal axis of Figure 4 seems to indicate that both differences were calculated (similarly for the axis of Figure 6), however both the text in page 23 and 25 as well as the caption to Figure 4 and 6 is ambiguous. Similarly, the middle panel of Figure 4 and the bottom two panels in Figure 5 are rather confusing and do not present information on the quantities being estimated in an unambiguous way.

   c) **Is the bootstrap procedure described in sufficient detail to allow reproduction?**

   I believe that the explanation of the calculation of a bootstrap CI provided in page 21 clarified greatly the description of the bootstrap procedure given in page 20 and provided enough detail for reproduction.

2. **Are the descriptions of statistical concepts clear and accurate?**
   a) **Are the description of statistical significance and significance testing clearly and sufficiently described to assist the layperson in understanding the analysis?**

   In general I think the document does a very good job at presenting statistical concepts to the layperson. The idea of a sample being a representative of the population, the concept of hypothesis testing, the interpretation of the results of an hypothesis test, and the concept statistical significance were all well described. To my opinion, in certain parts the document is not completely precise from a statistical point of view, and I think that a revision of the document to address and correct these slight imprecisions would be ideal.
   For example, on page 13 when the document discusses the derivation of confidence intervals, the way the text is written seems to imply that all confidence intervals are derived based on sampling distributions and Central Limit Theorem. While all confidence intervals are derived based on the asymptotic behavior of the sampling distributions, the Central Limit Theorem is a theorem that states the asymptotic behavior of the sampling distribution only of the mean of independent random variables. Thus it could only be used to derive confidence intervals of parameters that can be thought as the mean of a sequence of independent random variables. Calculation of the confidence intervals for other parameters, such as for example the variance, is not based on the Central Limit Theorem, although it is based on the asymptotic behavior of the sampling distribution of the variance.
   A second small imprecision is on page 18 when the document discusses assessing the air quality variability: in section 2.2.2.3 it uses the incorrect language "the CI of the sample mean": confidence intervals are not intervals for the sample estimators, but they are intervals for the population parameters. Hence, there "the CI of the sample mean" should be replaced with "CI of the mean".
   Besides these small imprecision, the description of statistical concepts is quite clear.

   b) **Do the examples provided in the TBD illustrate the concepts of statistics sufficiently for the layperson to understand the analysis?**

   I think that the examples in the document are instrumental for the layperson to completely grasp and understand the statistical concepts presented in the document. I also think that they are well explained and presented.

3. **Are the assumptions and choices in the analysis clearly described?**
a) Are the assumption and choices in the analysis sufficiently documented?
I don't think that the assumptions underlying the analyses are always sufficiently discussed. For example, an underlying assumption of bootstrapping, at least in the implementation of bootstrapping used in the analysis reported in the document, is that the data is considered to be observations of independent random variables. The

document does not explicitly state this underlying assumption, which will translate into assuming that ozone and PM2.5 daily monitoring values at a given sites are independent. This is a strong assumption underlying bootstrapping that the document does not mention openly.
On the other hand, other choices, such as bootstrapping the data within each year independently, resampling data from each 3-month period have been clearly explained and documented.

b) **Does the document sufficiently describe the sensitivity of results to the choices and assumptions in the analysis? For example, are the technical considerations that support the policy decision to aggregate the variability to a single national value clearly articulated?**

I have found this part of the document (e.g. Section 4) very unclear and not well explained, especially in comparison with the rest of the document. To my opinion sensitivity of the results to the choices and assumptions of the analyses are not discussed at all, and I think that these two points should be addressed in a revised version of the document.

4. **Are the procedure appropriate for the analytical goals?**
   a) **Is bootstrapping an appropriate technique to quantify the variability in the air quality design value statistics? Is the bootstrapping analysis a reasonable approach to inform a policy determination of Significant Impact Level (e.g. threshold levels)?**

I think that in a nutshell, as general procedure, bootstrap is an appropriate technique to quantify the variability in the air quality design value statistics, especially given that the design value statistics are based on percentiles of the distributions. Thus, given that the sampling distributions of the DV might not be available, bootstrapping can be a mean to quantify the variability and thus derive CI. I also believe that bootstrapping analysis is a reasonable approach to determine Significant Impact Level.
My point of contention with the analysis is that I am not sure that I completely agree with the way bootstrap has been implemented. In particular, I believe that ozone and PM2.5 concentration values at a site are fairly correlated from day to day, and thus the air quality data for a given site might display a significant auto-correlation at lag 1 (meaning that concentrations of ozone measured at a site a day apart are very likely significantly correlated), and might have a significant auto-correlation at longer lags depending on the season. Bootstrapping, in the way it has been implemented in the document, according to the document description, is based on the assumption that the observations are independent, which might not be the case for ozone concentrations. The sampling frequency of PM2.5 concentrations at the monitoring sites might render the PM2.5 data independent, however it is an assumption that should be verified.
Thus, while on a conceptual level, I think that bootstrapping could be used as a reasonable approach for deriving SILs, I think that in the actual

implementation of the bootstrapping method, it needs to be attested whether the observed ozone and PM2.5 concentration data within each 3-month period is independent. In case the assumption of independence is violated, bootstrapping method for temporally correlated data should be used in deriving the re-sampled datasets.

5. In your assessment is there need for further analysis or clarification? Do you have suggestions for improving the document?

As mentioned in the reply to Charge Question 4 above, I believe that there is need for further analysis. In particular I think that the issue of temporal autocorrelation in the data at each site has to be investigated and necessary correction to the bootstrap techniques should be implemented.

In terms of improvement to the documents, I think that the first two sections of the documents are well written and presented and, except for the few corrections suggested above, I do not see much need for improvements in those sections. I believe that the presentation of the results in Section 3 could be improved by clearly stating what are the statistics computed. Finally, as mentioned in the reply to question 3, I believe that Section 4 of the document is quite unclear and the document would improve greatly if a more exhaustive explanation of the considerations in Section 4 is provided.

United States Environmental Protection Agency	Office of Air Quality Planning and Standards Air Quality Analysis Division Research Triangle Park, NC	Publication No. EPA-454/S-18-001 March, 2018